SCIENTIFIC REPORTS
natureresearch

OPEN

What does fluorine do to a protein? Thermodynamic, and highly-resolved structural insights into fluorine-labelled variants of the cold shock protein

Hannah Welte[1,2], Tiankun Zhou[3], Xenia Mihajlenko[1], Olga Mayans[2,3] & Michael Kovermann[1,2,4*]

Fluorine labelling represents one promising approach to study proteins in their native environment due to efficient suppressing of background signals. Here, we systematically probe inherent thermodynamic and structural characteristics of the Cold shock protein B from *Bacillus subtilis* (*Bs*CspB) upon fluorine labelling. A sophisticated combination of fluorescence and NMR experiments has been applied to elucidate potential perturbations due to insertion of fluorine into the protein. We show that single fluorine labelling of phenylalanine or tryptophan residues has neither significant impact on thermodynamic stability nor on folding kinetics compared to wild type *Bs*CspB. Structure determination of fluorinated phenylalanine and tryptophan labelled *Bs*CspB using X-ray crystallography reveals no displacements even for the orientation of fluorinated aromatic side chains in comparison to wild type *Bs*CspB. Hence we propose that single fluorinated phenylalanine and tryptophan residues used for protein labelling may serve as ideal probes to reliably characterize inherent features of proteins that are present in a highly biological context like the cell.

Proteins are key operating elements of complex biological systems, such as cells. These macromolecules control a multiplicity of chemical processes and are a central characteristic of living systems[1]. The determination of the three-dimensional structure of a protein and the identification of its potential interacting partners is crucial to understand the function of these macromolecules in biological systems. However, most studies aiming to unravel the intracellular functionality of proteins are performed *in vitro*, using dilute experimental conditions that are distant from the crowded cellular environments that proteins experience in living organisms. The total intracellular macromolecular concentration can be estimated with up to about 400 g/l[2,3] whereas individual proteins are often present in tiny amounts. Thus, performing protein studies using *in vitro* conditions neglects the intracellular environment and does not account for a variety of effects like e.g. crowding[4], confinement[5] or quinary interactions[6]. In the recent years, significant progress has been made on the study of proteins in native cellular environments. In this regard, high-resolution NMR spectroscopy is a promising technique that offers indispensable insights into the structure, dynamics, and stability of the protein under study in an intracellular context[7–9]. The specific labelling of proteins using fluorine[10,11] may represent a successful strategy to perform *in cellula* NMR experiments. Once the protein has been successfully labelled using fluorine, the determination of kinetic and thermodynamic parameters (such as the change in enthalpy, ΔH, the change in heat capacity, ΔC_P, upon protein unfolding, and the folding and unfolding rate constants) is possible even using cell-like conditions[12–14]. Thus, the labelling of proteins using fluorine allows conducting a broad variety of NMR spectroscopic experiments focusing on ^{19}F resonances. This approach offers several advantages and simplifications compared to experiments commonly used

[1]Department of Chemistry, Universitätsstrasse 10, Universität Konstanz, DE-78457, Konstanz, Germany. [2]Graduate School Chemical Biology KoRS-CB, Universitätsstrasse 10, Universität Konstanz, DE-78457, Konstanz, Germany. [3]Department of Biology, Universitätsstrasse 10, Universität Konstanz, DE-78457, Konstanz, Germany. [4]Zukunftskolleg, Universitätsstrasse 10, Universität Konstanz, DE-78457, Konstanz, Germany. *email: michael.kovermann@uni-konstanz.de

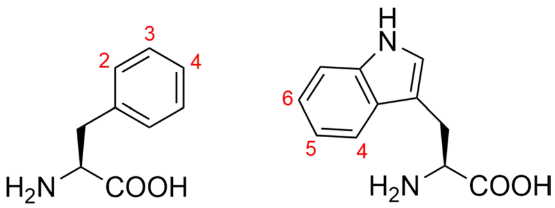

Figure 1. Numbering of sites used for single fluorine labelling in 2-^{19}F-, 3-^{19}F-, and 4-^{19}F-phenylalanine (left) and 4-^{19}F-, 5-^{19}F-, and 6-^{19}F-tryptophan (right). The structures have been created by using ChemDraw18 (www.perkinelmer.com).

in biomolecular NMR spectroscopy that focus on ^{1}H, ^{13}C, and ^{15}N nuclei[14,15]. In this respect, fluorine has a natural abundance of 100%, a high sensitivity (83% compared to ^{1}H), a large range of chemical shifts and no natural occurrence in proteins enabling to efficiently suppress undesired background signals. Thereby, structural, dynamical, and functional protein information can be obtained at atomic resolution in complex biological environments, even if the protein of interest is only present in a relatively low concentration[12,15,16]. We do not conceal that fluorine possesses large chemical shift anisotropy[17] which may potentially impede assignment and interpretation of fluorine NMR spectra. However, as leading examples of the applicability of ^{19}F-derivatized proteins, fluorotryptophan has been incorporated into the carbohydrate binding protein lectin from *Ralstonia solanacearum* enabling monitoring its interaction with ligands at atomic detail by X-ray crystallography[18], whereas ^{19}F NMR spectroscopy has been successfully applied to reveal the exchange dynamics present in antibody-antigen binding, an information that is not accessible using X-ray crystallography[19]. Furthermore, selective fluorine-labelling has been successfully used to derive distance restraints for a protein based on ^{19}F paramagnetic relaxation enhancement experiments[20] whereas the interaction sites present in protein-ligand complexes have been elucidated by applying ^{19}F pseudocontact shift analysis[21]. An elegant combination of X-ray crystallography, NMR spectroscopy, and MD simulations has been recently applied to unravel the distribution of conformational states along the reaction pathway of a homodimeric enzyme[22]. Pomerantz and co-workers have successfully utilized fluoroaromatic amino acids for the implementation into proteins to report on the structure-activity relationship[23] focusing on screening of small molecule-protein interactions based on protein-observed fluorine NMR spectroscopy[24]. Moreover, incorporation of difluorotyrosine has been successfully employed enabling to report on tyrosine phosphorylation[25] or to probe distinct conformational states of a protein which are related to signalling by applying ^{19}F NMR spectroscopy[26]. In a recent study, an NMR based comparison of the incorporation of 2- and 3-fluorotyrosine into a KIX domain has been presented[27].

The outstanding question in the application of ^{19}F-NMR based methodologies is how much does the ^{19}F-modification impact the inherent properties of the protein under study, particularly its atomic three-dimensional structure, conformational dynamics, and its overall thermodynamic stability. Addressing this question is of high importance as the fluorine-labelled protein variant, and not the wild type, is used in *in cellula* NMR spectroscopy to report on the native structural and dynamical features of proteins *in vivo*. In this context, it has been shown that extensively fluorinated amino acids can be particularly effective in increasing protein stability[28] due to the increase in buried hydrophobic surface area as identified in structures solved by X-ray crystallography[29]. Kitevski-LeBlanc and co-workers have shown that ^{19}F enrichment in fluoro-phenyalanine labelled calmodulin results in an increasing protein disorder which can be diminished by a decreased level of fluorination[30]. Other work has shown that the structural integrity of a small single domain protein is conserved when one fluoro-phenylalanine is incorporated[31] and that fluoro-tryptophan labelling of various sites in fluoroacetate dehalogenase does not modify its three-dimensional structural characteristics compared to the wild type[22].

In previous work, we have used auxotrophic *E. coli* strains to incorporate fluorinated phenylalanine (Phe) or tryptophan (Trp) amino acid residues into the Cold shock protein B from *Bacillus subtilis* (*Bs*CspB). Within these residues, we used different sites for the incorporation of the fluorine atoms, generating a total of three Phe and three Trp modified amino acids (Fig. 1)[32]. Using this tool kit, we were then able to successfully prepare six different ^{19}F-labelled *Bs*CspB variants (2-^{19}F-Phe-, 3-^{19}F-Phe-, 4-^{19}F-Phe-, 4-^{19}F-Trp-, 5-^{19}F-Trp- and 6-^{19}F-Trp-*Bs*CpB) at a milligram scale and in high purity enabling high resolution one-dimensional ^{1}H, ^{19}F and two-dimensional ^{1}H-^{15}N correlation NMR spectroscopy of the samples mentioned above[32]. *Bs*CspB is a relatively small protein of 67 amino acids length that belongs to the cold shock protein family[33–35]. It folds into five beta-strands that form a beta-barrel structure[36,37], which binds single-stranded DNA and RNA[38,39]. Notably, *Bs*CspB shows an overall thermodynamic stability of about $\Delta G° = 10\,\text{kJ mol}^{-1}$ at $T = 298\,\text{K}$ and possesses fast unfolding, k_u, and refolding, k_f, rate constants of about $k_u = 12\,\text{s}^{-1}$ and $k_f = 1070\,\text{s}^{-1}$, respectively[33]. Here, we focus on the precise determination of thermodynamic and kinetic parameters of all six fluorine-labelled protein variants in comparison to wild type *Bs*CspB. A combination of fluorescence, kinetic stopped flow, NMR spectroscopy, and X-ray crystallography has been applied to obtain an integrated understanding of potential effects of fluorine-labelling in proteins. Our findings show that fluorine-labelling of proteins utilizing singly modified amino acids like Phe or Trp does not cause a detectable alteration of thermodynamic and structural properties of *Bs*CspB. The present study closes an important gap in the basic characterization of fluorine-labelled proteins and underlines that this methodology may serve as an optimal tool to study proteins in their native complex biological environment applying high-resolution NMR spectroscopy.

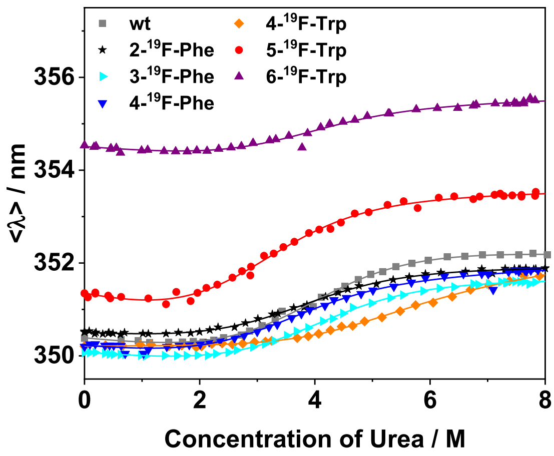

Figure 2. Dependence of average emission wavelength on the concentration of urea obtained from fluorescence spectra for all protein variants probed in this study (rectangles represent wild type protein, stars represent 2-^{19}F-Phe-, triangles with the tip to right represent 3-^{19}F-Phe-, triangles with the tip to bottom represent 4-^{19}F-Phe-, diamonds represent 4-^{19}F-Trp-, circles represent 5-^{19}F-Trp-, and triangles with the tip to top represent 6-^{19}F-Trp-variant). The straight lines represent best fits of Eq. (3) to the experimental data sharing the same value for the cooperativity of unfolding, *m*. Results of this fitting procedure are summarized in Table 1.

Protein	ΔG^0/kJ/mol	C_M/M
wt *Bs*CspB	11.1 ± 0.5	3.8 ± 0.2
2-^{19}F-Phe-*Bs*CspB	10.0 ± 0.7	3.5 ± 0.2
3-^{19}F-Phe-*Bs*CspB	10.5 ± 0.6	3.6 ± 0.2
4-^{19}F-Phe-*Bs*CspB	9.6 ± 0.6	3.3 ± 0.2
4-^{19}F-Trp-*Bs*CspB	12.7 ± 1.1	4.4 ± 0.4
5-^{19}F-Trp-*Bs*CspB	9.0 ± 0.5	3.1 ± 0.2
6-^{19}F-Trp-*Bs*CspB	11.0 ± 0.7	3.8 ± 0.2

Table 1. Analysis of equilibrium unfolding of wild type and fluorine-labelled variants of *Bs*CspB using *fluorescence spectroscopy* (Fig. 2) to determine the overall thermodynamic stability, ΔG^0. The cooperativity of protein unfolding has been used as global parameter in Eq. (3) taking all seven folded-to-unfolding transitions into account and has been determined to $m = -2.9 \pm 0.1$ kJ/(mol M). The midpoint of folded-to-unfolding transition, C_M, has been calculated using $C_M = \Delta G^0/m$.

Results and Discussion

Thermodynamic stability probed by fluorescence in equilibrium. Intrinsic fluorescence spectroscopy has been applied to probe the overall thermodynamic stability of the six differently fluorine-labelled variants of *Bs*CspB. For this purpose, folded-to-unfolding transitions of *Bs*CspB samples were chemically induced with increasing amounts of urea and monitored using fluorescence spectroscopy in equilibrium. Specifically, fluorescence emission spectra of wild type and 2-^{19}F-Phe-, 3-^{19}F-Phe-, 4-^{19}F-Phe-, 4-^{19}F-Trp-, 5-^{19}F-Trp-, and 6-^{19}F-Trp-labelled *Bs*CspB were measured in urea concentrations ranging from $c_{urea} = 0$ M to $c_{urea} = 8$ M. Note that all seven intrinsic Phe residues Phe9, Phe15, Phe17, Phe27, Phe30, Phe38, and Phe45 present in *Bs*CspB have been equally fluorine-labelled in 2-^{19}F-Phe-, 3-^{19}F-Phe- or 4-^{19}F-Phe-modified protein samples[32] (Fig. S1A) whereas 4-^{19}F-Trp-, 5-^{19}F-Trp-, and 6-^{19}F-Trp-*Bs*CspB possess Trp8 as single site of the modification (Fig. S1B). Fluorescence spectra showed that, upon unfolding, the maximal fluorescence emission intensity decreased (Figs. S2A–S8A) and shifted to significantly larger wavelengths (Figs. S2B–S7B) for both wild type and fluorine-labelled variants. The signal shifting to larger wavelengths regarding fluorescence emission of tryptophans is consistent with the aromatic group becoming more exposed to the polar solvent upon unfolding of the protein. The properties observed here for all six differently fluorine-labelled variants of *Bs*CspB regarding fluorescence emission upon chemical unfolding have been also reported for the wild type[33]. Solely 6-^{19}F-Trp-*Bs*CspB does not show a dependence of the maximum wavelength in fluorescence emission, λ^{max}, on the concentration of urea (Fig. S8B) but preserving a decrease in fluorescence emission intensity by an increasing concentration of urea (Fig. S8A). The lack in the increase of the maximum wavelength in fluorescence emission observed for the 6-^{19}F-Trp-labelled variant can be anticipated as λ^{max} is about 360 nm already in the absence of urea (Fig. S8B). However, λ^{max} observed for 6-^{19}F-Trp-*Bs*CspB has been also approached for remaining variants of *Bs*CspB at high concentrations of urea (Figs. S2B–S7B). The fluorescence emission intensity is lowest for 4-^{19}F-Trp-*Bs*CspB among all probed fluorine-labelled protein variants (Fig. S6A). This observation is based on the inherent low fluorescence quantum yield of 4-^{19}F-Trp[40] which is about 100 times lower compared to Trp, 5-^{19}F-Trp, and 6-^{19}F-Trp (Fig. S9). For fluorescence data analysis the intensity emission averaged wavelength, $<\lambda>$, has been determined (Fig. 2). This procedure is based on two rationales. Firstly, the maximum of fluorescence emission of *Bs*CspB variants shifts by an increasing concentration of urea to larger wavelengths (see above). Secondly, the maximum of fluorescence emission is dependent on the site used for inserting fluorine into free tryptophan residues (Fig. S9). Thus the free form of native Trp shows a maximum in fluorescence emission at $\lambda^{max} = 353$ nm, free 5-^{19}F-Trp at $\lambda^{max} = 360$ nm, free 6-^{19}F-Trp at $\lambda^{max} = 360$ nm, and free 4-^{19}F-Trp $\lambda^{max} = 376$ nm using $\lambda = 280$ nm for excitation going along with observations on free fluorotryptophans performed before using $\lambda = 295$ nm for excitation[40].

The intensity averaged emission wavelength $<\lambda>$ is calculated using Eq. (1)

$$<\lambda> = \sum_{i=320.5\text{ nm}}^{380\text{ nm}} (I_i\lambda_i) / \sum_{i=320.5\text{ nm}}^{380\text{ nm}} (I_i), \tag{1}$$

where I_i is the intensity of fluorescence emission at wavelength λ_i[41].

This method of fluorescence data analysis accounts for both potential changes in the shape of acquired fluorescence emission spectra and a potential shift in wavelength of the maximum in fluorescence emission, respectively. Note that measuring the refractive index of the buffer accounts for the precise determination of the concentration of urea which has been individually used (Eq. 2).

Applying a two state folding-to-unfolding model[42] (Eq. 3) to the experimentally obtained fluorescence data yields the free energy against unfolding, $\Delta G°$, of wild type *Bs*CspB and all six variants differing in the site used for fluorine labelling. The cooperativity of protein unfolding, m, is then used as a global fitting parameter enabling a direct comparison of the overall thermodynamic stability of all seven different protein variants probed in the present study. This approach results in $m = -2.9 \pm 0.1$ kJ/(mol M), which agrees well with m values which have been reported for *Bs*CspB before, $m = -3.2 \pm 0.1$ kJ/(mol M)[33,43]. The free energy of protein unfolding ranges between $\Delta G° = 9.0 \pm 0.5$ kJ/mol obtained for 5-^{19}F-Trp-BsCspB and $\Delta G° = 12.7 \pm 1.1$ kJ/mol for 4-^{19}F-Trp-*Bs*CspB (Table 1). Thus, the overall thermodynamic stability of fluorine-labelled *Bs*CspB variants reproduces the stability of wild type *Bs*CspB, determined here using the same type of data analysis ($\Delta G° = 11.1 \pm 0.5$ kJ/mol) (Table 1). Note that linear fitting of fluorescence emission intensities characterizing the transition region between folded and unfolded state has been also performed for all variants of *Bs*CspB[44] (Fig. S10, Table S1). This procedure ensures that the determination of the overall thermodynamic stability $\Delta G°$ is not primarily influenced by missing baselines representing the folded as well as the unfolded state as it has been shown by Alexander and Pace[44]. Notably, both approaches used for the analysis of fluorescence emission intensities yield consistent results for $\Delta G°$ (Table S1).

Summing up, fluorine labelling of either tryptophan or phenylalanine residues does not have a major effect on the overall thermodynamic stability of cold shock protein B when monitoring fluorescence emission dependent on an increasing concentration of urea. The overall thermodynamic stability determined for 4-^{19}F-Trp-*Bs*CspB is highest comparing all variants probed in the present study (Table 1) and deviates in $\Delta G°$ by 1.6 kJ/mol from wild type *Bs*CspB. This property of 4-^{19}F-Trp-*Bs*CspB has been one rationale to elucidate the structural features of this fluorine-labelled variant of *Bs*CspB by using X-ray crystallography in more detail. Focusing on phenylalanine modified variants, 4-^{19}F-Phe-*Bs*CspB deviates most in $\Delta G°$ from wild type *Bs*CspB by 1.5 kJ/mol. This property of 4-^{19}F-Phe-*Bs*CspB has been one rationale for the crystallization of this variant enabling to obtain atomically resolved information (see below).

Thermodynamic stability probed by NMR spectroscopy. The characterization of the thermodynamic stability of differently fluorine-labelled variants of *Bs*CspB has been expanded to folded-to-unfolding transitions induced by increasing the temperature. The thermal denaturation of the six ^{19}F-labelled *Bs*CspB variants and wild type *Bs*CspB has been monitored by one-dimensional ^{1}H (Fig. S11) and ^{19}F NMR spectroscopy (holds for 4-^{19}F-Phe-*Bs*CspB) to independently verify the experimental results obtained for chemical denaturation monitored using intrinsic fluorescence emission (see above). Advantageously, NMR spectroscopy provides direct spectroscopic information about the folded-to-unfolding transition of the entire protein under investigation whereas fluorescence spectroscopy operates as a local reporter of aromatic residues only. Thus, one-dimensional ^{1}H and ^{19}F NMR spectra have been acquired for different variants of *Bs*CspB including the wild type in the temperature range $T = 291$ K to $T = 330$ K (Figs. S11A,B–S17A,B). The fitting of Eq. (4) to the experimentally obtained data representing the fraction of unfolded protein gives access to the change in heat capacity, ΔC_P, the change in enthalpy, ΔH, and the temperature midpoint characterizing the folded-to-unfolding transition, T_M, for all variants of *Bs*CspB which have been probed here. Note that the change in heat capacity has been fixed to $\Delta C_P = 5.8$ kJ/(mol K) here as it has been specifically determined for *Bs*CspB using ^{1}H NMR spectroscopy before relying on both heat and cold denaturing[45]. The change in enthalpy has been used as a global parameter for all variants of *Bs*CspB when fitting Eq. (4) to the experimental data acquired by one-dimensional ^{1}H spectroscopy whereas T_M has been individually determined for all variants (Fig. 3A, Table 2). This procedure leads to $\Delta H = 197 \pm 2$ kJ/mol and values for T_M ranging between $T_M = 315.6$ K (2-^{19}F-Phe-*Bs*CspB) and $T_M = 320.8$ K (4-^{19}F-Trp-*Bs*CspB). Note that wild type *Bs*CspB shows $T_M = 316.8$ K indicating that there exists no significant difference in thermodynamic stability to the variously fluorine-labelled variants of *Bs*CspB. This result independently verifies the observation made for the chemical denaturation of *Bs*CspB variants presented before. The fitting of Eq. (4) to the data obtained for 4-^{19}F-Phe-*Bs*CspB using one-dimensional ^{19}F NMR spectroscopy yield $T_M = 318.4$ K and $\Delta H = 176 \pm 6$ kJ/mol (Fig. 3B) confirming the thermodynamic analysis done by using ^{1}H NMR spectroscopic data which has illuminated $T_M = 318.5$ K (Table 2). Note that a precise thermodynamic analysis of ^{19}F NMR detected folding-to-unfolding transitions of 2-^{19}F-Phe-, 3-^{19}F-Phe-, 4-^{19}F-Trp-, 5-^{19}F-Trp-, and 6-^{19}F-Trp-*Bs*CspB has been prohibited due to spectral indistinguishability of signals representing either the native state or the unfolded ensemble of the protein under study.

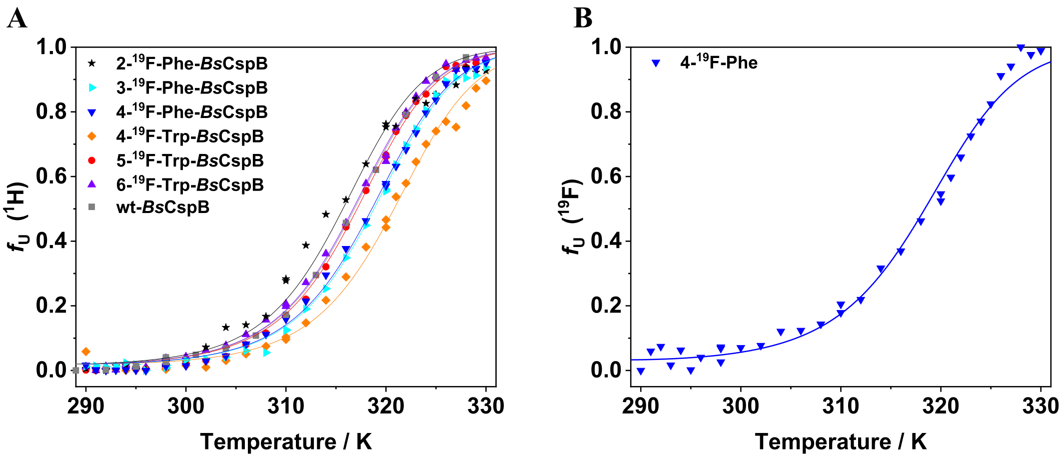

Figure 3. Folded-to-unfolding transitions monitored by using one-dimensional ^{1}H (**A**) and ^{19}F (**B**) NMR spectroscopy. The thermal denaturation has been probed for wild type (rectangles), 2-^{19}F-Phe- (stars), 3-^{19}F-Phe- (triangles with the tip to right), 4-^{19}F-Phe- (triangles with the tip to bottom), 4-^{19}F-Trp- (diamonds), 5-^{19}F-Trp- (circles), and 6-^{19}F-Trp-variant (triangles with the tip to top). The straight lines represent best fits of Eq. (4) to the experimental data sharing the same value for the change in enthalpy, ΔH. Results of this fitting procedure are summarized in Table 2.

Protein	T_M/K
wt *Bs*CspB	316.8 ± 0.2
2-^{19}F-Phe-*Bs*CspB	315.6 ± 0.1
3-^{19}F-Phe-*Bs*CspB	318.7 ± 0.1
4-^{19}F-Phe-*Bs*CspB	318.5 ± 0.1
4-^{19}F-Trp-*Bs*CspB	320.8 ± 0.1
5-^{19}F-Trp-*Bs*CspB	317.1 ± 0.1
6-^{19}F-Trp-*Bs*CspB	316.6 ± 0.1
4-^{19}F-Phe-*Bs*CspB (^{19}F)	318.4 ± 0.2

Table 2. Analysis of equilibrium unfolding of wild type and fluorine-labelled variants of *Bs*CspB using one-dimensional ^{1}H and ^{19}F *NMR spectroscopy* to determine the midpoint of folded-to-unfolding transition, T_M. The change in enthalpy of protein unfolding at T_M, $\Delta H(T_M)$, has been used as global parameter in Eq. (4) taking all seven ^{1}H-detected folded-to-unfolding transitions into account and has been determined to $\Delta H(T_M) = 197 \pm 2$ kJ/mol. For ^{19}F detection, the change in enthalpy has been determined to $\Delta H(T_M) = 176 \pm 6$ kJ/mol. The change in heat capacity has been fixed to $\Delta C_P = 5.8$ kJ/(mol K) as it has been specifically determined for *Bs*CspB using NMR spectroscopy before relying on both heat and cold denaturing[45].

The thermodynamic parameters calculated using thermal denaturation probed with NMR spectroscopy (Table 2) show excellent agreement with those derived by chemical denaturation monitored using intrinsic fluorescence (Table 1). Thus, 4-^{19}F-Trp-*Bs*CspB shows a small increase in overall thermodynamic stability (approx. 4 Kelvin) compared to wild type *Bs*CspB whereas 2-^{19}F-Phe-*Bs*CspB and 5-^{19}F-Trp-*Bs*CspB variants show a small decrease or uniformity in $\Delta G°$ (Table 1) and T_M (Table 2), respectively.

Unfolding and refolding kinetics probed by kinetic stopped-flow fluorescence. The thermodynamic analysis of all six differently fluorine-labelled variants has been further extended to kinetic experiments using a stopped-flow fluorescence device. This setup enables to monitor the change in fluorescence on a millisecond time scale induced by rapid mixing of two solutions representing either folding or unfolding conditions[46]. A monoexponential function has been applied to obtain the apparent rate constant, k_{obs}, to account for refolding (Fig. S18) and unfolding (Fig. S19) kinetics of fluorine-labelled variants and wild type *Bs*CspB. Having values for k_{obs} representing final concentrations of urea in the mixing cell ranging between $c_{urea} = 2.3$ M and $c_{urea} = 7.3$ M in hands, Eq. (5) has been applied to obtain rate constants for refolding, k_f, and unfolding, k_u, in absence of any denaturant, respectively. Moreover, Eq. (5) enables the determination of the slope of the unfolding, m_u, and the refolding limb, m_f, respectively. The fitting of Eq. (5) to the experimental data has been performed assuming a global value for the cooperativity of folding, $m = m_f + m_u$, for all differently fluorine-labelled variants of *Bs*CspB including wild type protein. The Chevron plot analysis reveals a two-state folding process for all variants of *Bs*CspB which have been kinetically probed (Fig. 4). The cooperativity of folding has been determined to $m = -2.8 \pm 0.8$ kJ/(mol M) matching the value of $m = -2.9 \pm 0.1$ kJ/(mol M) which has been independently determined in the present study by monitoring the folded-to-unfolding transition of all variants of *Bs*CspB using equilibrium fluorescence (Fig. 2, Table 1). Comparing the kinetic rate constants for unfolding of fluorine-labelled variants of *Bs*CspB ($31\,s^{-1} \le k_u \le 57\,s^{-1}$) does not reveal

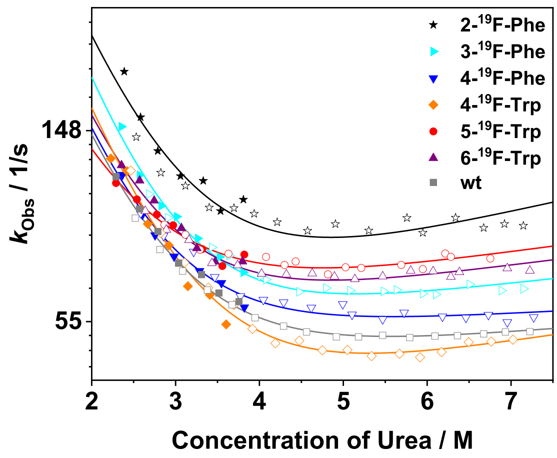

Figure 4. Dependence of the apparent rate constant on the concentration of urea characterizing unfolding (open symbols) and refolding (closed symbols) of fluorine-labelled variants of *Bs*CspB (stars represent 2-^{19}F-Phe-, triangles with the tip to right represent 3-^{19}F-Phe-, triangles with the tip to bottom represent 4-^{19}F-Phe-, diamonds represent 4-^{19}F-Trp-, circles represent 5-^{19}F-Trp-, and triangles with the tip to top represent 6-^{19}F-Trp-variant), and wild type protein (represented using rectangles). The straight lines represent best fits of Eq. (5) to the experimental data sharing the same value for the cooperativity of folding, $m = m_f + m_u$. Results of this fitting procedure are summarized in Table 3.

Protein	k_u/s^{-1}	k_f/s^{-1}	m_u/kJ/(molM)	m_f/kJ/(molM)	ΔG^0/kJ/mol
wt *Bs*CspB	40 ± 1	1050 ± 20	0.1	−2.9	8.1 ± 0.1
2-^{19}F-Phe-*Bs*CspB	56 ± 2	2000 ± 100	0.2	−3.0	8.9 ± 0.2
3-^{19}F-Phe-*Bs*CspB	42 ± 1	1600 ± 40	0.2	−3.0	9.0 ± 0.1
4-^{19}F-Phe-*Bs*CspB	44 ± 1	1050 ± 30	0.1	−2.9	7.9 ± 0.1
4-^{19}F-Trp-*Bs*CspB	31 ± 1	1450 ± 30	0.2	−3.0	9.5 ± 0.1
5-^{19}F-Trp-*Bs*CspB	57 ± 1	770 ± 20	0.1	−2.9	6.4 ± 0.1
6-^{19}F-Trp-*Bs*CspB	48 ± 1	1160 ± 30	0.2	−3.0	7.9 ± 0.1

Table 3. Analysis of unfolding and refolding kinetics of wild type and fluorine-labelled variants of *Bs*CspB using *kinetic stopped-flow fluorescence*. The cooperativity of folding, $m = m_f + m_u$, has been used as global parameter in Eq. (5) taking all seven variants of *Bs*CspB into account and has been determined to $m = -2.8 \pm 0.8$ kJ/(mol M). The value of ΔG^0 has been computed by using Eq. (6).

a significant difference to wild type *Bs*CspB ($k_u = 40 \pm 1\ s^{-1}$, Table 3). We note that refolding kinetics for differently fluorine-labelled variants of *Bs*CspB using $c_{urea} < 2.3$ M could not be reliably acquired. However, obtaining refolding rate constants for all differently fluorine-labelled variants of *Bs*CspB by applying Eq. (5) enables a comparison among individual k_f values in a qualitative manner as all kinetic data have been analyzed in the same way. Thus, the refolding rate constants which have been determined for all fluorine-labelled variants of *Bs*CspB are slightly higher (except for 5-^{19}F-Trp-*Bs*CspB) compared to $k_f = 1050 \pm 20\ s^{-1}$ obtained for wild type *Bs*CspB (Table 3). The determination of the rate constants accounting for refolding und unfolding enables to apply Eq. (6) to compute the difference in free energy, $\Delta G°$, of all variants of *Bs*CspB. As a result, these values are highly similar (Table 3) ranging between $\Delta G° = 6.4 \pm 0.1$ kJ/mol (5-^{19}F-Trp-*Bs*CspB) and $\Delta G° = 9.5 \pm 0.1$ kJ/mol (4-^{19}F-Trp-*Bs*CspB) covering the thermodynamic stability determined for wild type ($\Delta G° = 8.1 \pm 0.1$ kJ/mol). Comparing the values of $\Delta G°$ determined for all differently fluorine-labelled variants of *Bs*CspB by applying a kinetic setup with the overall thermodynamic stability obtained in equilibrium using chemical (Fig. 2, Table 1) or thermal denaturation (Fig. 3, Table 2) reveals two main features. Firstly, the overall thermodynamic stability of *Bs*CspB does not change significantly if fluorine has been attached either to all seven phenylalanine residues or to its single tryptophan residue. Secondly, it has been consistently shown by three independent experimental methods that 4-^{19}F-Trp-*Bs*CspB exhibits a small increase in $\Delta G°$ by about 1 kJ/mol and in T_M by about 4 Kelvin whereas 5-^{19}F-Trp-*Bs*CspB shows a moderate decrease in $\Delta G°$ by about 1.5 kJ/mol and a conserved value of T_M compared to wild type *Bs*CspB. Focusing on fluorine labelled phenylalanine variants, 4-^{19}F-Phe-*Bs*CspB has a thermodynamic stability which is about 1 kJ/mol lower as the value for $\Delta G°$ which has been determined for 2-^{19}F-Phe-*Bs*CspB and 3-^{19}F-Phe-*Bs*CspB, respectively (Tables 1 and 3).

We have extended the analysis of the kinetic data set to the information included in the amplitude of the monoexponential function following the refolding and unfolding of differently fluorine-labelled variants of *Bs*CspB. Thus, we have observed that the final values of the kinetics reporting on refolding and unfolding of fluorine-labelled variants do not converge as illustrated by 4-^{19}F-Phe-*Bs*CspB (Fig. S20A,B). The refolding kinetics for all fluorine variants probed here show consistently lower amplitudes compared to unfolding kinetics

(Fig. S20C–G). Note that the endpoint analysis performed for wild type *Bs*CspB based on kinetic stopped-flow data shows an unfolding transition which is anticipated (Fig. S20H). Consequently, we have examined the feature of non-converging endpoints seen for refolding and unfolding of fluorine-labelled *Bs*CspB variants by performing a set of control experiments. Firstly, 4-^{19}F-Phe-*Bs*CspB has been mixed with denaturing buffer leading to $c_{\text{urea}} = 2.6\,\text{M}$. Furthermore, unfolded 4-^{19}F-Phe-*Bs*CspB present in $c_{\text{urea}} = 7\,\text{M}$ has been mixed with native buffer leading to $c_{\text{urea}} = 2.6\,\text{M}$, too. The fluorescence emission of both unfolded and refolded 4-^{19}F-Phe-*Bs*CspB has been immediately measured. As a result, the fluorescence intensity representing refolded 4-^{19}F-Phe-*Bs*CspB is lower compared to the fluorescence intensity observed for unfolded 4-^{19}F-Phe-*Bs*CspB (Fig. S21A). Quantitatively, the fluorescence intensity monitoring unfolding of 4-^{19}F-Phe-BsCspB is about 30% higher compared to the fluorescence intensity monitored for protein refolding (Fig. S21A). The higher fluorescence intensity observed for unfolding of 4-^{19}F-Phe-*Bs*CspB compared to refolding matches the difference seen in fluorescence amplitude observing folding kinetics of 4-^{19}F-Phe-*Bs*CspB independently acquired at the stopped-flow instrument (Fig. S20A). Secondly, potential slow unfolding or refolding of 4-^{19}F-Phe-*Bs*CspB taking place on a second-to-minute time scale cannot be observed as the fluorescence emission spectra monitored directly after mixing and after 30 minutes look identical (Fig. S21A). Thirdly, monitoring refolding and unfolding of wild type *Bs*CspB by using identical experimental conditions reveals a different result. Both refolding and unfolding of wild type *Bs*CspB observed in $c_{\text{urea}} = 2.6\,\text{M}$ lead to almost identical profiles in fluorescence emission intensity (Fig. S22). Fourthly, we have applied high-resolution NMR spectroscopy to further unravel the origin of the gap present in amplitude analysis of the kinetic data observed for fluorine-labelled BsCspB variants (Fig. S21B). Thus, the reversibility of folded-to-unfolding reaction of 4-^{19}F-Phe-*Bs*CspB has been quantitatively probed by acquiring one-dimensional ^{1}H NMR data on both unfolded and refolded samples, respectively (Fig. S21B). The fraction of native 4-^{19}F-Phe-*Bs*CspB, f_{N}, can be determined by calculating the ratio $f_{\text{N}} = I_{\text{N}}/(I_{\text{N}} + I_{\text{N+U}})$ using integrals representing native, I_{N}, as native and unfolded, $I_{\text{N+U}}$, signals of *Bs*CspB. The analysis of both the unfolding and the refolding reaction of 4-^{19}F-Phe-*Bs*CspB leads to almost identical values for f_{N}, namely $f_{\text{N}} = 0.191$ and $f_{\text{N}} = 0.189$, respectively. This result shows that 4-^{19}F-Phe-*Bs*CspB folds fully reversibly and that there is no apparent long-term folding rate constant additionally present which may explain the gap observed in the amplitude analysis of the kinetic data. Taken these results together, we conclude that adding singly fluorinated tryptophan or phenylalanines to *Bs*CspB inherently changes the properties in tryptophan fluorescence emission of this protein. This change in fluorescence emission can be observed when comparing fluorine-labelled *Bs*CspB which has been refolded with the unfolded counterpart. Contrary, applying high-resolution NMR spectroscopy does not reveal structural differences in fluorine-labelled *Bs*CspB when comparing refolded with unfolded protein which has been modified by fluorotryptophan or fluorophenylalanines.

Structure determination of 4-^{19}F-Phe-*Bs*CspB and 4-^{19}F-Trp-*Bs*CspB by X-ray crystallography. Probing fluorine-labelled variants of *Bs*CspB using fluorescence and NMR spectroscopy has independently shown that 4-^{19}F-Trp-*Bs*CspB has a slightly increased thermodynamic stability, whereas $\Delta G°$ of 4-^{19}F-Phe-*Bs*CspB is slightly decreased compared to wild type protein. To investigate potential structural changes in 4-^{19}F-Trp-*Bs*CspB and 4-^{19}F-Phe-*Bs*CspB, we have elucidated their atomic structures using X-ray crystallography to 2.05 Å and 2.1 Å resolution, respectively (Fig. 5, Table S2). The samples crystallized in a closely related crystallographic lattice, containing one single molecular copy of modified *Bs*CspB in the asymmetric unit and sharing a same space group symmetry. The global RMSD for all C_{α} atoms for the 4-^{19}F-Trp-*Bs*CspB and 4-^{19}F-Phe-*Bs*CspB structures calculated in this work is 0.23 Å, indicating that the fold trace of the two variants is virtually identical. A comparison of the variants here studied with a previously reported crystal structure of wild type *Bs*CspB elucidated in a different crystallographic symmetry (PDB ID: 1CSP) revealed $\text{RMSD}_{\text{C}\alpha}$ values of 0.52 Å and 0.37 Å for 4-^{19}F-Trp-*Bs*CspB and 4-^{19}F-Phe-*Bs*CspB, respectively (Fig. 5). The excellent agreement across all structures confirmed that ^{19}F derivatization did not induce any detectable alterations of the fold. Furthermore, wild type and ^{19}F-derivative structures displayed identical conformational rotamers for the modified Trp and Phe residues, indicating that local structural distortions had also not taken place (Fig. 5). When comparing wild type and derivatives, the RMSD values for all atoms in the modified residues were: 0.48 Å for residue Trp8, 0.35 Å for Phe9, 0.30 Å for Phe15, 0.15 Å for Phe17, 0.43 Å for Phe27, 0.2 Å for Phe30, 0.65 Å for Phe38, and 0.3 Å for Phe49. Based on this excellent agreement, we conclude that the introduction of the ^{19}F-labels in *Bs*CspB did not induce structural changes of significance even when seven residues were modified.

Conclusion

Synergistic combination of equilibrium and kinetic fluorescence with NMR spectroscopy and X-ray crystallography. We have presented an integrated approach to structurally and dynamically characterize fluorine-labelled proteins using a repertoire of orthogonal biophysical methodologies. Fluorescence experiments conducted in equilibrium following the folded-to-unfolding reaction of fluorine-labelled *Bs*CspB induced by chemical denaturation as well as monitoring the refolding and unfolding of these samples by fluorescence emission in a time-resolved manner elucidated only small differences in overall thermodynamic stability. This conservation in thermodynamic stability regarding wild type behaviour could be confirmed by monitoring the thermal denaturation of fluorine-labelled *Bs*CspB using one-dimensional ^{1}H and ^{19}F NMR spectroscopy. The crystallization and structure determination of fluorine modified *Bs*CspB corroborated the findings presented above by showing an almost perfect overlay between wild type and ^{19}F-Phe or ^{19}F-Trp modified protein variants taking both backbone and side chain atoms into account. To our knowledge, this represents the first three-dimensional structure determination of a fluorophenylalanine-labelled protein possessing more than a single fluorination site which has been reported so far. Note that the solution structure of a fluorinated side chain labelled villin headpiece subdomain comprising 35 residues has been presented before[31]. Thus, the coherent characterization of fluorine-labelled *Bs*CspB done here consistently show only mild effects protein labelling causes on structural and

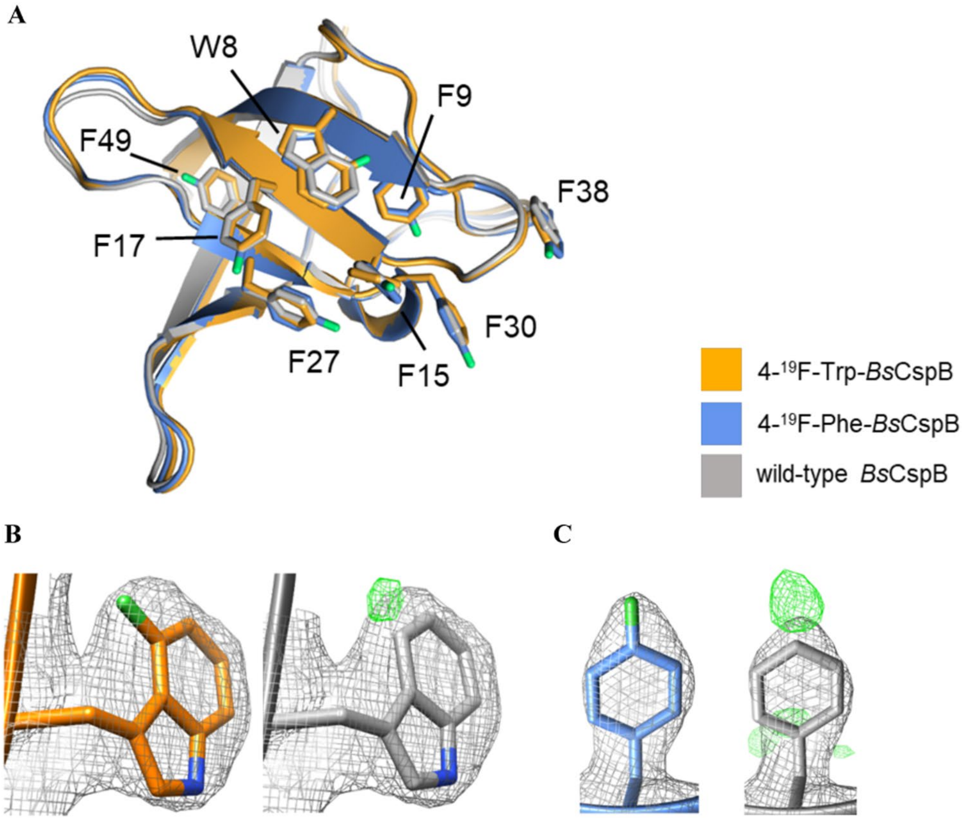

Figure 5. Three-dimensional structures of ^{19}F-derivatized *Bs*CspB. (**A**) The crystal structures of 4-^{19}F-Trp-*Bs*CspB (colored in orange) and 4-^{19}F-Phe-*Bs*CspB (colored in blue) are shown superimposed onto wild type *Bs*CspB (pdb code 1NMG, colored in grey). The side chains of ^{19}F-derivatized residues in this study are displayed and labelled according to their sequence numbering. (**B**,**C**) Electron density maps of residues ^{19}F-Trp8 (B) and ^{19}F-Phe49 (as representative) are shown. $(2mF_{obs} - DF_{calc})\alpha_{calc}$ maps contoured at 1σ are colored in grey. $(mF_{obs} - DF_{calc})$ maps calculated using phases from refined models with omitted ^{19}F-derivatized residues are shown in green color, contoured at 3σ. The structures have been created by using Chimera[59], version 1.14 (www.cgl.ucsf.edu/chimera).

dynamic properties by using singly fluorinated amino acids, even when multiple sites in the protein have been used for modification. The integrated approach for the determination of the impact fluorine-labelling has on proteins converges to the main finding. In fact, incorporation of singly fluorinated Trp or Phe into *Bs*CspB induces, if at all, only slight changes in structural and dynamic parameters. Interestingly, these slight changes are independent of the number and sites of fluorine atoms which have been inserted into *Bs*CspB: one fluorine atom used for ^{19}F-Trp labelling (sequence position 8) or seven fluorine atoms used for ^{19}F-Phe labelling (sequence positions 9, 15, 17, 27, 30, 38, and 49). This modest impact of protein labelling using fluorine has been partially presented for other proteins but only by using either a limited number of experimental techniques or a reduced number of different fluorinated amino acids which have been used for protein labelling. In this regard, a combination of CD and NMR spectroscopy showed that the three-dimensional structure and thermodynamic stability of GB1 protein is not significantly affected when 5-^{19}F-Trp has been used for labelling[47]. Moreover, the structure of 5-^{19}F-Trp- and 6-^{19}F-Trp-labelled annexin V has been determined by X-ray crystallography[48] elucidating only minimal changes in the local protein geometry compared to non-labelled protein and slight changes in thermal melting observed by CD spectroscopy which are on the same order as elucidated here for fluorine-labelled *Bs*CspB. Similarly, the crystal structure of 5-^{19}F-Trp-labelled triosephosphate isomerase showed no discrepancies in local and global structural properties comparing labelled with non-labelled protein[49], the same property has been reported for anthrax protective antigen indicating that 5-^{19}F-Trp-labelling minimally perturb structural properties seen for wild type protein[50] or for fluoroacetate dehalogenase utilizing also 5-^{19}F-Trp-labelling[22]. Specialized studies have focused e.g. on relaxation properties of fluorinated amino acids in free and protein-bound form[51], general NMR parameters of isolated fluorine-labelled amino acids[52] or the application of homonuclear ^{19}F-^{19}F EXSY NMR on a fluorine-labelled receptor protein[53]. We are aware that we have solely probed one protein in this study. However, the experimental design has been done following a highly systematic strategy by incorporating six different fluorinated amino acids into *Bs*CspB. Such fluorine-labelled variants of *Bs*CspB have been subsequently probed applying various orthogonal biophysical techniques which led to a consistent result. For this reason, we believe that the present study closes an important gap in the characterization of fluorine-labelled proteins by obtaining a convergent view on the impact that the insertion of fluorine into proteins has. We propose that single fluorine-labelled tryptophan and phenylalanine residues may serve as ideal candidates for the incorporation into proteins enabling experiments to understand protein performance in highly biological contexts like cell lysates or even *in cellula*.

Methods

Protein expression and purification. The six variants of the cold shock proteins B from *Bacillus subtilis* differing in the position of the fluorine (2-^{19}F-Phe-, 3-^{19}F-Phe-, 4-^{19}F-Phe-, 4-^{19}F-Trp-, 5-^{19}F-Trp-, and 6-^{19}F-Trp-*Bs*CspB, Figs. 1 and S1) were expressed in *E. coli* cells by using pET24a CspB and pAR1219 vectors as described previously[32] (DSMZ 12779 strain was used for the three Phe variants whereas the strain CAG 18455 7371 was used for the three Trp variants). Subsequently, an established protocol for protein purification was applied[32]. The concentration, *c*, of the purified proteins was determined by measuring the absorbance at $\lambda = 280$ nm, A_{280}, of the protein solution in a $d = 1$ cm long cuvette in an UV/Vis spectrometer (Agilent 8453 UV-visible Spectroscopy System, Agilent Technologies) employing extinction coefficients of $\varepsilon^{\circ 280} = 2705\,\mathrm{M^{-1}\,cm^{-1}}$ (4-^{19}F-Trp *Bs*CspB), $\varepsilon^{280} = 2887\,\mathrm{M^{-1}\,cm^{-1}}$ (5-^{19}F-Trp *Bs*CspB), $\varepsilon^{280} = 2575\,\mathrm{M^{-1}\,cm^{-1}}$ (6-^{19}F-Trp *Bs*CspB), and $\varepsilon^{280} = 5800\,\mathrm{M^{-1}\,cm^{-1}}$ (2-^{19}F-Phe *Bs*CspB, 3-^{19}F-Phe *Bs*CspB, 4-^{19}F-Phe *Bs*CspB, and wild type *Bs*CspB)[32]. The Lambert Beer law, $A_{280} = c*d*\varepsilon^{280}$, was then applied to the determination of *c*. Fluorine labelling efficiency can be specified with >95%[32].

Fluorescence spectroscopy in steady state. Fluorescence experiments performed in steady state were conducted on a FP-8500 Spectrofluorometer (Jasco). The final concentration of the protein was set to $c = 1\,\mu$M in all experiments. Samples were thermally equilibrated for at least 30 minutes and measured under stirring condition at a temperature $T = 298$ K, using 20 mM sodiumcacodylate, pH 7.0. The individual folded-to-unfolding transitions of *Bs*CspB variants were monitored between $c = 0$ M and $c = 8$ M urea applying 34 data points. Fluorescence spectra were acquired as triplicates in the wavelength (λ) range 290 nm to 400 nm in steps of 0.5 nm by using a wavelength for excitation of $\lambda = 280$ nm. The concentration of urea, $c_{\text{urea,}}$ was determined by measuring the refractive index (Eq. (2))[54].

$$c_{\text{urea}} = 117.66 * \Delta n + 29.753 * \Delta n^2 + 185.56 * \Delta n^3, \tag{2}$$

where Δn reflects the difference in the refractive index of the buffer solution taking absence and presence of urea into account.

All equilibrium protein folded-to-unfolding transitions were background subtracted and measured in duplicate.

The Eq. (3) was applied to determine the difference in free energy between the folded and the unfolded state of *Bs*CspB, ΔG^0, and the cooperativity of protein unfolding, *m*

$$\langle \lambda \rangle = \frac{(g_{\text{N}} + m_{\text{N}} * c_{\text{urea}}) + (g_{\text{U}} + m_{\text{U}} * c_{\text{urea}}) * \left(\exp\left(-\frac{\Delta G^0}{RT}\right) + \frac{m * c_{\text{urea}}}{RT}\right)}{1 + \left(\exp\left(-\frac{\Delta G^0}{RT}\right) + \frac{m * c_{\text{urea}}}{RT}\right)}, \tag{3}$$

where $\langle \lambda \rangle$ represents the intensity averaged emission wavelength[41], $g_{\text{N/U}}$ and $m_{\text{N/U}}$ account for the baselines of the folded and the unfolded state, *R* is the universal gas constant, c_{urea} is the concentration of urea, and *T* is the absolute temperature[42].

NMR spectroscopy. All one-dimensional ^{1}H NMR spectra were acquired at protein concentrations ranging between $c = 200\,\mu$M and $c = 650\,\mu$M and measured in 20 mM sodiumcacodylate containing 90% H_2O and 10% D_2O (v/v) at pH = 7.0. NMR data were collected on an 800 MHz Bruker Avance NEO NMR spectrometer equipped with a TCI cryogenically cooled probe possessing a proton channel which allows tuning of the ^{19}F resonance at $\omega_{\text{L}}{}^{19\text{F}} = 753$ MHz. The proton resonance frequency of trimethylsilylproanoic acid (TMSP) was used for direct referencing of all ^{1}H spectra. One-dimensional ^{19}F NMR spectra were indirectly referenced by using the information obtained for referencing of protons. The processing of NMR data used TOPSPIN 4.0.3 software (Bruker Biospin, Germany).

The determination of the thermodynamic stability of the different variants of *Bs*CspB was monitored by one-dimensional ^{1}H- and ^{19}F- NMR spectroscopy (holds for 4-^{19}F-Phe-*Bs*CspB) at different temperatures ranging between $T = 291$ K and $T = 330$ K. Two ranges differing in chemical shifts were used for the determination of the fraction of unfolded protein, f_{U}, being present at different temperatures. The first range, I_{N}, covers chemical shifts between 0.59 ppm and 0.14 ppm representing signals seen for the folded state. The second range, $I_{\text{N+U}}$, covers chemical shifts between 0.697 ppm and 1.064 ppm representing signals indicating both the folded and the unfolded state, respectively. Using the ratio $I_{\text{N}}/(I_{\text{N}} + I_{\text{N+U}})$ enables the precise determination of the total population of the folded state, f_{N}, at any temperature[45]. The population of the unfolded state, f_{U}, can now be determined assuming a two-state folding scenario as described for *Bs*CspB, $f_{\text{U}} = 1 - f_{\text{N}}$[33]. The fraction of unfolded protein, f_{U}, has been subsequently used to determine the temperature midpoint, T_{M}, of the folded-to-unfolding transition of different variants of *Bs*CspB by using Eq. (4)

$$f_{\text{U}}(T) = \frac{\exp\left(-\frac{\Delta H_{\text{U}}(T_{\text{M}})\left(\frac{T_{\text{M}} - T}{T_{\text{M}}}\right) - \Delta C_{\text{P}}\left(T_{\text{M}} - T + T * \ln\left(\frac{T}{T_{\text{M}}}\right)\right)}{RT}\right)}{1 + \exp\left(-\frac{\Delta H_{\text{U}}(T_{\text{M}})\left(\frac{T_{\text{M}} - T}{T_{\text{M}}}\right) - \Delta C_{\text{P}}\left(T_{\text{M}} - T + T * \ln\left(\frac{T}{T_{\text{M}}}\right)\right)}{RT}\right)}, \tag{4}$$

where ΔH_{U} represents the van't Hoff enthalpy of unfolding at T_{M}, ΔC_{P} the change in heat capacity between folded und unfolded state, *R* is the universal gas constant, and *T* is the absolute temperature[45].

Kinetic measurements using fluorescence stopped-flow methodology. Kinetic fluorescence spectroscopic data were collected by using a SX20 Stopped Flow Spectrometer (Applied Photophysics, UK). After excitation at a wavelength of 280 nm the folding kinetics was recorded by the change of fluorescence above a wavelength of 320 nm using a cut off filter. All single mixing experiments were performed in 20 mM sodiumcacodylate at pH = 7.0 and $T = 298$ K. The protein solution of $c = 15\,\mu$M present in the native buffer was diluted 11-fold with urea solutions (in 20 mM sodiumcacodylate, pH = 7.0) of different concentrations leading to final concentrations of urea in the measuring cell ranging between $c_{\text{urea}} = 2.6$ M and 7.3 M to detect kinetics of protein unfolding. Refolding kinetics of unfolded protein of $c = 15\,\mu$M present in $c_{\text{urea}} = 7$ M and 20 mM sodiumcacodylate, pH = 7.0 were followed by an 11-fold dilution with urea solutions (in 20 mM sodiumcacodylate, pH = 7.0) at different concentrations, leading to final concentrations of urea in the measuring cell ranging between $c_{\text{urea}} = 0.6$ M and 3.9 M. All folding and refolding kinetics were measured 12 times under identical conditions and averaged. Data processing used the Pro Data Viewer software (Applied Photophysics, UK).

Kinetic data were analysed using a monoexponential function. The apparent rate constant, k_{obs}, has been plotted logarithmically as a function of c_{urea}. Data analysis applying Eq. (5) enabled the determination of the rate constants of protein refolding k_f and unfolding k_u, respectively, assuming a two state folding process

$$\ln(k_{\text{obs}}) = \exp\left(\ln(k_f) + \frac{m_f * c_{\text{urea}}}{RT}\right) + \exp\left(\ln(k_u) + \frac{m_u * c_{\text{urea}}}{RT}\right). \tag{5}$$

Here, m_f and m_u represent the limbs for protein refolding and unfolding leading to the cooperativity of folding $m = m_f + m_u$, which was independently determined before using Eq. (3), R is the universal gas constant, and T is the absolute temperature[33].

Finally, the rate constants for refolding, k_f, and unfolding k_u, can be used to determine the overall thermodynamic stability, ΔG^0, of the protein under investigation. Thus the difference in free energy of a protein sensing folding or unfolding conditions can be obtained by using Eq. (6)

$$\Delta G^0 = -RT * ln\, K = -RT * ln\frac{[U]}{[N]} = -RT * ln\frac{k_u}{k_f}, \tag{6}$$

where K is the equilibrium constant, R is the universal gas constant, and T is the absolute temperature[33].

X-ray crystallography. Protein samples 4-^{19}F-Phe-*Bs*CspB and 4-^{19}F-Trp-*Bs*CspB were concentrated to 20 mg/ml in 20 mM sodium cacodylate pH 7.0 and 100 mM NaCl and crystallized using a Gryphon liquid dispenser (Art Robbins instruments) on 96-well Intelli-plates (Art Robbins instruments) employing the sitting drop method at 18 °C. Crystals of 4-^{19}F-Phe-*Bs*CspB grew from 100 mM CHES pH 9.5, 1 M sodium citrate tribasic. Crystals of 4-^{19}F-Trp-*Bs*CspB grew from 0.1 M BIS-TRIS propane pH 7.0, 1.2 M sodium citrate tribasic dehydrate. Drop consisted of a 150 nl:150 nl protein:precipitate ratio. For X-ray data cryo-collection, crystals were vitrified in LN_2 in mother liquor supplemented with 25% [v/v] glycerol. X-ray diffraction data were collected at the Swiss Light Source synchrotron (Villigen) and processed using XDS/XSCALE[55]. Phasing was by molecular replacement in PHASER[56] using the crystal structure of wild type CspB (PDB ID: 1CSP) as search model. Manual model building was in COOT[57] and refinement used PHENIX.refine[58]. Chemical libraries for the modified residues 4-^{19}F-Phe (PFF) and 4-^{19}F-Trp (4FW) were obtained from the Protein Data Bank. Visual structure comparisons and calculations of root-mean-square deviations (RMSD) were performed in Chimera[59].

Received: 5 December 2019; Accepted: 29 January 2020;
Published online: 14 February 2020

References

1. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philos. T R. Soc. A* **361**, 1205–1222, https://doi.org/10.1098/rsta.2003.1194 (2003).
2. Zimmerman, S. B. & Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia-Coli. *J. Mol. Biol.* **222**, 599–620, https://doi.org/10.1016/0022-2836(91)90499-V (1991).
3. Mika, J. T. & Poolman, B. Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotech.* **22**, 117–126 (2011).
4. Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604, https://doi.org/10.1016/S0968-0004(01)01938-7 (2001).
5. Zhou, H. X., Rivas, G. N. & Minton, A. P. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophysics* **37**, 375–397, https://doi.org/10.1146/annurev.biophys.37.032807.125817 (2008).
6. Gnutt, D. *et al.* Stability effect of quinary interactions reversed by single point mutations. *J. Am. Chem. Soc.* **141**, 4660–4669 (2019).
7. Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nat.* **458**, 106 (2009).
8. Theillet, F.-X. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nat.* **530**, 45 (2016).
9. Tanaka, T. *et al.* High-Resolution Protein 3D Structure Determination in Living Eukaryotic Cells. *Angewandte Chemie International Edition* (2019).
10. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The preparation of 19F-labeled proteins for NMR studies. *Methods enzymology* **380**, 400 (2004).
11. Gerig, J. Fluorine NMR of proteins. *Prog. Nucl. Magnetic Reson. Spectrosc.* **26**, 293–370 (1994).
12. Li, C. G. *et al.* Protein F-19 NMR in Escherichia coli. *J. Am. Chem. Soc.* **132**, 321–327, https://doi.org/10.1021/ja907966n (2010).
13. Sun, X., Dyson, H. J. & Wright, P. E. Fluorotryptophan Incorporation Modulates the Structure and Stability of Transthyretin in a Site-Specific Manner. *Biochemistry-Us* **56**, 5570–5581, https://doi.org/10.1021/acs.biochem.7b00815 (2017).
14. Schuler, B., Kremer, W., Kalbitzer, H. R. & Jaenicke, R. Role of entropy in protein thermostability: Folding kinetics of a hyperthermophilic cold shock protein at high temperatures using F-19 NMR. *Biochemistry-Us* **41**, 11670–11680, https://doi.org/10.1021/bi0262931 (2002).
15. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of F-19-Tryptophan for protein NMR spectroscopy. *Chem. Commun.* **48**, 10681–10683, https://doi.org/10.1039/c2cc35347d (2012).

16. Li, H. L. & Frieden, C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and F-19 NMR. *P Natl Acad. Sci. USA* **104**, 11993–11998, https://doi.org/10.1073/pnas.0705253104 (2007).
17. Hull, W. E. & Sykes, B. D. Fluorotyrosine Alkaline Phosphatase: Internal Mobility of Individual Tyrosines and the Role of Chemical Shift Anisotropy as a ~gFNuclear Spin Relaxation Mechanism in Proteins. *J. Mol. Biol.* **98**, 121–153 (1975).
18. Tobola, F. *et al*. Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *Acs Chem. Biol.* **13**, 2211–2219, https://doi.org/10.1021/acschembio.8b00377 (2018).
19. Acchione, M. *et al*. Specific Fluorine Labeling of the HyHEL10 Antibody Affects Antigen Binding and Dynamics. *Biochemistry-Us* **51**, 6017–6027, https://doi.org/10.1021/bi300455t (2012).
20. Matei, E. & Gronenborn, A. M. 19F paramagnetic relaxation enhancement: a valuable tool for distance measurements in proteins. *Angew. Chem. Int. Ed.* **55**, 150–154 (2016).
21. Zimmermann, K. *et al*. Localization of ligands within human carbonic anhydrase II using 19 F pseudocontact shift analysis. *Chem. Sci.* **10**, 5064–5072 (2019).
22. Kim, T. H. *et al*. The role of dimer asymmetry and protomer dynamics in enzyme catalysis. *Sci.* **355**, eaag2355, https://doi.org/10.1126/science.aag2355 (2017).
23. Divakaran, A., Kirberger, S. E. & Pomerantz, W. C. K. SAR by (Protein-Observed) 19F NMR. *Acc. Chem. Res.* **52**, 3407–3418 (2019).
24. Gee, C. T., Arntson, K. E., Koleski, E. J., Staebell, R. L. & Pomerantz, W. C. K. Dual Labeling of the CBP/p300 KIX domain for 19F NMR leads to identification of a new small molecule binding site. *ChemBioChem* **19**, 963–969 (2018).
25. Li, F. *et al*. A Genetically Encoded 19F NMR Probe for Tyrosine Phosphorylation. *Angew. Chem. Int. Ed.* **52**, 3958–3962 (2013).
26. Yang, F. *et al*. Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and 19F-NMR. *Nature. Commun.* **6**(8202), 1–15 (2015).
27. Ycas, P. D., Wagner, N., Olsen, N. M., Fu, R. & Pomerantz, W. C. K. 2-Fluorotyrosine is a valuable but understudied amino acid for protein-observed 19F NMR. *Journal of Biomolecular NMR*, https://doi.org/10.1007/s10858-019-00290-0 (2019).
28. Buer, B. C., Meagher, J. L., Stuckey, J. A. & Marsh, E. N. G. Comparison of the structures and stabilities of coiled-coil proteins containing hexafluoroleucine and t-butylalanine provides insight into the stabilizing effects of highly fluorinated amino acid side-chains. *Protein Sci.* **21**, 1705–1715, https://doi.org/10.1002/pro.2150 (2012).
29. Buer, B. C., Meagher, J. L., Stuckey, J. A. & Marsh, E. N. G. Structural basis for the enhanced stability of highly fluorinated proteins. *P Natl Acad. Sci. USA* **109**, 4810–4815, https://doi.org/10.1073/pnas.1120112109 (2012).
30. Kitevski-LeBlanc, J. L., Evanics, F. & Prosser, R. S. Optimizing 19F NMR protein spectroscopy by fractional biosynthetic labeling. *J. Biomolecular NMR* **48**, 113–121 (2010).
31. Cornilescu, G. *et al*. Solution structure of a small protein containing a fluorinated side chain in the core. *Protein Sci.* **16**, 14–19 (2007).
32. Welte, H. & Kovermann, M. Targeted expression and purification of fluorine labelled cold shock protein B by using an auxotrophic strategy. *Protein Expr. Purif.* **157**, 86–91, https://doi.org/10.1016/j.pep.2019.02.006 (2019).
33. Schindler, T., Herrler, M., Marahiel, M. A. & Schmid, F. X. Extremely Rapid Protein-Folding in the Absence of Intermediates. *Nat. Struct. Biol.* **2**, 663–673, https://doi.org/10.1038/nsb0895-663 (1995).
34. Schindelin, H., Herrler, M., Willimsky, G., Marahiel, M. A. & Heinemann, U. Overproduction, Crystallization, and Preliminary-X-Ray Diffraction Studies of the Major Cold Shock Protein from Bacillus-Subtilis, Cspb. *Proteins* **14**, 120–124, https://doi.org/10.1002/prot.340140113 (1992).
35. Schnuchel, A. *et al*. Structure in Solution of the Major Cold-Shock Protein from Bacillus-Subtilis. *Nat.* **364**, 169–171, https://doi.org/10.1038/364169a0 (1993).
36. Schindelin, H., Marahiel, M. A. & Heinemann, U. Universal Nucleic Acid-Binding Domain Revealed by Crystal-Structure of the Bacillus-Subtilis Major Cold-Shock Protein. *Nat.* **364**, 164–168, https://doi.org/10.1038/364164a0 (1993).
37. Schindler, T. *et al*. Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from Bacillus subtilis. *Proteins* **30**, 401–406, doi: 10.1002/(Sici)1097-0134(19980301)30:4<401::Aid-Prot7>3.0.Co;2-L (1998).
38. Zeeb, M. *et al*. Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucleic Acids Res.* **34**, 4561–4571 (2006).
39. Sachs, R., Max, K. E., Heinemann, U. & Balbach, J. RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *Rna* **18**, 65–76 (2012).
40. Wong, C.-Y. & Eftink, M. R. Incorporation of tryptophan analogues into staphylococcal nuclease, its V66W mutant, and Δ137− 149 fragment: Spectroscopic studies. *Biochemistry-Us* **37**, 8938–8946 (1998).
41. Royer, C. A., Mann, C. J. & Matthews, C. R. Resolution of the Fluorescence Equilibrium Unfolding Profile of Trp Aporepressor Using Single Tryptophan Mutants. *Protein Sci.* **2**, 1844–1852, https://doi.org/10.1002/pro.5560021106 (1993).
42. Santoro, M. M. & Bolen, D. W. Unfolding Free-Energy Changes Determined by the Linear Extrapolation MethodMethod .1. Unfolding of Phenylmethanesulfonyl Alpha-Chymotrypsin Using Different Denaturants. *Biochemistry-Us* **27**, 8063–8068, https://doi.org/10.1021/bi00421a014 (1988).
43. Köhn, B. & Kovermann, M. Macromolecular crowding tunes protein stability by manipulating solvent accessibility. *ChemBioChem*, https://doi.org/10.1002/cbic.201800679 (2018).
44. Alexander, S. S. & Pace, C. N. Comparison of Denaturation of Bovine Beta-Lactoglobulins-a and B and Goat Beta-Lactoglobulin. *Biochemistry-Us* **10**, 2738-+ (1971).
45. Szyperski, T., Mills, J. L., Perl, D. & Balbach, J. Combined NMR-observation of cold denaturation in supercooled water and heat denaturation enables accurate measurement of Delta C-p of protein unfolding. *Eur. Biophys. J. Biophy* **35**, 363–366, https://doi.org/10.1007/s00249-005-0028-4 (2006).
46. Khan, F., Chuang, J. I., Gianni, S. & Fersht, A. R. The kinetic pathway of folding of barnase. *J. Mol. Biol.* **333**, 169–186 (2003).
47. Campos-Olivas, R., Aziz, R., Helms, G. L., Evans, J. N. & Gronenborn, A. M. Placement of 19F into the center of GB1: effects on structure and stability. *Febs Lett.* **517**, 55–60 (2002).
48. Minks, C., Huber, R., Moroder, L. & Budisa, N. Atomic mutations at the single tryptophan residue of human recombinant annexin V: effects on structure, stability, and activity. *Biochemistry-Us* **38**, 10649–10659 (1999).
49. Rozovsky, S., Jogl, G., Tong, L., McDermott, A. E. & Solution-state, N. M. R. investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *J. Mol. Biol.* **310**, 271–280 (2001).
50. Chadegani, F. *et al*. 19F nuclear magnetic resonance and crystallographic studies of 5-fluorotryptophan-labeled anthrax protective antigen and effects of the receptor on stability. *Biochemistry-Us* **53**, 690–701 (2014).
51. Lu, M., Ishima, R., Polenova, T. & Gronenborn, A. M. 19 F NMR relaxation studies of fluorosubstituted tryptophans. *J Biomol Nmr*, 1–9 (2019).
52. Dürr, U. H. N., Grage, S. L., Witter, R. & Ulrich, A. S. Solid state 19F NMR parameters of fluorine-labeled amino acids. Part I: Aromatic substituents. *J. Magnetic Reson.* **191**, 7–15 (2008).
53. Sušac, L., Eddy, M. T., Didenko, T., Stevens, R. C. & Wüthrich, K. A2A adenosine receptor functional states characterized by 19F-NMR. *Proc. Natl Acad. Sci.* **115**, 12733–12738 (2018).
54. Warren, J. R. & Gordon, J. A. On Refractive Indices of Aqueous Solutions of Urea. *J. Phys. Chem-Us* **70**, 297-&, https://doi.org/10.1021/j100873a507 (1966).
55. Kabsch, W. Xds. *Acta Crystallogr. Sect. D: Biol. Crystallography* **66**, 125–132 (2010).

56. McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M. & Storoni, L. PHASER crystallographic software. *J. Appl. Cryst* (2007).
57. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D: Biol. Crystallography* **66**, 486–501 (2010).
58. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
59. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements
We thank Birgit Köhn for providing samples of wild type *Bs*CspB and Professor Magnus Wolf-Watz for fruitful discussion. Financial support from the Juniorprofessurenprogramm handled by the Baden-Württemberg Stiftung (grant ID: 647/16) is gratefully acknowledged.

Author contributions
M.K. designed the study. H.W., T.Z., and X.M. performed experiments and collected data. H.W., T.Z., O.M., and M.K. analyzed data. H.W., and M.K. wrote the manuscript. All authors have read and critically revised drafts for intellectual contents and provided approval for publication.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-59446-w.

Correspondence and requests for materials should be addressed to M.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.