Structural and thermodynamic analyses of the β-to-α transformation in RfaH reveal principles of fold-switching proteins

Philipp K. Zuber[1]†, Tina Daviter[2]‡, Ramona Heißmann[1], Ulrike Persau[1], Kristian Schweimer[1], Stefan H. Knauer[1]*

[1] Biochemistry IV – Biophysical Chemistry, University of Bayreuth, Universitaetsstrasse 30, 95447 Bayreuth, Germany.

[2] Birkbeck, University of London, Malet Street, Bloomsburry, London WC1E 7HX, United Kingdom.

† Present address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, United Kingdom

‡Present address: The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, United Kingdom.

* For correspondence: stefan.knauer@uni-bayreuth.de

Abstract

The two-domain protein RfaH, a paralog of the universally conserved NusG/Spt5 transcription factors, is regulated by autoinhibition coupled to the reversible conformational switch of its 60-residue C-terminal KOW domain between an α-hairpin and a β-barrel. In contrast, NusG/Spt5-KOW domains only occur in the β-barrel state. To understand the principles underlying the drastic fold switch in RfaH, we elucidated the thermodynamic stability and the structural dynamics of two RfaH- and four NusG/Spt5-KOW domains by combining biophysical and structural biology methods. We find that the RfaH-KOW β-barrel is thermodynamically less stable than that of most NusG/Spt5-KOWs and we show that it is in equilibrium with a globally unfolded species, which, strikingly, contains two helical regions that prime the transition towards the α-hairpin. Our results suggest that transiently structured elements in the unfolded conformation might drive the global folding transition in metamorphic proteins in general.

Introduction

Fundamental understanding of how proteins fold has ever been one of the most important questions in structural biology and it is still not answered, despite recent progress in protein structure prediction (Jumper et al., 2021; Tunyasuvunakool et al., 2021). Since the formulation of the "thermodynamic hypothesis of protein folding" by Anfinsen (Epstein et al., 1963) it has been generally accepted that the amino acid sequence of a protein determines its three-dimensional structure and that a protein adopts only a single folded conformation, which is referred to as physiological state and which corresponds to its global energy minimum. This conformation, in turn, fulfills one distinct function. While this "one sequence – one structure – one function" dogma holds true for most well-folded (globular) proteins, it has been challenged by several discoveries over the past decades. Among those are, for instance, (i) moonlighting proteins, which fulfill two completely unrelated functions (Jeffery, 2014, 1999), (ii) intrinsically disordered proteins (IDPs), which do not adopt a defined secondary or tertiary structure at all, but sample an ensemble of sterically allowed conformations instead (van der Lee et al., 2014), and (iii), most strikingly, metamorphic proteins (also referred to as fold-switching proteins), which can reversibly interconvert between at least two well-defined conformations, sometimes in response to a molecular signal (Murzin, 2008).

The free energy landscape of globular, well-folded proteins is often portrayed as a rugged funnel, with the "rim" corresponding to the multitude of random-coil structures of the "unfolded state" (U state) and the deepest point (global minimum in Gibb's free energy, G), representing the "native" or "physiological" state (N state). IDPs, in contrast, exhibit a rather flat energy landscape and no specific conformation is favored, i.e. significantly populated. Fold-switching proteins are thought to reside in-between these two scenarios, that is their energy landscape may be funnel-like, but it

55 shows at least two major minima, each representing a distinct, well-folded conformation. The
56 various conformations of a fold-switching protein may differ in the following aspects: (i) the type
57 of secondary structure (α-helices, β-strands, turns, random coil), (ii) the extent of secondary
58 structure elements, and (iii) the tertiary structure, usually in combination with (i) and/or (ii)
59 (Dishman and Volkman, 2018; Kim and Porter, 2021). Additionally, these states often exhibit
60 different quaternary structures, e.g. monomeric in one state *vs*. multimeric in another state.

61 A particularly intriguing example of fold-switching proteins is the transcription factor RfaH from
62 *E. coli* (*Ec*RfaH), a member of the universally conserved family of NusG (bacteria) / Spt5 (archaea
63 and eukaryotes) proteins (Werner, 2012). NusG/Spt5 proteins exhibit a modular structure with
64 several domains. Bacterial NusG consists of at least an N-terminal domain and a C-terminal
65 Kyrpides, Ouzounis, Woese (KOW) domain connected by a flexible linker (Werner, 2012). Spt5
66 proteins contain a NusG-like N-terminal (NGN) domain and one (archaea) or several (eukaryotes)
67 KOW domains (Werner, 2012). All structurally characterized NusG/Spt5-KOW domains adopt a
68 five-stranded β-barrel structure (Figure 1A; see e.g. (Klein et al., 2011; Meyer et al., 2015; Mooney
69 et al., 2009; Zuber et al., 2018)).

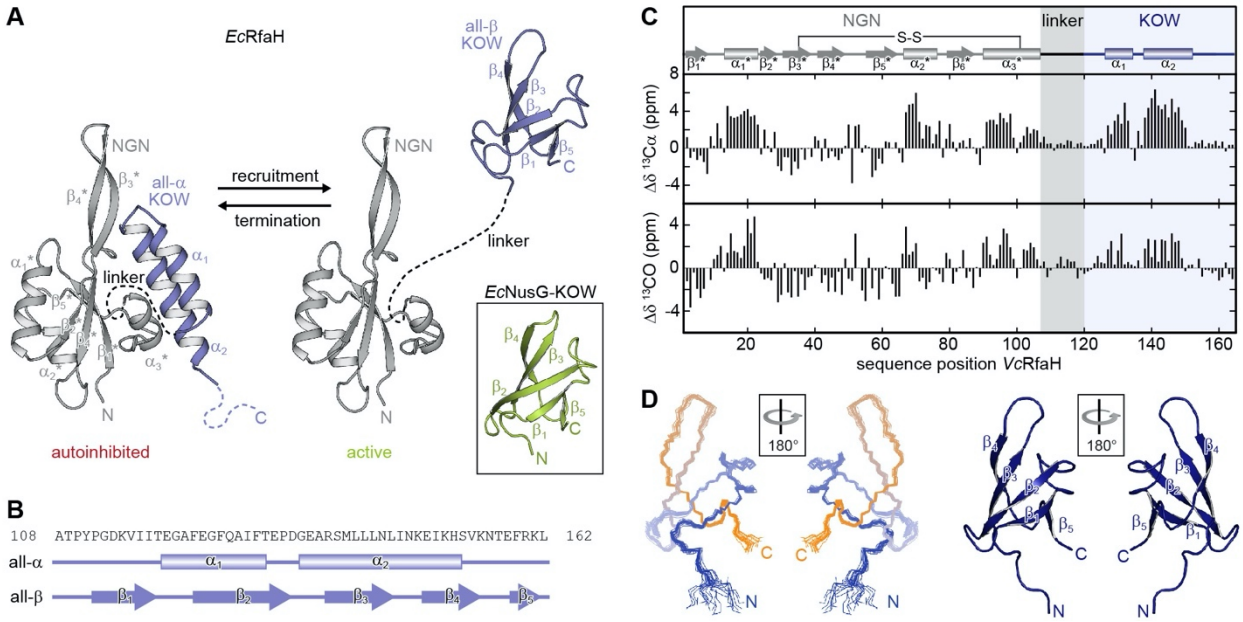

Figure 1. Fold-switching within the NusG/RfaH family. (**A**) Cartoon representation of *Ec*RfaH in the closed, autoinhibited state (left; protein data bank identifier (PDB-ID): 5OND) and in the open, active conformation (right; PDB-ID all-β *Ec*RfaH-KOW: 2LCL) as well as of *Ec*NusG-KOW (boxed; PDB-ID: 2JVV). Unstructured regions are shown as dashed lines, termini are labeled. (**B**) Secondary structures of *Ec*RfaH-KOW in the all-α and the all-β state. Tubes indicate α-helical elements, arrows represent β-strands. The amino acid sequence is shown above. (**C**) Secondary chemical shift of *Vc*RfaH. The plots show the difference between the observed chemical shift and the corresponding predicted random coil value of ^{13}Cα (top) and ^{13}CO (bottom). Positive values indicate helical, negative values elongated (β-sheet) structures, and values close to zero are observed for random coil-like structures. The secondary structure elements inferred from the analysis are shown above the graphs (code for secondary structure elements as in (B)). The position of the identified disulfide bridge (see also **Figure 1 – Figure supplement 1A, B)** is indicated. (**D**) Left: Ribbon representation of the 20 lowest energy structures of *Vc*RfaH-KOW (PDB-ID: 6TF4). Right: Cartoon representation of the lowest energy structure. β-strands and termini are labeled.
Figure supplement 1. Disulfide bridge formation in *Vc*RfaH.
Figure supplement 2. Structure comparison of KOW domains used in this study.

*Ec*RfaH is an operon-specific paralog of *E. coli* NusG (*Ec*NusG) and – just like *Ec*NusG – consists of an NGN domain that is loosely connected with a KOW domain *via* a flexible 15 amino acid long linker. However, in free *Ec*RfaH *Ec*RfaH-KOW folds as an α-helical hairpin (all-α state) that interacts with the *Ec*RfaH-NGN domain. Thus, the binding site for RNA polymerase (RNAP) at the domain interface on *Ec*RfaH-NGN is masked and *Ec*RfaH is locked in an autoinhibited state

93 (Belogurov et al., 2007). Upon recruitment to a transcription elongation complex pausing at an
94 *operon polarity suppressor* (*ops*) site, *Ec*RfaH is activated (Artsimovitch and Landick, 2002;
95 Zuber et al., 2019): the domains dissociate and the liberated *Ec*RfaH-KOW refolds into a NusG-
96 KOW-like β-barrel (all-β state; Figure 1A, B) (Burmann et al., 2012; Zuber et al., 2019).

97 The refolding occurs spontaneously as soon as the domains are separated and *Ec*RfaH-KOW, when
98 produced as an isolated domain, also adopts the all-β state, implying that the all-α fold is only
99 stable in the presence of *Ec*RfaH-NGN (Burmann et al., 2012; Tomar et al., 2013). Each of the
100 *Ec*RfaH-KOW states has a specific function: the all-α state prevents off-target recruitment of
101 *Ec*RfaH and competition with the general transcription factor NusG (Belogurov et al., 2007),
102 whereas the all-β *Ec*RfaH-KOW serves as recruitment platform for ribosomes to activate
103 translation (Burmann et al., 2012; Zuber et al., 2019). Upon release from RNAP *Ec*RfaH is
104 transformed back into its autoinhibited state, i.e. the structural switch of *Ec*RfaH-KOW is fully
105 reversible (Zuber et al., 2019). *Ec*RfaH was not only considered a fold-switching protein, but
106 termed a "transformer protein" to emphasize, that a complete domain cycles reversibly between
107 two states with radically different stable secondary/tertiary structure and with each state
108 performing a distinct function (Knauer et al., 2012).

109 The fine-tuned mechanism used by *Ec*RfaH to control its functions may be widespread in nature
110 (Porter and Looger, 2018) and a recent study predicts that 24 % of the bacterial NusG family
111 members might exhibit similar reversible α-to-β transitions (Porter et al., 2022). However, the
112 molecular principles underlying the fold-switching process are only poorly understood. Here, we
113 present a comprehensive thermodynamic and structural analysis of six KOW domains from
114 NusG/Spt5/RfaH proteins from all domains of life. We combine circular dichroism (CD)
115 spectroscopy, differential scanning calorimetry (DSC), and solution-state nuclear magnetic

116 resonance (NMR) spectroscopy to gain insight into the mechanism and the dynamics of fold-
117 switching within the RfaH family on a molecular level and provide a rationale for the mechanism
118 of fold-switching proteins in general.
119

Results

Evolutionary conservation of fold-switching within the RfaH family

To date, only for *Ec*RfaH three-dimensional structures and comprehensive evidence for fold-switching are available (Belogurov et al., 2007; Burmann et al., 2012; Zuber et al., 2019), although other RfaH orthologs seem to employ a similar mechanism to carry out their function (Carter et al., 2004; Porter et al., 2022). Thus, we first asked whether this ability might be a general feature of RfaH proteins. We chose RfaH from *Vibrio cholerae* (*Vc*RfaH) for a structural analysis by solution-state NMR spectroscopy as it is evolutionarily remote from *Ec*RfaH (sequence identity *Ec*/*Vc*RfaH: 43.6 % (full-length) or 35.8 % (KOW domain), respectively). We first identified the secondary structure elements of the full-length protein by performing an NMR backbone assignment and calculating the secondary chemical shift for each ^{13}Cα and ^{13}CO atom, which depends on the main chain geometry (Figure 1C). In full-length *Vc*RfaH, the KOW domain exhibits two stretches with helical structure that are separated by about four residues and the overall pattern of secondary structure elements perfectly matches the one of autoinhibited *Ec*RfaH (Burmann et al., 2012), suggesting similar tertiary structures for *Ec*RfaH and *Vc*RfaH (compare to Figure 1A), but with helix α_3* being 1.5 turns longer in *Vc*RfaH. Interestingly, the Cα and Cβ atoms of C34 and C102 exhibit chemical shifts typical for cystines ((Sharma and Rajarathnam, 2000), Figure 1 – figure supplement 1A). These residues are located at the end of helix α_3* and in strand β_3*, respectively, and are, most probably, in close proximity, as indicated by the structure of *Ec*RfaH. The addition of a reducing agent to [^{2}H, ^{15}N, ^{13}C]-*Vc*RfaH led to drastic changes of the chemical shifts of C34 and C102 as well as residues in spatial proximity in a [^{1}H, ^{15}N]-heteronuclear single quantum coherence (HSQC) spectrum (Figure 1 – Figure supplement 1B). From this we conclude that C34 and C102 form a disulfide bridge, that covalently tethers the α_3*-

143 helix to the core of *Vc*RfaH-NGN, a feature absent in *Ec*RfaH. However, upon refolding from a
144 solution containing 8 M urea and reducing agent, ^{15}N-*Vc*RfaH adopted the same conformation as
145 before denaturation (Figure 1 – Figure supplement 1C), suggesting that the disulfide bridge is not
146 required for *Vc*RfaH to fold into the autoinhibited state.

147 Next, we determined the solution structure of the isolated *Vc*RfaH-KOW domain by NMR
148 spectroscopy. *Vc*RfaH-KOW also shows the 5-stranded β-barrel topology typical for NusG/Spt5-
149 KOW domains (Figure 1D and Table 1), with a Cα root mean square deviation (rmsd) of 1.4 Å as
150 compared to isolated *Ec*RfaH-KOW.

151 Although we do not present functional data on *Vc*RfaH here, these results strongly suggest that
152 *Vc*RfaH-KOW can also switch between an all-α and an all-β state and that *Vc*RfaH is thus, most
153 probably, also a transformer protein.

154

155 **The model systems**

156 The sequence of NusG/Spt5-KOW domains has been evolutionarily optimized to fold in only one
157 defined conformation. Consequently, in the case of RfaH-KOW, the ability to switch between the
158 all-α and the all-β state must be encoded within the primary structure, whereas the "decision"
159 which state to adopt solely depends on the availability of RfaH-NGN (Tomar et al., 2013).
160 Sequence alignments and bioinformatical approaches (Balasco et al., 2015; Bernhardt and
161 Hansmann, 2018; Gc et al., 2014; Joseph et al., 2019; Li et al., 2014; Shi et al., 2017; Xiong and
162 Liu, 2015) gave first hints why RfaH, in contrast to NusG, is a metamorphic protein and how the
163 structural switch might proceed. Yet, experimental evidence is still scarce. Thus, we analyzed
164 isolated KOW domains of six NusG/Spt5 or RfaH proteins to identify characteristic properties of
165 fold-switching proteins and to understand the molecular mechanisms underlying the refolding

166 mechanism of RfaH-KOW. Due to the fact that NusG proteins are universally conserved, we chose
167 NusG-KOWs from *E. coli* and *Mycobacterium tuberculosis* (*Ec*/*Mt*NusG-KOW), the Spt5-KOW
168 from the hyperthermophilic archaeon *Methanocaldococcus jannaschii* (*Mj*Spt5-KOW) and the
169 fifth KOW domain from human Spt5 (hSpt5-KOW5) as representative NusG-/Spt5-KOWs and
170 the *Ec*/*Vc*RfaH-KOWs as representatives for RfaH proteins. The constructs used are about 65
171 residues in length and contain the structured region and parts of the neighboring linker(s) (Figure
172 1 – Figure supplement 2A). All six domains exhibit the typical β-barrel topology (Figure 1 – Figure
173 supplement 2B) with major differences only in the loops or turns connecting the β-strands (Figure
174 1 – Figure supplement 2C).
175

176 **Thermal and chemical stability of the KOW domains**

177 Metamorphic proteins that switch between two stable conformations are expected to show two
178 main minima in their energy landscape, each corresponding to one of these states (Dishman and
179 Volkman, 2018). This implicates that (i) in order to control the structural interconversion, one of
180 the conformations has to be (de)stabilized according to a molecular signal, and (ii) the energy
181 minima cannot be as deep as the global minimum of a protein with a single, stable conformation
182 to avoid permanent trapping of one state. Consequently, the all-β RfaH-KOW should show a
183 limited thermodynamic stability to allow facile refolding to the all-α state when RfaH-NGN is
184 available after transcription termination. To test this hypothesis, we analyzed the thermal stability
185 of the six KOW domains by CD-based thermal denaturation experiments (Figure 2A) and by DSC
186 (Figure 2B) at pH 4 and pH 7. At pH 7 unfolding was reversible for all KOW domains except for
187 hSpt5-KOW5, which showed aggregation; the opposite effect was observed at pH 4 (Figure 2 –
188 Figure supplement 1). All observed unfolding transitions were analyzed with a two-state model to

determine the melting temperature, T_m, the enthalpy of unfolding at T_m, $\Delta H_u(T_m)$, and, in case of the DSC thermograms, the temperature-dependent difference in heat capacity between the N and U states, $\Delta C_p(T)$ (Figure 2C,D and Table 2).

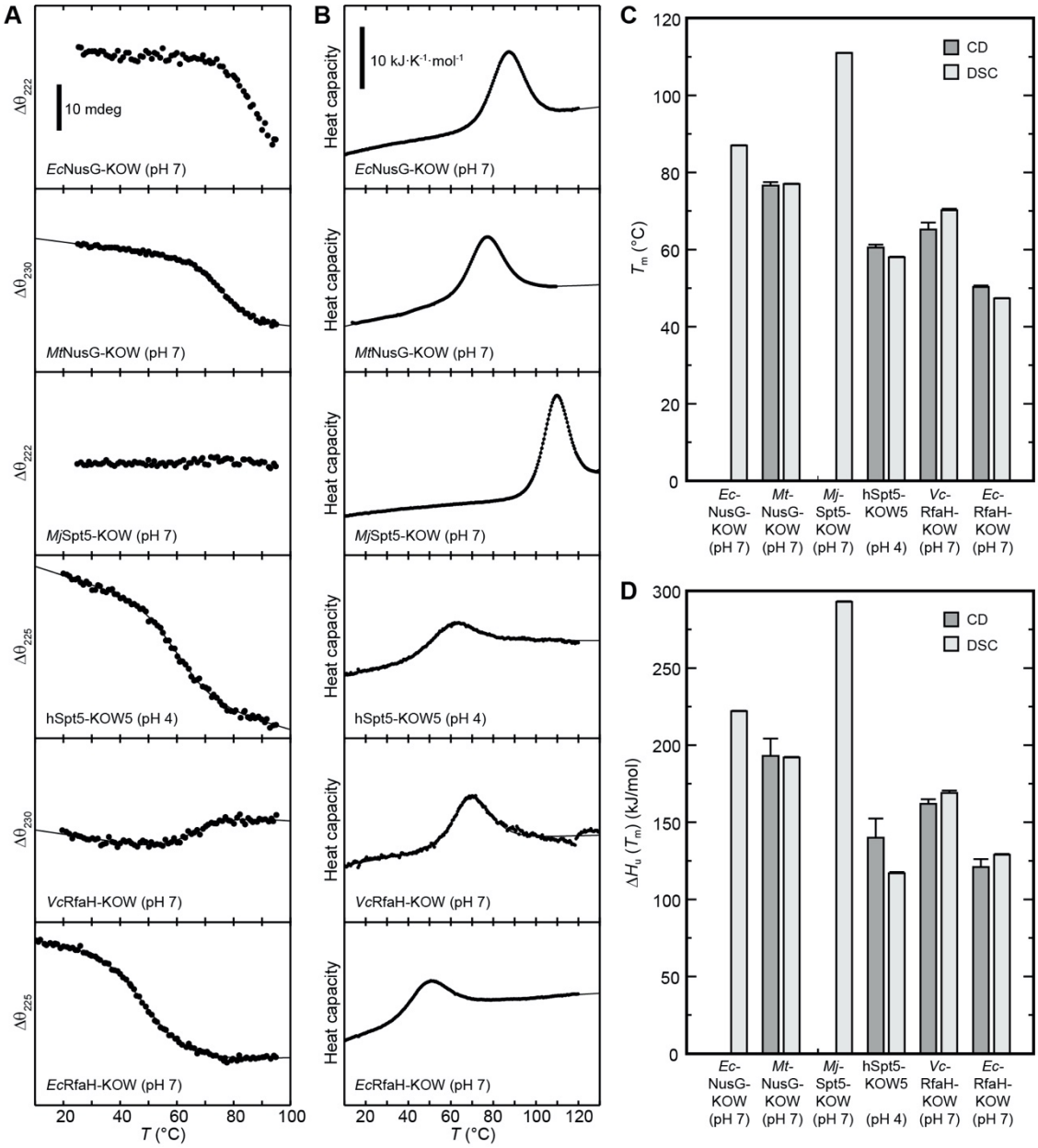

Figure 2. Thermal unfolding experiments of the six KOW domains. (**A**) Thermal unfolding monitored *via* change in the CD signal with a temperature gradient from 20 to 95 °C. The line corresponds to the best fit to a two state-unfolding model. Measurements were carried out with proteins in 10 mM K-acetate (pH 4.0) buffer for hSpt5-KOW5 and in 10 mM K-phosphate

(pH 7.0) buffer for all other domains. The wavelength for monitoring the transition was chosen based on the largest difference between the spectra of the folded and unfolded protein (for details see Materials & Methods). Data for *Ec*NusG-KOW was not fitted due to the lack of the baseline of the unfolded state. *Mj*Spt5-KOW could not be denatured at all. (**B**) Thermograms obtained from DSC measurements. All profiles are normalized to one molar of protein. The lines correspond to best fits to a two-state unfolding model that includes a T-dependent ΔC_p change. Buffers are as in (A). (**C,D**) T_m (**C**) and $\Delta H_u(T_m)$ (**D**) values derived from thermal unfolding experiments monitored by CD and DSC.
Source data 1. Data for thermal denaturation experiments for all KOW domains.
Figure supplement 1. Reversibility of thermal unfolding.

Due to the fact that the KOW domains are β-barrels the precision of the thermodynamic parameters determined by CD spectroscopy is not as high as for proteins with helical elements. Nevertheless, the results obtained by DSC and CD spectroscopy are in good agreement showing that *Ec*NusG-KOW, *Mt*NusG-KOW, and *Mj*Spt5-KOW have much higher T_m values (87 °C, 77 °C and 111 °C, respectively) than hSpt5-KOW5 (58-60 °C), *Ec*RfaH-KOW (47-50 °C), and *Vc*RfaH-KOW (65-70 °C). The same trend was observed for $\Delta H_u(T_m)$ values. Consequently, this data indicates that *Ec*NusG-KOW, *Mt*NusG-KOW, and *Mj*Spt5-KOW have a higher thermodynamic stability than Spt5-KOW5, *Ec*RfaH-KOW, and *Vc*RfaH-KOW.

To corroborate and complement the previous findings, we next performed far-UV CD-based chemical unfolding experiments at pH 4 and pH 7 using urea as denaturant (Figure 3A – F, left).

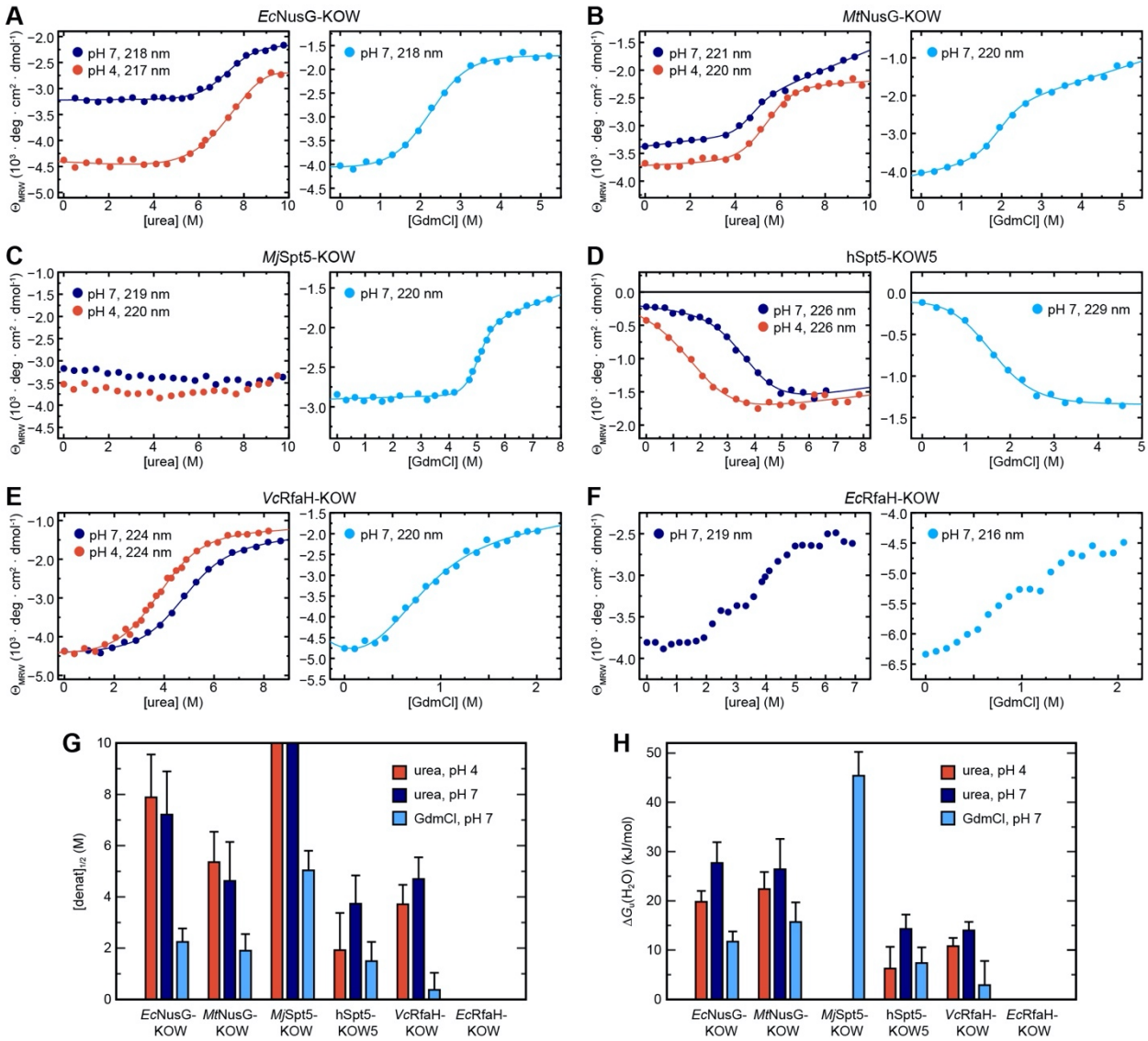

Figure 3. CD spectroscopy-based chemical equilibrium unfolding of the six KOW domains. (**A**) – (**F**): Change in Θ_{MRW} of the indicated protein domain upon over-night incubation with increasing concentrations of (left) urea in 10 mM K-acetate (pH 4.0; red circles) or 10 mM K-phosphate (pH 7.0; dark blue circles), respectively, and (right) GdmCl in 10 mM K-phosphate (pH 7.0; light blue circles). The detection wavelength is indicated and chosen based on the maximum difference between the spectra of the folded and unfolded state (for details see Materials and Methods). The lines correspond to the best fits to a two-state unfolding model, except for *Ec*RfaH-KOW, which exhibits a three-state unfolding behavior. (**G, H**) Comparison of $[\text{denat.}]_{1/2}$ values **(G)** and $\Delta G_u(H_2O)$ values **(H)** of the KOW domains derived from the chemical denaturation experiments shown in (A)-(E).

Source data 1. Data for chemical denaturation experiments for all KOW domains.

Figure supplement 1. Chemical unfolding of *Vc*RfaH-KOW monitored by change in Trp fluorescence.

Figure supplement 2. Reversibility of chemical denaturations for all KOW domains.

236
237 *Ec*NusG-KOW, *Mt*NusG-KOW, hSpt5-KOW5, and *Vc*RfaH-KOW show a sigmoidal unfolding
238 curve at either pH, indicative of a two-state unfolding process. Analysis of this data by the linear
239 extrapolation model yields transition midpoints ($[urea]_{1/2}$ values) and $\Delta G_u(H_2O)$ values that
240 confirm the relative order of the stability as determined by thermal denaturation (Figure 3G, H,
241 Table 3, and Figure 2C,D). For *Mj*Spt5-KOW only the native state baseline is observable at both
242 pH values, demonstrating that no denaturation could be achieved and that, consequently, this KOW
243 domain exhibits the highest thermodynamic stability (assuming an *m*-value comparable to that of
244 the other KOW domains, MjSpt5-KOW likely has a $\Delta G_u(H_2O)$ value > 30-40 kJ/mol). Notably,
245 we obtained a $\Delta G_u(H_2O)$ value for hSpt5-KOW5 at pH 7, showing that this domain has a stability
246 comparable to that of *Vc*RfaH-KOW at physiological pH (Table 3). As *Vc*RfaH-KOW, in contrast
247 to all other KOW domains in this study, contains a Trp residue an additional fluorescence-based
248 denaturation experiment was performed, and the obtained parameters are in good agreement with
249 the CD data (Table 3 and Figure 3 – Figure supplement 1A).
250 To complement the analysis, we repeated the unfolding experiments at pH 7 using guanidinium
251 chloride (GdmCl; Figure 3A – F, right, Table 3 and Figure 3 – Figure supplement 1B). As GdmCl
252 is a more potent denaturant than urea, we were now able to denature even *Mj*Spt5-KOW, giving a
253 $[GdmCl]_{1/2}$ value of 5.03 M, which is more than twice the value of the next stable protein. In
254 accordance with the urea-based unfolding experiments at pH 7, *Mj*Spt5-KOW, *Ec*NusG-KOW and
255 *Mt*NusG-KOW exhibit higher $\Delta G_u(H_2O)$ and $[denat.]_{1/2}$ values than *Vc*RfaH-KOW and hSpt5-
256 KOW5, although the relative order of stability of *Mt*NusG-KOW and *Ec*NusG-KOW is swapped.
257 This difference as well as the difference between the absolute $\Delta G_u(H_2O)$ values derived from the
258 urea- and GdmCl-based denaturations is a well-documented phenomenon and may be attributed
259 to the limited applicability of the linear extrapolation model for the analysis of denaturations by

260 GdmCl (see e.g. (Gupta et al., 1996; Makhatadze, 1999)). Thus, we base our conclusions on the
261 relative comparison of the obtained values. We finally note that chemical unfolding was
262 completely reversible in all cases (Figure 3 – Figure supplement 2).

263 Surprisingly, and in contrast to all other domains, *Ec*RfaH-KOW shows a more complex unfolding
264 curve in both urea- and GdmCl-based denaturation experiments at pH 7, with an additional plateau
265 at ≈3 M urea or ≈1 M GdmCl, respectively, between the N and U baselines (Figure 3F; no curve
266 could be obtained at pH 4 due to native state aggregation). This suggests that the unfolding of
267 *Ec*RfaH-KOW may be described *via* a three-step model including an observable equilibrium
268 intermediate that might play an important role in the fold-switching mechanism of *Ec*RfaH-KOW.

269 In summary, the poor spectroscopic properties of the analyzed domains limit the precision of the
270 absolute values of the thermodynamic parameters obtained from CD experiments. However, our
271 findings reveal clear differences in the global stability of the six domains and allow a grouping
272 into two classes: *Mj*Spt5-KOW and *Ec*/*Mt*NusG-KOW are considered as "stable domains",
273 whereas the β-barrel *Ec*/*Vc*RfaH-KOW as well as hSpt5-KOW5 show a reduced thermodynamic
274 stability.

275

276 **Regions that are unfolded in all-α RfaH-KOW are destabilized in the all-β conformation**

277 We next asked whether the less stable KOW domains also exhibit local differences in their stability
278 as compared to the NusG-KOWs and *Mj*Spt5-KOW. Therefore, we identified the backbone H-
279 bond pattern in the six domains and quantified the magnitude of the through H-bond coupling
280 constant, $^{h3}J_{NC'}$, by long-range HNCO NMR experiments (Table 4). This parameter is inversely
281 proportional to the length of the H-bond and the deviation from its optimum angle, thus reflecting
282 the H-bond strength (Grzesiek et al., 2004). To allow comparison between the six domains, we

grouped H-bonds that are located at equivalent positions of the β-barrels and ordered them according to their position in the individual β-sheets (Figure 4A, B).

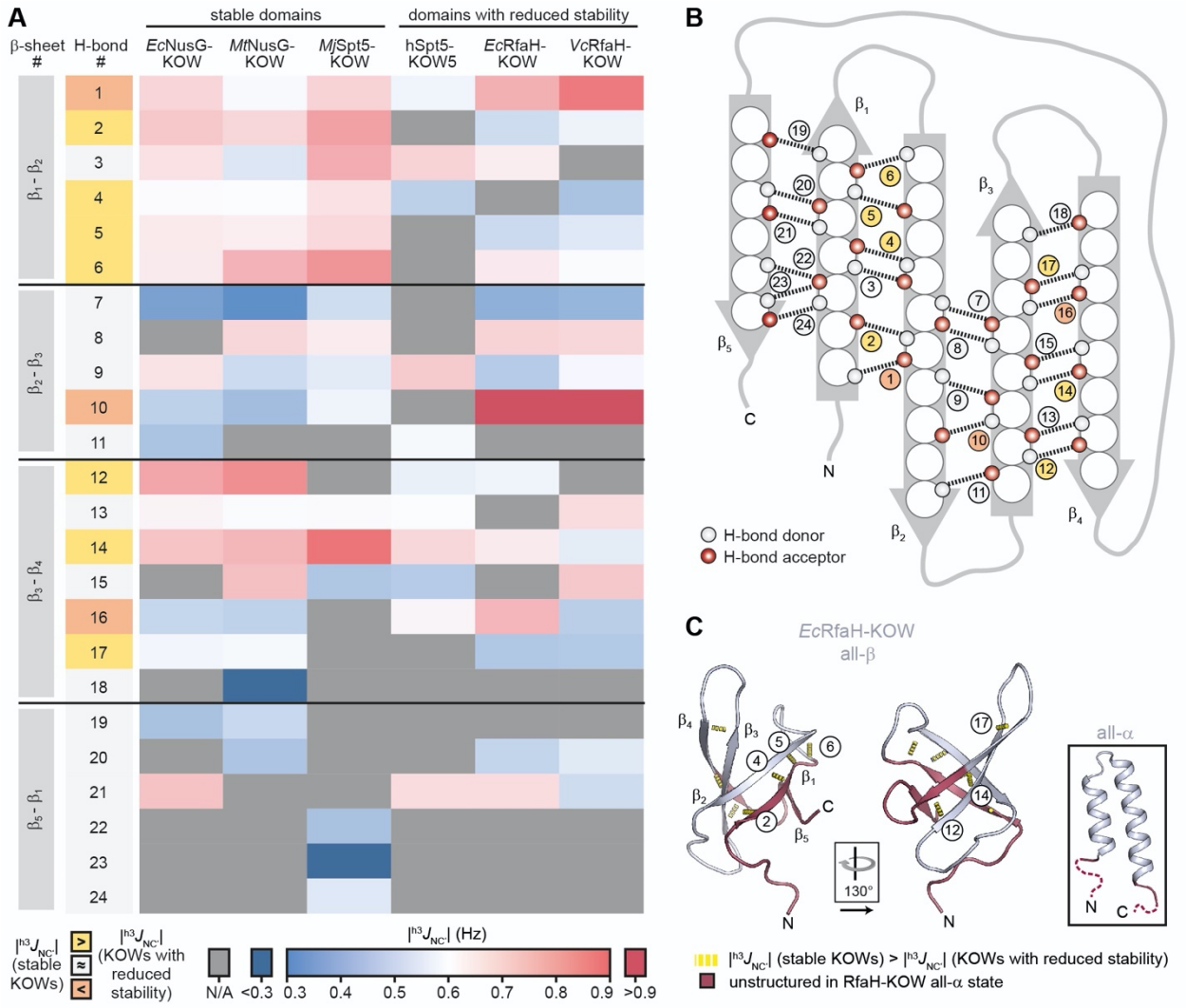

Figure 4. H-bond pattern and stability in the six KOW domains. (**A**) Heat map of the magnitude of the $^{h3}J_{NC'}$ coupling constants of the H-bonds determined by long-range HNCO NMR experiments. H-bonds that are located at equivalent positions are grouped and ordered according to their location in the respective β-sheet (position within the β-barrel as indicated in (B)), and colored according to their $|^{h3}J_{NC'}|$ value as indicated at the bottom. H-bond numbers highlighted in yellow: H-bonds that have lower $|^{h3}J_{NC'}|$ values for at least two of the domains with reduced thermodynamic stability compared to the stable domains; H-bond numbers highlighted in orange: H-bonds that have higher $|^{h3}J_{NC'}|$ values for at least two of the domains with reduced thermodynamic stability compared to the stable domains. (**B**) Scheme of the positions of the H-bonds (dashed lines) within the β-barrel. Amino acids are depicted as spheres. White and red circles represent H-bond donors and acceptors, respectively. H-bonds are color-coded as in (A). (**C**) Cartoon representation of all-β *Ec*RfaH-KOW (PDB-ID: 2LCL, gray). Regions that are

298 unstructured in the all-α conformation are colored in dark red. H-bonds that have lower $|^{h3}J_{NC'}|$
299 values for at least two of the domains with reduced thermodynamic stability compared to the stable
300 domains are shown as yellow dashed tubes and labeled. The relative orientation of the structures
301 is indicated. The inset shows the all-α *Ec*RfaH-KOW (PDB-ID: 5OND; gray; unstructured regions
302 at the termini are colored in dark red and correspond to the dark red regions in the all-β *Ec*RfaH-
303 KOW).
304

305 Most $|^{h3}J_{NC'}|$-values are in the range of 0.5 – 0.9 Hz, which is typical for H-bonds of β-sheets
306 (Grzesiek et al., 2004). In line with having the highest T_m, *Mj*Spt5-KOW often exhibits the highest
307 coupling constants, which is indicative of a highly rigid packing of the β-barrel. Strikingly,
308 *Mj*Spt5-KOW has three additional H-bonds between strands β_5 and β_1 (# 22-24), which provides
309 an extra stabilization of the C-terminal β-strand that may contribute to the high thermostability of
310 this protein. The "stable" domains (i.e. *Ec*/*Mt*NusG-KOW and *Mj*Spt5-KOW) show their strongest
311 H-bonds in two regions, namely between strands β_1:β_2 and β_3:β_4. In addition, most of these H-
312 bonds are more stable than corresponding H-bonds in *Ec*/*Vc*RfaH-KOW and hSpt5-KOW5,
313 implying that the H-bonds in the domains with reduced stability are more dynamic and on average
314 longer or involve a less optimal bonding angle. From this we conclude that in *Ec*/*Vc*RfaH-KOW
315 and hSpt5-KOW5 strands β_1 and parts of β_4 are less stably bound to the rest of the β-barrel than
316 in the stable domains. Moreover, together with the fact that β_1, the C-terminal half of β_4, and β_5
317 are disordered in the all-α state of the *Ec*/*Vc*RfaH-KOW (Figure 4C), this also reflects the
318 chameleonic folding behavior of these regions in the all-β state.
319

320 **hSpt5-KOW5, *Ec*- and *Vc*RfaH-KOW exchange with a globally unfolded conformer on the**
321 **ms-time scale**

322 To assess the folding mechanism of the KOW domains at the amino acid level, we performed an
323 NMR-based analysis of the structural dynamics of the six β-barrel proteins. As larger structural

324 rearrangements, such as folding events, mostly occur at the µs-ms time scale for small proteins or
325 are even slower (Maxwell et al., 2005), we focused on the analysis of the slow chemical exchange
326 regime. Therefore, we performed amide ^{15}N-based chemical exchange saturation transfer (CEST)
327 experiments (Vallurupalli et al., 2012). This method allows the sensitive detection and
328 characterization of sparsely populated states (= minor species; relative population p_B) that
329 exchange with a major species (relative population $p_A = 1 - p_B$) with a rate k_{ex} of 10 – 200 s^{-1}. The
330 detection is achieved by frequency-selective saturation along the ^{15}N dimension that is
331 "transferred" from the minor to the major species. This decreases the signal intensity of the major
332 species and then leads to an additional dip in the CEST profile (major species signal intensity
333 versus saturation frequency) next to the large major species minimum if there is a difference in the
334 resonance frequencies of the two species.

335 None of the CEST profiles of *Ec*NusG-KOW, *Mt*NusG-KOW, and *Mj*Spt5-KOW exhibits an
336 exchange peak (Figure 5A – C), demonstrating that these domains are stable on the ms time scale,
337 in agreement with their high thermodynamic stabilities (see above).

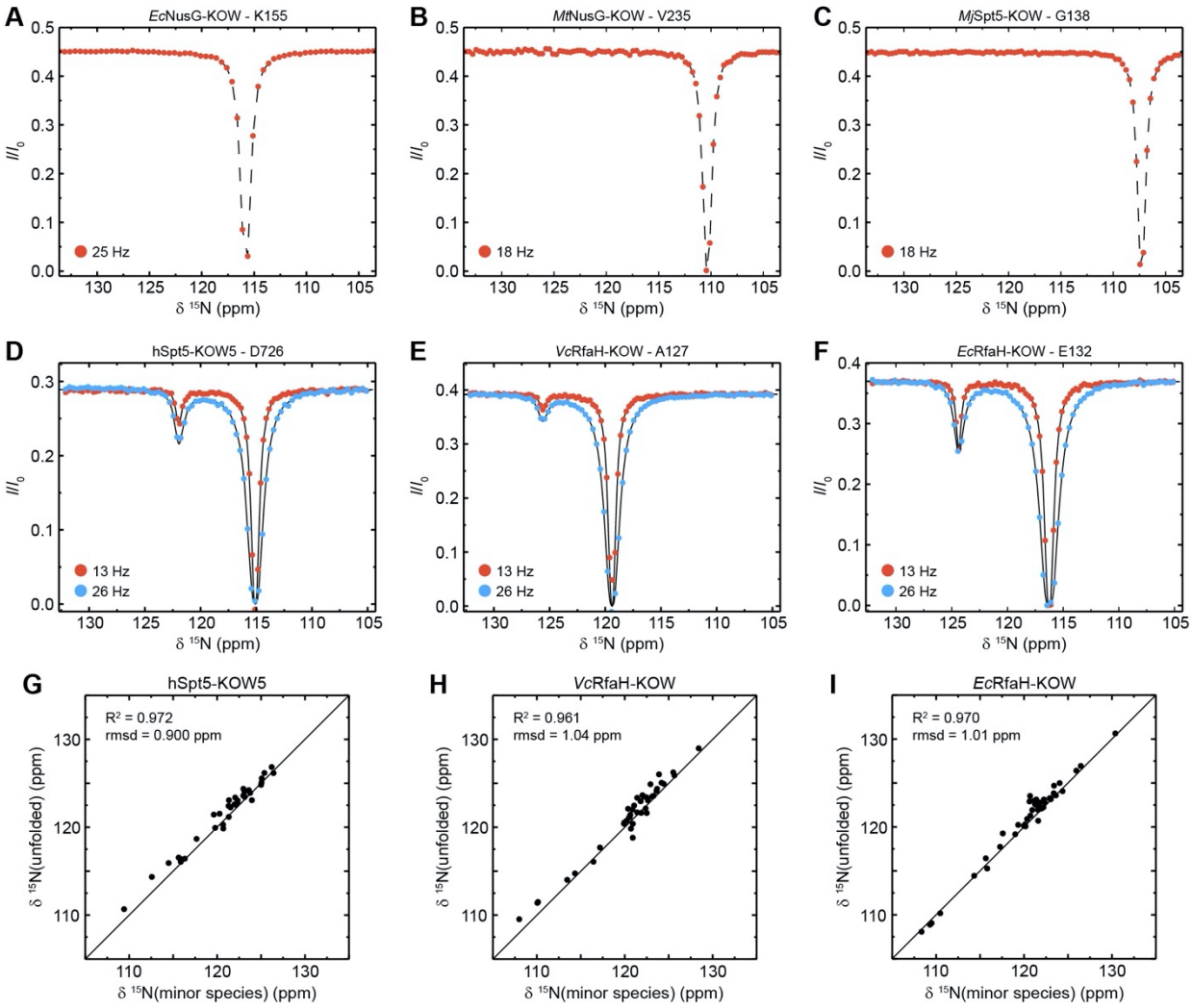

Figure 5. CEST analysis of the KOW domains. (**A**) – (**F**): Representative backbone ^{15}N-CEST profiles of the indicated KOW domain measured with one ((A) – (C)) or two ((D) – (F)) B_1 field strengths and an exchange time of 0.5 s. B_0 field for ((A)-(C)): 21.15 T; B_0 field for ((A)-(C)): 16.45 T. The lines in (D) – (F) are fits to a two-state exchange model. (**G**) – (**I**): Correlation plots showing the high similarity of the chemical shift of the minor CEST species and that of the corresponding random coil value. The latter were obtained by backbone assignment in 8 M urea (*Ec*RfaH-KOW) or are theoretical values (*Vc*RfaH-KOW, hSpt5-KOW5). The squared correlation coefficient and the rmsd between the two corresponding sets of chemical shifts are listed.
Source data 1. CEST fits for *Ec*RfaH-KOW, *Vc*RfaH-KOW and hSpt5-KOW5.
Source data 2. Experimentally determined chemical shift values of urea-denatured *Ec*RfaH-KOW and predicted random coil chemical shift values of *Vc*RfaH-KOW and hSpt5-KOW5.
Figure supplement 1. Extended CEST analysis of hSpt5-KOW5, *Vc*RfaH-KOW, and *Ec*RfaH-KOW.

353 In contrast, most CEST traces of hSpt5-KOW5, *Ec*RfaH-KOW, and *Vc*RfaH-KOW have a second
354 dip, indicating exchange with a second, low populated state (exemplary traces are shown in Figure
355 5D - F). Using a two-state exchange model we fitted all CEST traces that showed an exchange
356 signal individually to determine the residue-specific k_{ex} and p_B values. In all three cases, the k_{ex}/p_B
357 values appear to cluster in one region, suggesting a global, cooperative process (Figure 5 – Figure
358 supplement 1A). Thus, we next performed a global fit of all CEST traces for each of the three
359 proteins resulting in global rate constants and populations as well as lifetimes of the two states
360 (Table 5). This analysis yields a relatively high p_B value (5.50 %) but low k_{ex} (15.0 s^{-1}) for *Ec*RfaH-
361 KOW, a much lower p_B value (0.44 %) but higher k_{ex} (75.0 s^{-1}) for *Vc*RfaH-KOW, and p_B/k_{ex} values
362 of 0.85 % and 89.0 s^{-1} for hSpt5-KOW5.

363 To characterize the exchanging species structurally, we analyzed the chemical shifts of the minor
364 species. In all three cases, the minor species shifts show a very good correlation with those of a
365 completely unfolded conformation (Figure 5G – I; R^2 > 96 %, rmsd < 1.04 ppm). Note that the
366 chemical shifts for the unfolded state of *Ec*RfaH-KOW were obtained experimentally by backbone
367 assignment of the protein in 8 M urea, whereas those of *Vc*RfaH-KOW and hSpt5-KOW5 are
368 predicted values (see Materials and Methods for details). Determination of the relative populations
369 finally results in the equilibrium constant and the difference in Gibbs free energy, ΔG, separating
370 the energy levels of the two species (Table 5). As expected, these ΔG values are similar to those
371 obtained from the urea-based unfolding experiments at pH 7 (Table 3).

372 Taken together, the CEST experiments show that the folded all-β state of the isolated RfaH-KOWs
373 and also hSpt5-KOW5 is in equilibrium with a species that resembles an unfolded conformation.
374 As this state is easily accessible from the β-barrel, we conclude that the folding barrier separating
375 the two states cannot be too high as this would prohibit an exchange on the ms time scale.

376 **The unfolded conformers of *Ec*- and *Vc*RfaH-KOW contain transiently helical structures**

377 Although chemical shifts of the minor species of *Ec*RfaH-KOW nicely correlate with the chemical
378 shifts of the urea-unfolded protein (**Figure 5I**), there are some noticeable differences in the ^{15}N
379 chemical shifts ($\Delta\delta$ ^{15}N) of the two data sets (red bars in **Figure 6A, top panel**). In particular, two
380 regions (region 1: Q127 – T131, region 2: E136 – I150) show significant deviations of -1 to -3
381 ppm, indicating local residual structures in these regions. As the type of present (secondary)
382 structure cannot be derived from ^{15}N data, we recorded a CEST experiment on the ^{13}Cα carbons
383 of ^{13}C-*Ec*RfaH-KOW (**Figure 6 – Figure Supplement 1**) and calculated $\Delta\delta$ ^{13}Cα between the
384 observed minor species values and the random coil values obtained from the urea-unfolded protein
385 (red bars in **Figure 6A, bottom panel**). The deviations are positive in regions 1 and 2, indicating
386 the presence of helical structures at these sites. This is in agreement with secondary structure
387 predictions, which show that the Leu-rich motif (LLLNL) in region 2, where the deviations of δ
388 ^{15}N and δ ^{13}Cα are most pronounced, has high α-helical propensity (**Figure 6 – Figure**
389 **Supplement 2**; see also (Balasco et al., 2015)). Moreover, the two helical elements are located at
390 the positions of the two α-helices in the all-α form of *Ec*RfaH-KOW (compare **Figure 1B**). Due
391 to the presence of two dips the CEST profiles can be analyzed using a two-state model (minor *vs*
392 major species). Interestingly, the resulting ^{15}N transverse relaxation rates (R_2 values) of regions 1
393 and 2 in the minor species are significantly higher than corresponding rates in the β-barrel state
394 (**Figure 6A, mid panel**). Generally, one would expect that the minor species exhibits lower R_2
395 values as it is more flexible due to its largely unfolded nature (Farrow et al., 1995). The increased
396 relaxation rates thus indicate the presence of additional exchange processes on the intermediate to
397 fast chemical exchange (i.e. µs – ms) time scale. Consequently, the minor species itself seems to
398 be an ensemble of predominantely unfolded, fast interconverting structures with transient helical

in regions 1 and 2 rather than a static population. As no dips in addition to the ones of the minor and major species can be observed in the CEST profiles, the population of other states is low and beyond the detection limit of CEST experiments.

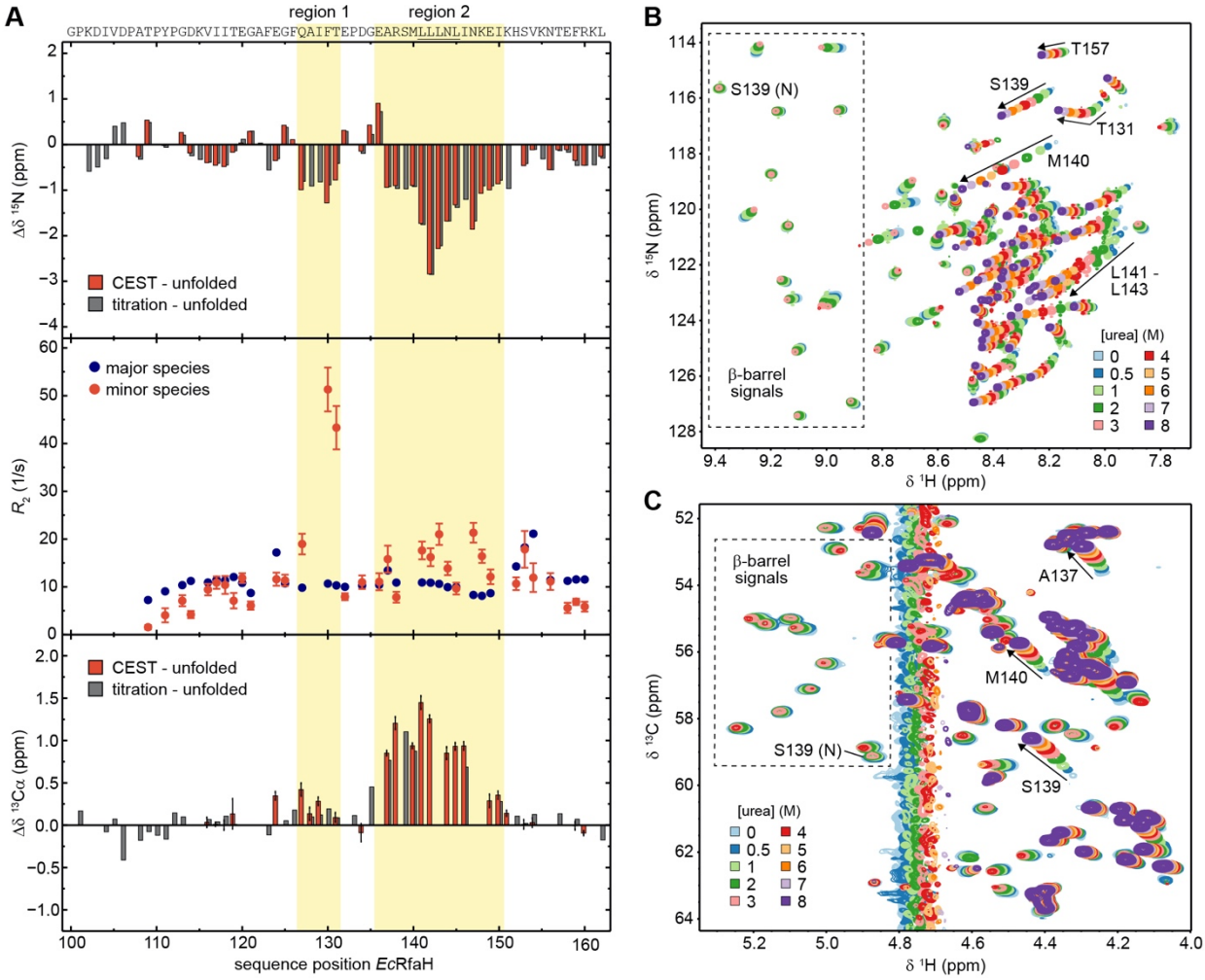

Figure 6. The minor species of *Ec*RfaH-KOW contains residual structure. (**A**) Deviations of the minor species of *Ec*RfaH-KOW from the urea-unfolded state. **Top row:** Sequence-dependent difference between the ^{15}N backbone amide chemical shifts of the minor species and of the value obtained by assignment in 8 M urea. The values for the minor species were either obtained from the CEST experiment (red bars, individual fits; "CEST – unfolded"), or by tracing back the chemical shift changes from 8 M to 0 M urea in the [^{1}H, ^{15}N]-HSQC-based urea titration (grey bars; "titration – unfolded"; see panel (B)). **Middle row:** R_2 values of the major species (*Ec*RfaH-KOW β-barrel; blue) and minor species (red), obtained from fitting the CEST profiles (global fit). Regions 1 and 2 of the minor species have R_2 values significantly higher than those of their corresponding β-barrel conformation indicating additional exchange processes, whereas N- and C-terminal regions have R_2 values lower than those of their corresponding β-barrel conformation, which is typical for random coil structures**. Bottom row:** Sequence dependent difference between

415 the ^{13}Cα chemical shifts of the minor species and of the value obtained by assignment in 8 M urea.
416 The values for the minor species were either obtained from the CEST experiment (red bars,
417 individual fits; "CEST – unfolded"), or by tracing back the chemical shift changes from 8 M to
418 0 M urea in the [^{1}H, ^{13}C]-ctHSQC-based urea titration (grey bars; "titration – unfolded"; see panel
419 (C)). The sequence of *Ec*RfaH-KOW is given above the diagram, the Leu-rich motif is underlined.
420 Regions 1 and 2 are highlighted. (**B, C**) NMR-based chemical equilibrium unfolding experiments
421 of *Ec*RfaH-KOW using urea as denaturant. The plots show an overlay of (B) [^{1}H, ^{15}N]-HSQC, and
422 (C) [^{1}H, ^{13}C]-ctHSQC spectra of [^{15}N, ^{13}C]-*Ec*RfaH-KOW, acquired in the presence of varying
423 urea concentrations. The system was buffered by 20 mM Na-phosphate (pH 6.5), 100 mM NaCl,
424 1 mM EDTA, 10 % (v/v) D_2O. Boxed regions mark signals corresponding to the β-barrel state
425 with the signal of S139 being labeled with "N" ("native"). Arrows and further labels indicate
426 signals of residues that exhibit strong chemical shift changes in the indirect dimension (^{15}N in (B),
427 ^{13}C in (C)). The spectra are colored as indicated.
428 **Source data 1.** ANS binding by *Ec*RfaH-KOW during urea-based denaturation.
429 **Figure supplement 1.** Exemplary traces of ^{13}Cα–CEST experiments of *Ec*RfaH-KOW.
430 **Figure supplement 2.** Secondary structure prediction for the six KOW domains.
431 **Figure supplement 3.** The minor species of *Vc*RfaH-KOW contains residual structure.
432 **Figure supplement 4.** The minor species of hSpt5-KOW5 is completely unfolded.
433 **Figure supplement 5.** The intermediate state of *Ec*RfaH-KOW is no equilibrium MG.
434 **Figure supplement 6.** Extended analysis of the urea-induced denaturation of *Ec*RfaH-KOW.
435

436 Like *Ec*RfaH-KOW, the minor species of *Vc*RfaH-KOW also seems to contain residual structure
437 (Figure 6 – Figure supplement 3A). As the unfolded state of this domain was not assigned
438 experimentally, predicted chemical shift values for the random coil structure were used for the
439 correlation plot (Figure 5H). When plotting the Δδ ^{15}N values versus the sequence position (Figure
440 6 – Figure supplement 3A), the resulting pattern resembles the one obtained for *Ec*RfaH-KOW
441 (compare Figure 6A, top panel). The regions around residues 103 - 125 (linker) and 155 - 165 (C-
442 terminus) show relatively low Δδ ^{15}N values, indicating a random coil structure, whereas the
443 region around residues 140 - 150 (corresponding to region 2 in *Ec*RfaH-KOW) exhibits
444 significantly increased Δδ ^{15}N values, suggesting residual structure, similar to *Ec*RfaH-KOW.
445 However, only very small minor species dips were observed in some traces of a CEST experiment
446 recorded on the ^{13}Cα carbons of ^{13}C-*Vc*RfaH-KOW (Figure 6 – Figure supplement 3B), which we
447 attribute to the very low population of the *Vc*RfaH-KOW minor species (0.4 %) that is at the

448 detection limit of the Cα-CEST experiment (which is less sensitive than the ^{15}N-CEST).
449 Consequently, we analyzed the CEST profiles only qualitatively. Unambiguous minor species dips
450 could be identified for amino acids predominantly located in the region with residual structure
451 with chemical shifts that are downfield-shifted as compared to random coil values (Figure 6 –
452 Figure supplement 3B), indicating the presence of helical elements. As for *Ec*RfaH-KOW, this is
453 in full agreement with secondary structure predictions, which suggest that all NusG/Spt5-KOW
454 domains adopt 4-5 β-strands whereas both RfaH-KOW domains exhibit propensities for both β-
455 strands and α-helices, especially in the regions with residual structure in the CEST minor species
456 (Figure 6 – Figure supplement 2). Taken together this data suggests that the *Vc*RfaH-KOW minor
457 species also contains transient residual helical structures.

458 The hSpt5-KOW5 domain is part of an "RNA clamp" during transcription elongation in eukaryotes
459 (Bernecky et al., 2017) and exhibits the typical β-barrel fold in all available structures. Strikingly,
460 hSpt5-KOW5 also exchanges with an unfolded species under non-denaturing conditions (Figure
461 5G), just as *Ec*RfaH-KOW and *Vc*RfaH-KOW. The magnitude of the differences between the
462 minor species ^{15}N chemical shifts and the predicted random coil values (Figure 6 – Figure
463 supplement 4A) is similar to that observed for *Vc*RfaH-KOW (Figure 6 – Figure supplement 3A).
464 Interestingly, the minor species' chemical shifts of a ^{13}Cα -CEST of ^{13}C-hSpt5-KOW5 clearly
465 indicate the absence of any substantial residual structure (Figure 6 – Figure supplement 4). In
466 contrast to all other KOW domains in this study, hSpt5-KOW5 is not located at the very C-terminus
467 of full-length hSpt5, but it is just one out of seven KOW domains being flanked by several
468 hundreds of residues at either terminus. Thus, the stability of this domain may be different in its
469 physiological environment. Taken together, this data suggests that hSpt5-KOW5 is a typical
470 monomorphic β-barrel and that its decreased stability, accompanied by the existence of a minor,

471 unfolded species, may be attributed to the absence of the neighboring domains, although we cannot
472 completely rule out that these features are real, intrinsic properties of hSpt5-KOW5 in the full-
473 length protein with (yet unknown) functional relevance.

474 As the completely unfolded state was only experimentally assigned for *Ec*RfaH-KOW we will
475 focus on this domain in the following analysis. Owing to its population of 5.5 % (Figure 5 – Figure
476 supplement 2), *Ec*RfaH-KOW's minor species should be detectable in standard HSQC spectra,
477 given a sufficiently high signal-to-noise ratio. As we observed a stable intermediate during the CD-
478 based chemical unfolding of *Ec*RfaH-KOW we aimed at analyzing the role of the minor species
479 during the chemical denaturation of *Ec*RfaH-KOW by recording [^{1}H, ^{15}N]- and [^{1}H, ^{13}C]-
480 correlation spectra of [^{15}N, ^{13}C]-labeled *Ec*RfaH-KOW in the presence of various urea
481 concentrations (0 – 8 M) (Figure 6B,C). In both spectra series, we observed a decrease in peak
482 intensity/volume of the β-barrel signals with increasing urea concentration (boxed regions in
483 Figure 6B,C), which is completed at ≈ 4 M urea, indicating that the first transition in the far-UV
484 CD-based chemical denaturation of *Ec*RfaH-KOW (Figure 3F) corresponds to the unfolding of its
485 β-barrel (tertiary) structure. This is also corroborated by near-UV CD spectroscopy-based
486 chemical denaturation experiments using urea or GdmCl, respectively, (Figure 6 – Figure
487 supplement 5A, B), which clearly show that the transition during the titration from 0 to ~3 M
488 urea/~1 M GdmCl is accompanied by a loss in tertiary structure. The possibility that the resulting
489 conformation corresponds to an equilibrium molten globule is, however, excluded due to its
490 inability to bind 8-anilino-1-naphthalenesulfonic acid (ANS, Figure 6 – Figure supplement 5C).

491 In order to identify signals corresponding to the minor species in the HSQC spectra of *Ec*RfaH-
492 KOW we started with the spectrum of the urea-unfolded protein (8 M urea, purple spectra in Figure
493 6B,C). Most of the corresponding signals shifted linearly with decreasing urea concentration and

494 also lost intensity at urea concentrations < 3 M (e.g. signal of S139 in Figure 6B,C). At 0 M urea,
495 finally, only a set of weak signals remained, which we identified as signals of the minor species as
496 these match the chemical shifts of the minor species identified in the CEST experiments (compare
497 red and grey bars in Figure 6A, top/bottom panels). Based on the linear transition between the
498 positions of the (urea) unfolded state towards the positions of the minor species signals we
499 conclude that addition of urea shifts the minor species' population towards the completely unfolded
500 state. Although we cannot assess if the minor species samples the completely unfolded state in the
501 absence of any denaturant, the increased ^{15}N R_2 values indicate additional exchange processes of
502 the minor species on the µs-ms time scale (Figure 6A, middle panel). Thus, we hypothesize that
503 the minor species can be described as an ensemble of exchanging sub-states, some corresponding
504 to the completely unfolded state "U" and some exhibiting residual helical structure, hereby referred
505 to as α-helical unfolding intermediate "Uα" with the minor species observed in the CEST
506 experiments being the average population under native conditions.

507 If this is true, the urea-induced chemical shift perturbations experienced by the minor species
508 signals in the [^{1}H, ^{15}N]-HSQCs can be explained by a combination of two effects: (i) change of
509 the chemical environment of the spins due to the presence of urea, which particularly affects δ ^{1}H
510 (see e.g. signal of T157 in Figure 6B), and (ii) change in the relative populations of the minor
511 species' sub-states towards the unfolded state, which mainly affects Δδ ^{15}N. Since the Hα/Cα
512 chemical shifts are relatively independent of the solvent conditions, their perturbations in the urea
513 denaturation series (Figure 6C) even better reflect the change in the ratio of the minor species' sub-
514 states. The shifting of the minor species' peaks in Figure 6C is completed at ≈7 M urea, implying
515 that the second transition in the far-UV CD-based unfolding experiment (Figure 3F) corresponds
516 to the denaturation of Uα. Interestingly, the R_2 values of residues in region 1 are more than twice

517 as high as those of residues in region 2 (Figure 6A) and, in the [^{1}H, ^{15}N]-HSQC-based denaturation
518 experiment (Figure 6B), the minor species' signals of residues in region 1 do not shift in a linear
519 manner as it is typical for two exchanging states. Instead, they show a curved transition that is
520 "kinked" at ≈ 2 M urea (see e.g. T131), implying a more complex unfolding process and thus
521 structural heterogeneity of this region. Although our experiments do not allow a precise structural
522 characterization of all states of the minor species, it may be described as an ensemble of largely
523 unfolded, interconverting structures with states U and Uα constituting the extrema.

524 Due to the fast chemical exchange between the *Ec*RfaH-KOW's U and Uα states during the
525 chemical denaturation, their relative populations in a certain titration step are encoded in the
526 chemical shift of the minor species signal, whereas the volume of the minor species peak is
527 proportional to the sum of the populations of both states (assuming similar transverse relaxation
528 rates for the species). The chemical shifts of Cα/Hα groups depend to a much lower extent on the
529 urea concentration in the sample than the chemical shifts of amide groups and therefore they
530 provide better measures for the exchange between U and Uα. To first quantify the decay of the all-
531 β conformation and the increase of the minor species during the urea denaturation we analyzed the
532 peak intensity of both species exemplarily for residue S139 in the [^{1}H, ^{13}C]-ctHSQC-based titration
533 (Figure 6C and Figure 6 – Figure supplement 6A). The resulting ΔG value of ≈ 7 kJ/mol between
534 the energy levels of major and minor species agrees well with the results from the CEST
535 experiment (7 kJ/mol). Additionally, the *m* value of 3.4 kJ/(mol M) is very similar to the *m* values
536 obtained for the other KOW domains by CD spectroscopy (Table 3), indicating that the minor
537 species is indeed close to a completely unfolded state with a small buried surface area.

538 The complete denaturation of the minor species, i. e. the transition of Uα to a fully unfolded state
539 U, can be followed in the [^{1}H, ^{13}C]-ctHSQC-based denaturation experiment by analyzing the
540 change of the minor species's chemical shifts from the positions in the absence of urea towards
541 those of the completely unfolded state. For example, the Hα/Cα correlation peaks of residues
542 A137, S139, or M140, which are located in region 2, clearly shift from regions typical for α-helical
543 structures (upfield ^{1}H, downfield ^{13}C relative to random coil values) to positions corresponding to
544 an unstructured conformation (downfield ^{1}H, upfield ^{13}C), and finally they localize next to the
545 signals of the Ala, Ser or Met residues that do not reside in regions with residual helical structure
546 (**Figure 6C**). Plotting the changes of the ^{13}Cα chemical shifts of A137, S139 and M140 versus
547 the urea concentration (**Figure 6 – Figure supplement 6B**) results in curves that resemble the
548 second half of an unfolding transition (Uα ⇌ U) and approach the baseline of the fully unfolded
549 state at ≈ 6 M urea. The absence of a baseline for Uα precludes a quantitative analysis, but it
550 indicates that the transition mid-point of the curve is probably close to or below 0 M urea. In
551 summary, the data of the NMR-based denaturation experiments (i) strongly support our hypothesis
552 that the minor species identified in the CEST experiments is an ensemble of fast interconverting,
553 mostly unfolded structures with U and Uα being the extreme and (ii) suggest that the minor species
554 might be an important intermediate during the refolding process.

555 **Discussion**

556 **Fold-switching is conserved among RfaH proteins**

557 Genes coding for RfaH orthologs can be found in many bacterial pathogens, including *Salmonella*,
558 *Klebsiella*, *Vibrio*, and *Yersinia* spp (Carter et al., 2004). Despite their divergent evolution, RfaH
559 proteins seem to have a conserved mechanism of action (Carter et al., 2004). To date only *Ec*RfaH
560 was structurally characterized in detail, revealing that this protein has unique structural features
561 classifying it as transformer protein (Belogurov et al., 2007; Burmann et al., 2012; Zuber et al.,
562 2019). Here, we show that *Vc*RfaH, an evolutionary quite divergent representative sharing 35.8 %
563 sequence identity with *Ec*RfaH, exhibits very similar structural properties, i.e. *Vc*RfaH-KOW, like
564 *Ec*RfaH-KOW, folds as α-hairpin in the full-length protein, but adopts a NusG-type β-barrel
565 conformation in its isolated form (Figure 1). Interestingly, in *Vc*RfaH helix α_3* is 1.5 turns longer
566 as compared to *Ec*RfaH and *Vc*RfaH has a disulfide bridge connecting strand β_3* and helix α_3*,
567 stabilizing this helix. These two features imply a stabilization of the domain interface and thus an
568 increased affinity between the domains as compared to *Ec*RfaH. This might also explain the
569 increased stability of the isolated *Vc*RfaH-KOW domain (≈ 14 kJ/mol), which compensates the
570 higher energy gain of the domain interaction. Further, the increased stability of the *Vc*RfaH-KOW
571 domain may be the cause for the sigmoid-shaped CD-based chemical denaturation curves, in
572 agreement with an apparent two-state unfolding process: global unfolding of the folded state
573 occurs at higher denaturant concentrations, where potential partly structured folding intermediates
574 are already largely destabilized and therefore escape detection. This conclusion is supported by the
575 Trp fluorescence-based denaturation data (Figure 3 – Figure supplement 1), suggesting that the
576 change in the CD signal is almost exclusively caused by the decay of the β-barrel conformation
577 and that the contribution of Uα to the change of the CD-signal is negligible. Nevertheless, we

578 conclude that *Vc*RfaH may be regulated by fold-switching just like *Ec*RfaH, and that this
579 metamorphic behavior is conserved in the class of RfaH proteins and may even be found in other
580 NusG paralogs, in agreement with a recent study that predicts that nearly 25 % of bacterial NusG
581 proteins might perform α ↔ β transitions similar to *Ec*RfaH (Porter et al., 2022).

582

583 **Model for the structural plasticity of RfaH**

584 *Ec*RfaH switches the conformation and function of its KOW domain in a reversible manner to
585 achieve a tight control of gene expression (Zuber et al., 2019). In free *Ec*RfaH, the α-helical hairpin
586 conformation is the preferred state of *Ec*RfaH-KOW, whereas domain separation or isolation of
587 *Ec*RfaH-KOW foster population of the all-β state in solution (Burmann et al., 2012), suggesting
588 that the all-α conformation is intrinsically unstable, but becomes the thermodynamic minimum in
589 free *Ec*RfaH due to interaction with *Ec*RfaH-NGN.

590 Interestingly, our thermodynamic analysis (Figures 2 and 3) of the isolated *Ec*RfaH-KOW domain
591 reveals that, although the all-β conformation is the preferred state in isolation, it is only marginally
592 stable, and it is in rapid equilibrium with an "unfolded" state, which is populated to a significant
593 extent, even under physiological conditions. The "unfolded" state is a mixture of random-coil-type
594 unfolded species U and species Uα containing two helical regions .

595 Based on our results, we suggest a model for the structural transitions of *Ec*RfaH-KOW (Figure
596 7).

597

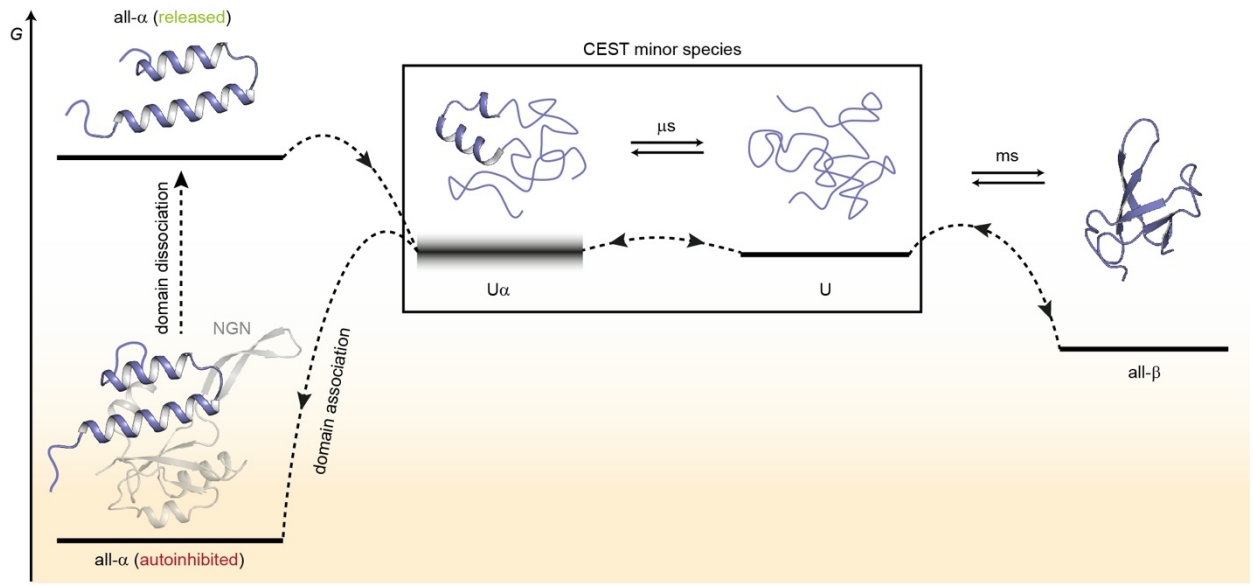

Figure 7. Model for the conformational plasticity of *Ec*RfaH-KOW. Qualitative Gibbs free energy level diagram and associated structures for the all-α to all-β transition of *Ec*RfaH-KOW and *vice versa*. In its ground state, i.e. the autoinhibited conformation, the energy of the all-α conformation of *Ec*RfaH-KOW is strongly lowered by the extensive inter-domain contacts with the *Ec*RfaH-NGN domain. Upon recruitment, the domains dissociate, the helical structure of the released KOW domain becomes destabilized in isolation, and rapidly decays towards an ensemble of mainly unfolded sub-states that interconvert on the μs time scale. Some of the sub-states correspond to the completely unfolded state (U) whereas others retain some residual (α-) helical elements (Uα). The scheme displays exemplary structures of these sub-states. Due to their fast structural interconversion, U and Uα may be grouped into a single macro-state/ensemble (as is the case during the CEST experiments) that exhibits helical structures for a limited amount of time and is otherwise unfolded. Uα is either marginally stable or even unstable (therefore, its energy level is blurred). The disordered conformation then allows for easy and rapid refolding to the all-β conformation. Due to their low thermodynamic stability, or even instability of all-β and Uα, respectively, the last two steps are reversible, that is the all-α state can be rapidly regained when the *Ec*RfaH-NGN domain becomes available for re-association after transcription termination.

In the autoinhibited state the all-α conformation of *Ec*RfaH-KOW corresponds to the minimum of the Gibbs free energy as it is stabilized by contacts to the *Ec*RfaH-NGN. During recruitment of *Ec*RfaH to an *ops*-paused elongation complex, the *Ec*RfaH-NGN:KOW interface is destabilized (most probably *via* an encounter complex), the domains dissociate and *Ec*RfaH-NGN is sequestered to RNAP (Zuber et al., 2019). The freed all-α *Ec*RfaH-KOW is not stable as *G* increases due to the loss of *Ec*RfaH-NGN contacts. Consequently, *Ec*RfaH-KOW unfolds,

resulting in an ensemble of rapidly interconverting sub-states. Some of these sub-states still contain two residual α-helical regions (intermediate Uα) that correspond to the tip of the α-hairpin in the all-α state, in agreement with hydrogen/deuterium exchange data, which indicate that the hairpin-tip is the most stable part of the all-α conformation (Galaz-Davison et al., 2020). Other sub-states represent the completely unfolded protein, which then rapidly refolds into the all-β form. Upon transcription termination *Ec*RfaH is released, and the process is reversed with unfolding of the β-barrel starting, most probably, by detaching β_1 and β_4/β_5 from the central strands as the corresponding H-bonds are the least stable ones (Figure 4). The U state is in equilibrium with Uα, where two α-helical regions that will later constitute the α-hairpin tip are formed transiently and may thus serve both as the nucleation point for the completion of the all-α structure and as starting point for recognition of its cognate binding site on the NGN. This mechanism ensures rapid re-autoinhibition and prevents aggregation of *Ec*RfaH. Although we did not analyze *Vc*RfaH as extensively as *Ec*RfaH, our results suggest that the *Vc*RfaH-KOW domain most likely employs a similar mechanism for its structural transformation, indicating that the presented model is a general scheme for RfaH proteins.

In support of our model, all computational studies on *Ec*RfaH found that the all-α conformation is stable only when in contact with the NGN. Modification of the strength of the *Ec*RfaH-NGN:KOW interface (Ramírez-Sarmiento et al., 2015) or deletion of the linker (Xun et al., 2016) destabilize the all-α fold and ultimately drive *Ec*RfaH-KOW into the β-barrel state. Moreover, the β-barrel fold is stable and corresponds to or is close to the energy minimum of the energy landscape of *Ec*RfaH-KOW, whereas the all-α fold rapidly unfolds and has a higher *G*-value than the all-β state (Balasco et al., 2015; Bernhardt and Hansmann, 2018; Gc et al., 2014; Joseph et al., 2019; Li et al., 2014; Xiong and Liu, 2015). Apart from these general concepts, most studies differ in several

645 key points, such as the extent to which the all-α state is populated in the isolated *Ec*RfaH-KOW,
646 or the precise folding pathway from all-α to all-β. Strikingly, a recent bioinformatical study very
647 nicely mirrors our data as the authors also observed a significant portion of transiently formed
648 helical structure within the unfolded state ensemble in their simulations (Seifi and Wallin, 2021).
649

650 **Requirements for fold-switching proteins**

651 Previous work on designed and naturally occurring fold-switching proteins has identified several
652 specific properties that make fold-switching proteins distinct from others (Bryan and Orban, 2010;
653 Porter and Looger, 2018). In this study, we show that RfaH meets all these requirements and is
654 thus a showcase example for fold-switching proteins:

655 (i) Reduced thermodynamic stability (Bryan and Orban, 2010). A diminished stability is both the
656 result of and key to the function of fold-switching proteins. As the fold-switching sequence must
657 be compatible with both adopted topologies, it can only be optimized to a certain extent to stabilize
658 one specific fold, ensuring that both conformations can be interconverted and that the structure is
659 not "trapped" in one state. This is reflected by a dual-funneled energy landscape with two main
660 minima, which are, however, not as deep as the global minimum of a stable protein. Our
661 comprehensive thermodynamic analysis (Figures 2 and 3) reveals that the all-β fold of both RfaH-
662 KOWs is indeed less stable than the bacterial and archaeal NusG/Spt5-KOW domains. As general
663 transcription factors, NusG/Spt5 proteins do not require an as-sophisticated regulation as RfaH
664 (Artsimovitch and Knauer, 2019) (for hSpt5-KOW5 see below) and thus benefit from a stable
665 structure to carry out their function. The difference in thermodynamic stability is especially
666 striking for *Ec*NusG-KOW and *Ec*RfaH-KOW as both belong to the same class of transcriptional
667 regulators, originate from the same organism and share a sequence identity of 35.8 % (43.6 % for

668 the full length proteins), yet underly completely different regulatory mechanisms that, in turn,
669 strongly depend on the difference of this thermodynamic parameter. As a result, *Ec*RfaH is tightly
670 regulated by autoinhibition coupled to the conformational transformation of a whole domain and
671 controls just a small set of specific genes whereas *Ec*NusG is a stable, monomorphic protein
672 involved in the transcription of most host genes.

673 (ii) Generation of new binding surfaces (Bryan and Orban, 2010). The regulation of
674 conformational transitions in fold-switching proteins is achieved by energetically stabilizing one
675 of the two conformations in response to a molecular trigger, resulting in a far more dynamic energy
676 landscape than that of well-folded, monomorphic proteins as the energy level of a particular
677 conformation strongly depends on the environment. This context-dependent stabilization of one
678 state is possible because the two different folds exhibit different surface topologies, each allowing
679 distinct interactions. The ability to selectively hide/expose "latent" binding sites within different
680 folds is also the most important function of fold-switching in general, as it enables a level of control
681 that cannot be achieved by other mechanisms. In RfaH, autoinhibition is coupled to a
682 conformational switch. In the autoinhibited state the all-α KOW interacts with the RfaH-NGN to
683 prevent off-target recruitment and interference with NusG (Belogurov et al., 2009) whereas the
684 refolded state allows simultaneous binding of RfaH to RNAP via RfaH-NGN and to the ribosome
685 via all-β RfaH-KOW to activate translation (Kang et al., 2018; Zuber et al., 2019).

686 (iii) Unfolded regions in one of the two states (Bryan and Orban, 2010). In RfaH-KOW, the central
687 β-strands β_2, β_3, and β_4 transform into two α-helices during the all-β-to-all-α transition and *vice*
688 *versa* (Figure 1B). However, the all-α KOW domain contains unstructured N- and C-termini,
689 whereas the corresponding regions form β-strands β_1 and β_4/β_5 in the all-β conformation (Figure
690 1B and Figure 4C). These disordered parts provide an entropic stabilization of the respective state

691 as they do not adopt a defined structure and the corresponding β-strands are less stably bound to
692 the rest of the β-barrel than in the stable KOW domains (Figure 4). A bioinformatic study indicated
693 that these regions of the CTD additionally stabilize the NGN:KOW interface by forming transient,
694 IDP-like interactions (Xun et al., 2016). We show that the structural interconversion between the
695 two RfaH states proceeds *via* a chiefly unfolded intermediate and we propose that the disordered
696 segments may help to facilitate and/or initiate this transition, similar to the mechanism suggested
697 for the human chemokine XCL1 (lymphotactin) (Tyler et al., 2011). Finally, disordered regions in
698 one state have the advantage that they can be evolutionary optimized to selectively stabilize one
699 of the two states of a fold-switch pair, whereas there is no need to fit a defined three-dimensional
700 structure in the other state. This is reflected by the secondary structure predictions of both *Ec*RfaH-
701 KOW and *Vc*RfaH-KOW which suggest propensities for both helical structures and β-strands in
702 the central region that indeed interconvert between α-helices and β-strands whereas only β-strands
703 are predicted for the termini (Figure 6 – Figure supplement 2).
704 (iv) Divergence in predicted and observed secondary structure (Porter and Looger, 2018).
705 Secondary structure predictions show that both *Vc*RfaH-KOW and *Ec*RfaH-KOW contain
706 stretches with high propensity for both β-strands and α-helical structures , whereas NusG/Spt5-
707 KOW domains are predicted to adopt 4-5 β-strands (Figure 6 – Figure supplement 2). Three
708 dimensional structures of the KOW domains of this study confirm that the NusG/Spt5-KOW
709 domains are indeed β-barrels whereas the fold of the RfaH-KOW domains depends on the context
710 (Figure 1 and Figure 1 – Figure supplement 2). Interestingly, this tendency is also visible in the
711 isolated KOW domain as the disordered regions in the all-α fold correspond to the β-strands that
712 are less stable in the RfaH-KOWs as compared to NusG-KOWs, whereas the helical propensity is
713 reflected in the structure of Uα. Importantly, secondary structure predictions strongly depend on

714 the underlying algorithms and, as can be seen for *Mt*NusG-KOW (Figure 6 – Figure supplement
715 2), ambiguity in the secondary structure predictions are a necessary, yet not sufficient condition
716 for a protein to switch folds.

717 (v) Cooperatively folding units (Porter and Looger, 2018). The folding cooperativity of *Ec*RfaH-
718 KOW depends on the presence of the *Ec*RfaH-NGN, i.e. in the absence of *Ec*RfaH-NGN *Ec*RfaH-
719 KOW folds cooperatively on its own. However, the cooperativity is generally rather low and the
720 activation barrier separating the "unfolded" and the folded states is small, allowing fast transitions.

721

722 **Fold-switching is a highly efficient principle of regulation with a steadily increasing**
723 **importance.**

724 To date about six fold-switching proteins have been studied in detail (summarized in refs.
725 (Dishman and Volkman, 2018; Lella and Mahalakshmi, 2017; Zamora-Carreras et al., 2020)), but
726 estimates suggest that up to 4 % of the proteins in the PDB may have the ability to switch folds
727 (Porter and Looger, 2018). Our study demonstrates which molecular mechanisms confer RfaH its
728 structural plasticity that allows operon-specific regulation without competing with its
729 monomorphic paralog NusG/Spt5. In line with our findings, a recent study on XCL1, another
730 model system for fold-switching proteins, identified very similar principles for the evolution and
731 design of fold-switching proteins (Dishman et al., 2021).

732

733 **Importance of a chiefly unfolded state in protein fold-switching**

734 In summary, our results highlight two key features in protein fold-switching: decreased
735 thermodynamic stability and defined local structures in "unfolded" intermediates. Diminished
736 stability is often thought to be detrimental for proteins as it favors non-native contacts and

promotes aggregation. However, it is essential to confer fold-switching proteins their conformational plasticity, and, as all transitions from and to the unfolded states are very fast, and the population of these states is rather low, fold-switchers can evade aggregation. Further, the capability of the "unfolded" state to harbor residual defined structures, e.g. α-helices, allows to pre-encode a second conformation that could be readily adopted upon a molecular signal.

745 **Materials and Methods**

746

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
bacterial strain	*Escherichia coli* BL21(DE3)	Novagen	N/A	
recombinant DNA reagent	List of recombinant plasmids used	Table 7		
sequence-based reagent	List of primers used	Table 6, Biolegio	PCR primers	
recombinant protein	*V. cholerae* RfaH	This work		See Materials & Methods, section "Production of recombinant proteins"
recombinant protein	*V. cholerae* RfaH-KOW	This work		See Materials & Methods, section "Production of recombinant proteins"
recombinant protein	*E. coli* RfaH-KOW	(Burmann et al., 2012) doi: 10.1016/j.cell.2012.05.042		
recombinant protein	*E. coli* NusG-KOW	(Burmann et al., 2010) doi: 10.1126/science.1184953		
recombinant protein	*M. tuberculosis* NusG-KOW	(Strauß et al., 2016)		

		doi: 10.1080/07391102.2015.1031700		
recombinant protein	*M. janaschii* Spt5-KOW	This work		See Materials & Methods, section "Production of recombinant proteins"
recombinant protein	human Spt5-KOW5 (G699-G754)	This work		See Materials & Methods, section "Production of recombinant proteins"
commercial kit	QIAquick Gel Extraction Kit	Qiagen	Cat#: 28706	
commercial kit	QIAprep Spin Miniprep Kit	Qiagen	Cat#: 27106	
chemical compound	$(^{15}NH)_4SO_4$	Sigma/Merck KGaA	Cat#: CS01-185_148	
chemical compound	D_2O	Euriso-Top GmbH	Cat#: 7789-20-0	
chemical compound	^{13}C-D-glucose	Euriso-Top GmbH	Cat#: CLM-1396-10	

chemical compound	Urea	Carl Roth GmbH & Co. KG	Cat#: 2317.1	
chemical compound	GdmCl	Carl Roth GmbH & Co. KG	Cat#: 0037.1	
chemical compound	DSS	Sigma	Cat#: T-8636	
chemical compound	ANS	Sigma/Merck KGaA	Cat#: 10417-5G-F	
software	Fit-o-Mat v0.752	(Möglich, 2018) doi: 10.1021/acs.jchemed.8b00649		
software	PyMol v. 1.7	The PyMOL Molecular Graphics System, Schrödinger, LLC.	https://pymol.org/2/	
software	NMRViewJ	One Moon Scientific, Inc.	http://www.onemoonscientific.com/nmrviewj	
software	ChemEx v. 0.6.1	(Vallurupalli et al., 2012) doi: doi:10.1021/ja3001419	https://github.com/gbouvignies/ChemEx	
other	quartz cuvette for CD spectroscopy, 1 mm	Hellma GmbH & Co. KG		See Materials & Methods, section CD spectroscopy

other	quartz cuvette for CD spectroscopy, 2 mm	Hellma GmbH & Co. KG		See Materials & Methods, section CD spectroscopy
other	quartz cuvette for fluorescence spectroscopy, 1 cm	Hellma GmbH & Co. KG		See Materials & Methods, section fluorescence spectroscopy

747

748

749 **Cloning**

750 The *Vc*RfaH expression vector pVS13 (*V. cholerae rfaH* from pHC301 (Carter et al., 2004) in
751 plasmid pTYB1 (NEB)) was a gift from I. Artsimovitch, The Ohio State University, Columbus,
752 OH. The C-terminal *Vc*RfaH residue, Thr165, is substituted by an Ala to ensure efficient cleavage
753 of the resulting chitin binding domain (CBD) intein fusion protein (see below). Expression
754 plasmids for *Vc*RfaH-KOW (residues E103-T165), hSpt5-KOW5 (residues G699-G754), and
755 *Mj*Spt5-KOW (residues K83-D147) were created by cloning of the corresponding gene regions
756 into vector pETGb1a (G. Stier, EMBL, Heidelberg, Germany) *via Nco*I and *BamH*I (*Vc*RfaH-
757 KOW), or *Nco*I and *EcoR*I (hSpt5-KOW5 and *Mj*Spt5-KOW) restriction sites, respectively.
758 Templates for PCR amplification were plasmids pHC301 (Carter et al., 2004) for *Vc*RfaH-KOW,
759 pOTB7_huSUPT5H (Zuber et al., 2018) for hSpt5-KOW5, and pGEX-2TK_*Mj*Spt5-KOW
760 ((Hirtreiter et al., 2010); kindly provided by F. Werner, University College London, UK) for
761 *Mj*Spt5-KOW. The primers used for cloning are listed in **Table 6**. All plasmids used in this study
762 are listed in **Table 7**.

763

764 **Production of recombinant proteins**

765 *Vc*RfaH was obtained from a CBD intein fusion protein encoded in plasmid pVS13, with
766 expression conditions and purification strategy as described for *E. coli* RfaH (Vassylyeva et al.,
767 2006). *Ec*NusG-KOW and *Mt*NusG-KOW were produced as previously described (Burmann et
768 al., 2010; Strauß et al., 2016). *Mj*Spt5-KOW, hSpt5-KOW5, *Ec*RfaH-KOW, and *Vc*RfaH-KOW
769 were obtained from Gb1 fusions with expression and purification conditions similar to that of
770 *Ec*RfaH-KOW (Burmann et al., 2012).

771 The quality of all recombinantly produced proteins was ensured according to the guidelines
772 established by ARBRE-MOBIEU and P4EU (https://arbre-mobieu.eu/guidelines-on-protein-
773 quality-control/) (de Marco et al., 2021). In brief, purity was checked by sodium dodecyl sulfate
774 polyacrylamide gel electrophoresis, the absence of nucleic acids by UV spectroscopy, the identity
775 by mass spectrometry and/or NMR spectroscopy, the folding state by CD and/or NMR
776 spectroscopy, and the absence of aggregation by analytical gel filtration or dynamic light
777 scattering.

778

779 **Isotopic labeling of proteins**

780 For the production of ^{15}N- and ^{15}N, ^{13}C-labelled proteins, *E. coli* cells were grown in M9 medium
781 (Green et al., 2012; Meyer and Schlegel, 1983) containing $(^{15}NH_4)_2SO_4$ (Sigma/Merck KGaA,
782 Darmstadt, Germany) or $(^{15}NH_4)_2SO_4$ and ^{13}C-D-glucose (Euriso-Top GmbH, Saarbrücken,
783 Germany), respectively, as sole nitrogen or carbon sources. Deuteration was achieved by
784 accustoming cells to M9 medium prepared with increasing concentrations of D_2O (0 %, 50 %,
785 100 % (v/v); Euriso-Top GmbH, Saarbrücken, Germany). Expression and purification protocols
786 were identical to those of the unlabeled proteins.

787

788 **NMR spectroscopy**

789 NMR experiments were conducted at Bruker Avance 600, Avance 700, Ascend Aeon 900, and
790 Ascend Aeon 1000 spectrometers, each equipped with room temperature (Avance 600) or
791 cryogenically cooled, inverse ^{1}H, ^{13}C, ^{15}N triple resonance probes (all other spectrometers). All
792 measurements were conducted in 5 mm tubes with a sample volume of 550 µl at 25 °C, if not
793 stated otherwise. NMR data was processed using in-house software and analyzed using
794 NMRViewJ (OneMoon Scientific).

795 Backbone resonance assignments for *Vc*RfaH, *Vc*RfaH-KOW, hSpt5-KOW5, *Mj*Spt5-KOW, and
796 urea-unfolded *Ec*RfaH-KOW were obtained using standard band-selective excitation short
797 transient (BEST) (Lescop et al., 2007; Schanda et al., 2006) Transverse Relaxation Optimized
798 Spectroscopy (TROSY)-based triple resonance experiments (Pervushin et al., 1997; Salzmann et
799 al., 1998). Additionally, carbon-detected CACO, CAN, and NCO experiments (Bermel et al.,
800 2005) were recorded for *Vc*RfaH-KOW. Side chain assignments for *Vc*RfaH-KOW were obtained
801 from CCH- and H(C)CH-TOCSY, HBHA(CO)NH, C(CO)NH, aromatic [^{1}H, ^{13}C]-HSQC and ^{13}C-
802 edited aromatic Nuclear Overhauser Enhancement Spectroscopy (NOESY) experiments (Sattler
803 et al., 1999). Three-dimensional assignment and NOESY experiments were acquired using a Non-
804 Uniform Sampling scheme with a sparsity of 25 - 50 %. Spectra were subsequently reconstructed
805 with in-house written software using the iterative soft thresholding algorithm (Hyberts et al.,
806 2012). The *Ec*RfaH-KOW, *Vc*RfaH-KOW, hSpt5-KOW5 and *Mj*Spt5-KOW samples contained
807 0.5 – 1 mM [^{15}N, ^{13}C]-labeled protein in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1
808 mM Ethylenediaminetetraacetic acid (EDTA), 10 % (v/v) D_2O. The *Ec*RfaH-KOW sample further
809 contained 6 M urea. Due to limited sample stability and poor quality of the initial spectra, *Vc*RfaH

810 (0.3 mM) was [^{2}H, ^{15}N, ^{13}C]-labeled and in an optimized buffer (25 mM Bis-Tris-Propane
811 (pH 6.5), 25 mM Na-Tartrate, 50 mM NaCl, 10 % (v/v) D_2O) and the measurements were
812 conducted at 20 °C. The Cα and CO secondary chemical shift for *Vc*RfaH was calculated as
813 difference between the observed chemical shift and the predicted random coil value (Wishart and
814 Sykes, 1994) using a deuterium correction as given in (Venters et al., 1996). Chemical shift
815 assignments for *Ec*NusG-KOW, *Mt*NusG-KOW and native *Ec*RfaH-KOW were taken from
816 previous studies (Burmann et al., 2012; Mooney et al., 2009; Strauß et al., 2016). The random coil
817 chemical shifts for characterization of the minor species in case of *Vc*RfaH-KOW and hSpt5-KOW
818 were calculated using the Poulsen IDP/IUP random coil chemical shifts calculator tool
819 (https://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS/).
820
821 Distance restraints for the structure calculation of *Vc*RfaH-KOW were obtained from standard ^{13}C-
822 , and ^{15}N-edited 3D NOESY experiments (Sattler et al., 1999) with mixing times of 120 ms.
823 NOESY cross signals were classified according to their intensities and converted to distance
824 restraints with upper limits of 3 Å (strong), 4 Å (medium), 5 Å (weak) and 6 Å (very weak),
825 respectively. Hydrogen bonds were identified from corresponding experiments (see below).
826 Psi/Phi angle restraints were obtained from the geometry dependence of the backbone chemical
827 shifts using TALOS (Cornilescu et al., 1999). The structure calculation was performed with
828 XPLOR-NIH version 2.1.2 using a three-stage simulated annealing protocol with floating
829 assignment of prochiral groups including a conformational database potential (Schwieters et al.,
830 2003). Structures were analyzed with XPLOR-NIH and PROCHECK-NMR (Laskowski et al.,
831 1996).

832 ^{15}N-based CEST experiments were conducted according to (Vallurupalli et al., 2012). All samples
833 contained ≈ 0.7 – 1 mM ^{15}N-labeled protein. For initial CEST experiments, the domains were in
834 20 mM HEPES (pH 7.5), 100 mM NaCl, 10 % (v/v) D_2O and a single CEST B_1 field (ν_1 = 18 -
835 25 Hz) during an exchange period of 500 ms was employed. Proteins showing an exchange peak
836 in their CEST profiles were further studied in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl,
837 1 mM EDTA, 10 % (v/v) D_2O to decrease amide proton-H_2O exchange. CEST experiments were
838 then recorded using two different B_1 fields (ν_1 = 13 Hz/26 Hz) and an exchange period of 500 ms.
839 The B_1 frequencies were calibrated using a 1D approach on an isolated signal (Guenneugues et al.,
840 1999). The CEST traces obtained at 13/26 Hz were fitted simultaneously according to a two-state
841 exchange model using ChemEx (version 0.6.1, (Vallurupalli et al., 2012)). Due to the
842 monodisperse distribution of the resulting k_{ex}/p_B values (**Table 5**), the CEST traces were then fitted
843 globally, yielding a global k_{ex} and p_B value. Only those CEST profiles were included in the global
844 fit that showed a $\Delta\omega > 1$ ppm. ^{13}Cα-CEST experiments were recorded on [^{15}N, ^{13}C]-labeled protein
845 samples using a [^{1}H, ^{13}C] constant-time (ct) HSQC based approach (Bouvignies et al., 2014). To
846 maximize the number of analyzable signals, the proteins were in 20 mM Na-phosphate (pH 6.5),
847 100 mM NaCl, 1 mM EDTA, 99.9 % (v/v) D_2O (pH uncorrected for D_2O). In this case, the
848 chemical shift was referenced *via* 0.5 mM internal DSS. The experiment was performed at a single
849 B_1 field strength (25 Hz) at an exchange period of 500 ms. The CEST traces obtained for [^{1}H, ^{13}C]-
850 *Ec*RfaH-KOW were fitted with ChemEx.
851 NMR-based chemical denaturation experiments of the KOW domains were done by recording
852 [^{1}H, ^{15}N]-HSQC and [^{1}H, ^{13}C]-ctHSQC spectra of 80 µM [^{15}N, ^{13}C]-*Ec*RfaH-KOW in 20 mM Na-
853 phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA, 10 % (v/v) D_2O buffer containing 0 – 8 M urea.
854 The chemical shifts were referenced to 0.5 mM internal DSS.

855 For the NMR-based refolding experiment of *Vc*RfaH under reducing conditions a [^{1}H, ^{15}N]-HSQC
856 spectrum of ^{15}N-*Vc*RfaH in refolding buffer (50 mM Na-phosphate (pH 6.5), 50 mM NaCl, 2 mM
857 DTT) was recorded before the protein was incubated in refolding buffer containing 8 M urea for
858 24 hours. Having recorded another [^{1}H, ^{15}N]-HSQC spectrum urea was removed by stepwise
859 dialysis against 4 l of refolding buffer containing 4 M, 2 M, 1 M, 0.5 M, and 0 M urea, respectively
860 (2-4 hours for the first four steps and overnight for the last step). Finally, a [^{1}H, ^{15}N]-HSQC
861 spectrum of the refolded protein was recorded.
862 Hydrogen bonds were identified from 2D or 3D long range (LR) TROSY based HNCO
863 experiments as previously described (Cordier et al., 2008). All samples contained [^{15}N, ^{13}C]-
864 labeled proteins at 0.7 – 1 mM in 20 mM Na-phosphate (pH 6.5), 100 mM NaCl, 1 mM EDTA,
865 10 % (v/v) D_2O.
866
867
868 **CD spectroscopy**
869 CD data were collected at a Jasco J-1100 spectrometer (Jasco Deutschland GmbH, Pfungstadt,
870 Germany), using quartz cuvettes (Hellma GmbH & Co. KG, Müllheim, Germany). CD spectra
871 were normalized (**Equation 1**) to obtain the mean residue-weighted ellipticity (Θ_{MRW}):

$$\Theta_{MRW} = \frac{100 \cdot \theta}{N \cdot c \cdot d} \qquad (1)$$

872 θ is the ellipticity in mdeg, *N* the number of amino acids, *c* the protein concentration in mM, and
873 *d* the pathlength of the cuvette in cm.
874 Thermal unfolding and refolding curves were obtained by measuring the CD signal of 15 µM
875 (≈ 0.1 mg/ml) protein buffered by either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate
876 (pH 4.0), respectively, in a 1 cm quartz cuvette upon heating to 95 °C and subsequently re-cooling

877 to the initial temperature. The scan speed was 1 °C/min, the dwell time 1 min, and the integration
878 time 4 s. Checking the reversibility of thermal unfolding and determination of the wavelength used
879 for temperature transition curves was done by recording far-UV CD spectra at 25 °C, then 95 °C,
880 and after subsequent re-cooling to 25 °C in a 2 mm pathlength cuvette using 25 µM protein
881 solutions in either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0). The wavelength to
882 follow a thermal transition corresponds to the wavelength > 215 nm with the largest difference in
883 the CD signal between folded and unfolded state and was chosen for each transition individually.
884 Using wavelengths < 215 nm led to noisy signals at high temperatures and resulted in non-
885 interpretable data.
886
887
888 Changes in ellipticity (θ) upon thermal unfolding were analyzed with a two-state model using Fit-
889 o-Mat version 0.752 (Möglich, 2018) to obtain the melting temperature (T_m) and enthalpy change
890 at T_m ($\Delta H_u(T_m)$) of the transition (both fit parameters) (**Equation 2**):
891
892 $$\theta = f_N \cdot \left(y_N + m_N \cdot (T - T_m)\right) + (1 - f_N) \cdot \left(y_U + m_U \cdot (T - T_m)\right) \quad (2)$$
893 with T being the absolute temperature in K, y_N and y_U the y-intercepts, and m_N and m_U the slopes
894 of the N- and U-state baselines, respectively. f_N is the fraction of folded molecules, which is related
895 to the equilibrium constant K_u according to **Equation 3**:
896
$$f_N = \frac{1}{1 + K_u} \quad (3)$$
897

898 Finally, K_u is related to the change in Gibb's free energy of the unfolding reaction (ΔG_u) and

899 $\Delta H_u(T_m)$ by **Equation 4**:

900

$$K_u = e^{-\Delta G_u/(RT)} \text{with} \Delta G_u = \Delta H_u(T_m) - \frac{T}{T_m} \cdot \Delta H_u(T_m) \quad (4)$$

R is the ideal gas constant.

CD-based chemical equilibrium unfolding experiments were performed at 25 °C. Urea (BioScience Grade; ≈ 10 M) and GdmCl (≈ 8 M; both from Carl Roth GmbH & Co. KG, Karlsruhe, Germany) stock solutions were prepared according to (Pace et al., 1990). Far-UV CD unfolding experiments were conducted using a 1 mm cuvette. All points of the unfolding curves were obtained from individual samples, each containing 40 - 60 µM (≈ 0.25 – 0.4 mg/ml) protein in either 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0), respectively. All samples were equilibrated over-night. The denaturant concentration of each sample was determined refractrometrically after CD data acquisition.

As for the thermal transitions, the wavelength to follow a chemical denaturation corresponds to the wavelength > 215 nm with the largest difference in the CD signal between folded and unfolded state and was chosen for each transition individually (wavelengths < 215 nm led to noisy signals and non-interpretable data at high denaturant concentrations).

Unfolding curves that indicate a two-state transition were analyzed using the linear extrapolation method (Santoro and Bolen, 1988) with Fit-o-Mat version 0.752 (Möglich, 2018) to obtain $\Delta G_u(H_2O)$ and the m-value (**Equation 5**):

$$S = f_N \cdot (y_N + m_N \cdot [\text{denat}]) + (1 - f_N) \cdot (y_U + m_U \cdot [\text{denat}]) \quad (5)$$

S is the signal derived from far-UV CD spectroscopy (i.e. the Θ_{MRW}-value), intrinsic Trp fluorescence (for *Vc*RfaH-CTD), or the normalized peak volumes of the [^{1}H, ^{13}C]-ctHSQC

923 major/minor species signals for *Ec*RfaH-KOW residue S139, respectively. [denat] is the
924 denaturant (i.e. urea or GdmCl) concentration in M, y_N and y_U are the y-intercepts, and m_N and m_U,
925 the slopes of the N- and U-state baselines, respectively. f_N is given by **Equation 3**. In this case, K_u
926 is defined as (**Equation 6**):
927

$$K_u = e^{-\Delta G/(RT)} \quad \text{with} \quad \Delta G = \Delta H(H_2O) - m \cdot [\text{denat}] \tag{6}$$

928

929 Finally, the $[\text{denat}]_{1/2}$ value is obtained by (**Equation 7**):

930 $$[\text{denat}]_{\frac{1}{2}} = \frac{\Delta G(H_2O)}{m} \tag{7}$$

931 Near-UV CD unfolding experiments of *Ec*RfaH-KOW were conducted using a 1 cm quartz cuvette
932 and 0.5 mM protein in 10 mM K-phosphate (pH 7.0). As the exchange between folded and
933 unfolded state is reasonably fast ($k_{ex} \approx 15\ s^{-1}$ at 0 M urea/GdmCl), all points were obtained from a
934 titration of the initial denaturant-free protein sample with a 10 M urea or 8 M GdmCl solution in
935 10 mM K-phosphate (pH 7.0). The sample was then incubated for 5 min at 25 °C to reach
936 equilibrium. Curves were smoothed mathematically using a Savitzky-Golay filter.
937 To probe reversibility of chemical unfolding and validate incubation times used to reach
938 equilibrium, proteins were dialyzed against 20 mM NH_4HCO_3 (pH 7.0) buffer, shock-frozen,
939 lyophilized and subsequently solved in 10 mM K-phosphate (pH 7.0) or 10 mM K-acetate (pH 4.0)
940 with or without 10 M urea/8 M GdmCl, respectively. CD samples containing the identical
941 denaturant concentration (1–2 samples in pre-transition region, 1 at $[\text{denaturant}]_{1/2}$, 1 in post-
942 transition region) were then prepared from the native or unfolded proteins. All samples were
943 equilibrated over-night; far-UV CD spectra were then recorded using a 1 mm quartz cuvette.
944

945 **Fluorescence spectroscopy**

946 Fluorescence spectra were recorded at 25 °C using a Peltier-controlled Fluorolog-3 fluorimeter
947 (Horiba Europe GmbH, Oberursel, Germany) equipped with a 1 cm quartz cuvette (Hellma GmbH
948 & Co. KG, Müllheim, Germany). Samples for chemical denaturation of *Vc*RfaH-KOW contained
949 ≈ 11 μM protein and were prepared as described for the far-UV CD samples. The *Vc*RfaH-KOW
950 Trp residue was excited at 295 nm; emission spectra were then recorded from 300 to 400 nm with
951 slit widths between 2.65/2.65 nm and 2.8/2.8 nm (excitation/emission) and an integration time of
952 0.2 s. Analysis of the resulting denaturation curve was performed as described for CD data.
953 ANS (Sigma/Merck KGaA, Darmstadt, Germany) interaction experiments were conducted by
954 preparing a urea denaturation series of *Ec*RfaH-KOW (final concentration: 5 μM) as described for
955 the CD-based unfolding experiments, equilibrating over-night and adding ANS at a
956 fluorophore:protein ratio of 100:1. Fluorescence spectra were then recorded from 410 to 650 nm
957 following excitation at 395 nm with slit widths of 2.6/2.6 nm (excitation/emission) and 0.1 s
958 integration time. A control experiment was conducted with identical experiment and instrument
959 setup, respectively, but samples lacking protein. The obtained fluorescence at a given wavelength
960 was then plotted against the urea concentration of the respective sample.

961

962 **DSC**

963 The KOW domains were in either 10 mM K-acetate (pH 4.0; hSpt5-KOW5) or 10 mM K-
964 phosphate (pH 7.0; all other domains), respectively. Given a lack of Trp residues in most domains,
965 the protein concentration was determined *via* absorption at 205 nm using the molar extinction
966 coefficient (ε_{205}) as calculated by the Protein Calculator tool (Anthis and Clore, 2013).

967 Initial DSC experiments were carried out on a MicroCal VP-DSC instrument (MicroCal/Malvern
968 Panalytical, Malvern, UK; active volume: 509 µl). The samples were vacuum degassed at room
969 temperature just before the measurements. Prior to the protein-buffer scans, several buffer-buffer
970 scans were performed. All thermograms were recorded at a scan rate of 1.5 K/min under an excess
971 pressure of 30 psi in passive feed-back mode from ≈ 10 °C to 110 °C or 130 °C (*Mj*Spt5-KOW5),
972 respectively. The unfolding was calometrically reversible for *Ec*NusG-KOW, *Mt*NusG-KOW,
973 *Mj*Spt5-KOW and *Ec*RfaH-KOW (data not shown). hSpt5-KOW5 aggregated at pH 7.0 upon
974 unfolding at all tested concentrations, whereas *Vc*RfaH-KOW aggregated at concentrations >
975 0.2 mg/ml.
976 We repeated the measurements for all proteins but *Mt*NusG-KOW using a MicroCal VP-Capillary
977 DSC instrument (Malvern Panalytical, Malvern, UK; active volume 137 µl). The thermograms
978 were obtained at a heating rate of 1.5 K/min with excess pressure (30 psi) and at mid gain feed-
979 back mode. Buffer-buffer runs were done prior to the protein measurements. Thermograms were
980 recorded from ≈ 5 °C to 130 °C. The protein concentration was 0.2 – 1 mg/ml for *Ec*NusG-KOW,
981 0.25 – 1 mg/ml for *Mj*Spt5-KOW, 0.15 – 0.25 mg/ml for hSpt5-KOW5, 0.2 – 1 mg/ml for *Ec*RfaH-
982 KOW and 0.1 – 0.15 mg/ml for *Vc*RfaH-KOW. The measurement for hSpt5-KOW5 was carried
983 out with 10 mM K-acetate (pH 4.0), all other KOW domains were in 10 mM K-phosphate
984 (pH 7.0).
985 The obtained raw DSC data (VP-DSC data for *Mt*NusG-KOW, VP-Capillary DSC data for all
986 other KOW domains) was scan-rate normalized, the corresponding buffer-buffer baseline was
987 subtracted, and the thermograms were then normalized to one mol of protein. To extract the
988 thermodynamic parameters, the data was fitted to a two-state unfolding model including a
989 temperature-dependent change in heat capacity from native to unfolded state (Viguera et al., 1994).

990 The temperature dependence of the native state heat capacity ($C_{p,0}$) is assumed to be linear
991 (**Equation 8**; note that $C_{p,0}$ contains an instrument-specific offset), whereas the difference in heat
992 capacity to the unfolded state ($\Delta C_{p,u}(T)$) is approximated by a parabolic function (**Equation 9**):
993

$$C_{p,0} = a_0 + b_0 \cdot T \tag{8}$$

$$\Delta C_{p,u}(T) = a + b \cdot T + c \cdot T^2 \tag{9}$$

994 The value for the pre-factor of the quadratic term, *c,* was obtained by calculating the theoretical
995 partial molar heat capacity, $C_p(T)$, of the unfolded state for each of the six protein domains at 5,
996 25, 50, 75, 100 and 125 °C, respectively, according to Makhatadze and Privalov (Makhatadze and
997 Privalov, 1990). Then, the values for $C_p(T)$ were plotted over the temperature and a parabolic
998 function was fitted, yielding *c*.
999 The concentration-normalized heat capacity (C_p) then is the sum of $C_{p,0}$, the change of the
1000 "internal" heat capacity that depends on the fraction of the protein in the folded and unfolded state
1001 (i.e. the equilibrium constant K_u), δC_p^{int}, and the excess heat absorption of the unfolding reaction
1002 δC_p^{exc} (**Equation 10**):

$$C_p = C_{p,0} + \delta C_p^{int} + \delta C_p^{exc} \tag{10}$$

1003

1004 With δC_p^{int} and δC_p^{exc} given in **Equation 11**:
1005

$$\delta C_p^{int} = \Delta C_{p,u} \cdot \frac{K_u}{1 + K_u} \text{ and } \delta C_p^{exc} = \frac{\left(\Delta H_u(T)\right)^2}{RT^2} \cdot \frac{K_u}{(1 + K_u)^2} \tag{11}$$

1006

1007 K_u is related to the change in Gibbs energy of the unfolding reaction ($\Delta G_u(T)$) by (**Equation 12**):
1008

$$K_{\mathrm{u}} = e^{-\Delta G_{\mathrm{u}}(T)/(RT)} \text{with} \Delta G_{\mathrm{u}}(T) = \Delta H_{\mathrm{u}}(T) - T \cdot \Delta S_{\mathrm{u}}(T) \tag{12}$$

The temperature-dependent enthalpy and entropy change ($\Delta H_{\mathrm{u}}(T)$, and $\Delta S_{\mathrm{u}}(T)$, respectively) are given by **Equations 13** and **14**:

$$\Delta H_{\mathrm{u}}(T) = \Delta H_{\mathrm{u}}(T_{\mathrm{m}}) + a \cdot (T - T_{\mathrm{m}}) + \frac{b}{2} \cdot (T^2 - T_{\mathrm{m}}^2) + \frac{c}{3} \cdot (T^3 - T_{\mathrm{m}}^3) \tag{13}$$

$$\Delta S_{\mathrm{u}}(T) = \frac{\Delta H_u(T_m)}{T_{\mathrm{m}}} + a \cdot ln\left(\frac{T}{T_{\mathrm{m}}}\right) + b \cdot (T - T_{\mathrm{m}}) + \frac{c}{2} \cdot (T^2 - T_{\mathrm{m}}^2) \tag{14}$$

During fitting of C_{p}, parameters a_0, b_0, a, b, T_{m} and $\Delta H_{\mathrm{u}}(T_{\mathrm{m}})$ were allowed to float, while c was kept constant.

Acknowledgments

We thank I. Artsimovitch for providing pVS13 and F. Werner for providing pGEX-2TK_*Mj*Spt5-KOW. Our sincere thanks go to F.X. Schmid for detailed discussions and valuable comments on the manuscript. We also thank A. Matagne and Birgitta M Wöhrl for helpful discussions, C. M. Johnson for performing DSC measurements, A. Häussermann for carrying out the NMR-based refolding experiment of *Vc*RfaH, and A. Hager for technical support, as well as the Northern Bavarian NMR Centre (NBNC) for providing access to NMR spectrometers. The project was supported by the German Research Foundation grant Ro617/21-1 (to P. Rösch) and COST action CA 15126 ARBRE-MOBIEU (to PKZ, TD, SHK).

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: Coordinates for *Vc*RfaH-KOW have been deposited to the Protein Databank (ID: 6TF4). Chemical shifts have been deposited in the Biological Magnetic Resonance Databank under the following accession numbers: #28039 (hSpt5-KOW5), #28040 (*Mj*Spt5-KOW), #28041 (*Vc*RfaH) and #34450 (*Vc*RfaH-CTD). All other data needed to evaluate the conclusions in the paper are present in the paper, in the Supplemental Material, and/or as source data. Additional data related to this paper and plasmids may be requested from the authors (the latter require a material transfer agreement).

1044 **References**

1045 Anthis NJ, Clore GM. 2013. Sequence-specific determination of protein and peptide
1046 concentrations by absorbance at 205 nm. *Protein Science* **22**:851–858. doi:10.1002/pro.2253
1047 Artsimovitch I, Knauer SH. 2019. Ancient Transcription Factors in the News. *MBio* **10**.
1048 doi:10.1128/mBio.01547-18
1049 Artsimovitch I, Landick R. 2002. The Transcriptional Regulator RfaH Stimulates RNA Chain
1050 Synthesis after Recruitment to Elongation Complexes by the Exposed Nontemplate DNA Strand.
1051 *Cell* **109**:193–203. doi: 10.1016/s0092-8674(02)00724-9
1052 Balasco N, Barone D, Vitagliano L. 2015. Structural conversion of the transformer protein RfaH:
1053 new insights derived from protein structure prediction and molecular dynamics simulations. *J*
1054 *Biomol Struct Dyn* **33**:2173–2179. doi:10.1080/07391102.2014.994188
1055 Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. 2009. Functional
1056 specialization of transcription elongation factors. *EMBO J* **28**:112–122.
1057 doi:10.1038/emboj.2008.268
1058 Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG,
1059 Artsimovitch I. 2007. Structural Basis for Converting a General Transcription Factor into an
1060 Operon-Specific Virulence Regulator. *Mol Cell* **26**:117–129. doi:10.1016/j.molcel.2007.02.021
1061 Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. 2005. Complete
1062 Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew*
1063 *Chem Int Ed* **44**:3089–3092. doi:10.1002/anie.200461794
1064 Bernecky C, Plitzko JM, Cramer P. 2017. Structure of a transcribing RNA polymerase II-DSIF
1065 complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* **24**:809–815.
1066 doi:10.1038/nsmb.3465

1067 Bernhardt NA, Hansmann UHE. 2018. Multifunnel Landscape of the Fold-Switching Protein
1068 RfaH-CTD. *J Phys Chem B* **122**:1600–1607. doi:10.1021/acs.jpcb.7b11352
1069 Bouvignies G, Vallurupalli P, Kay LE. 2014. Visualizing Side Chains of Invisible Protein
1070 Conformers by Solution NMR. *J Mol Biol* **426**:763–774. doi:10.1016/j.jmb.2013.10.041
1071 Bryan PN, Orban J. 2010. Proteins that switch folds. *Curr Opin Struct Biol* **20**:482–488.
1072 doi:10.1016/j.sbi.2010.06.002
1073 Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R,
1074 Artsimovitch I, Rösch P. 2012. An α Helix to β Barrel Domain Switch Transforms the
1075 Transcription Factor RfaH into a Translation Factor. *Cell* **150**:291–303.
1076 doi:10.1016/j.cell.2012.05.042
1077 Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P. 2010. A
1078 NusE:NusG Complex Links Transcription and Translation. *Science* **328**:501–504.
1079 doi:10.1126/science.1184953
1080 Carter HD, Svetlov V, Artsimovitch I. 2004. Highly Divergent RfaH Orthologs from Pathogenic
1081 Proteobacteria Can Substitute for *Escherichia coli* RfaH both In Vivo and In Vitro. *JBacteriol*
1082 **186**:2829–2840. doi:10.1128/JB.186.9.2829-2840.2004
1083 Cordier F, Nisius L, Dingley AJ, Grzesiek S. 2008. Direct detection of N−H···O=C hydrogen
1084 bonds in biomolecules by NMR spectroscopy. *Nat Protoc* **3**:235–241.
1085 doi:10.1038/nprot.2007.498
1086 Cornilescu G, Delaglio F, Bax A. 1999. Protein backbone angle restraints from searching a
1087 database for chemical shift and sequence homology. *J Biomol NMR* **13**:289–302.
1088 doi:10.1023/A:1008392405740
1089 de Marco A, Berrow N, Lebendiker M, Garcia-Alai M, Knauer SH, Lopez-Mendez B, Matagne

1090 A, Parret A, Remans K, Uebel S, Raynal B. 2021. Quality control of protein reagents for the
1091 improvement of research data reproducibility. *Nat Commun* **12**:2795. doi:10.1038/s41467-021-
1092 23167-z
1093 Dishman AF, Tyler RC, Fox JC, Kleist AB, Prehoda KE, Babu MM, Peterson FC, Volkman BF.
1094 2021. Evolution of fold switching in a metamorphic protein. *Science* **371**:86–90.
1095 doi:10.1126/science.abd8700
1096 Dishman AF, Volkman BF. 2018. Unfolding the Mysteries of Protein Metamorphosis. *ACS*
1097 *Chem Biol* **13**:1438–1446. doi:10.1021/acschembio.8b00276
1098 Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure
1099 prediction server. *Nucleic Acids Res* **43**:W389–W394. doi:10.1093/nar/gkv332
1100 Epstein CJ, Goldberger RF, Anfinsen CB. 1963. The Genetic Control of Tertiary Protein
1101 Structure: Studies With Model Systems. *Cold Spring Harb Symp Quant Biol* **28**:439–449.
1102 doi:10.1101/SQB.1963.028.01.060
1103 Farrow NA, Zhang O, Forman-Kay JD, Kay LE. 1995. Comparison of the Backbone Dynamics
1104 of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer.
1105 *Biochemistry* **34**:868–878. doi:10.1021/bi00003a021
1106 Galaz-Davison P, Molina JA, Silletti S, Komives EA, Knauer SH, Artsimovitch I, Ramírez-
1107 Sarmiento CA. 2020. Differential Local Stability Governs the Metamorphic Fold Switch of
1108 Bacterial Virulence Factor RfaH. *Biophys J* **118**:96–104. doi:10.1016/j.bpj.2019.11.014
1109 Gc JB, Bhandari YR, Gerstman BS, Chapagain PP. 2014. Molecular Dynamics Investigations of
1110 the α-Helix to β-Barrel Conformational Transformation in the RfaH Transcription Factor. *J Phys*
1111 *Chem B* **118**:5101–5108. doi:10.1021/jp502193v
1112 Green MR, Sambrook J, Sambrook J. 2012. Molecular cloning: a laboratory manual, 4th ed. ed.

1113 Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

1114 Grzesiek S, Cordier F, Jaravine V, Barfield M. 2004. Insights into biomolecular hydrogen bonds

1115 from hydrogen bond scalar couplings. *Prog Nucl Magn Reson Spectrosc* **45**:275–300.

1116 doi:10.1016/j.pnmrs.2004.08.001

1117 Guenneugues M, Berthault P, Desvaux H. 1999. A Method for Determining B_1 Field

1118 Inhomogeneity. Are the Biases Assumed in Heteronuclear Relaxation Experiments Usually

1119 Underestimated? *J Magn Reson* **136**:118–126. doi:10.1006/jmre.1998.1590

1120 Gupta R, Yadav S, Ahmad F. 1996. Protein Stability: Urea-Induced versus Guanidine-Induced

1121 Unfolding of Metmyoglobin. *Biochemistry* **35**:11925–11930. doi:10.1021/bi961079g

1122 Hirtreiter A, Damsma GE, Cheung ACM, Klose D, Grohmann D, Vojnic E, Martin ACR,

1123 Cramer P, Werner F. 2010. Spt4/5 stimulates transcription elongation through the RNA

1124 polymerase clamp coiled-coil motif. *Nucleic Acids Res* **38**:4040–4051. doi:10.1093/nar/gkq135

1125 Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. 2012. Application of iterative

1126 soft thresholding for fast reconstruction of NMR data non-uniformly sampled with

1127 multidimensional Poisson Gap scheduling. *J Biomol NMR* **52**:315–327. doi:10.1007/s10858-012-

1128 9611-z

1129 Jeffery CJ. 2014. An introduction to protein moonlighting. *Biochem Soc Trans* **42**:1679–1683.

1130 doi:10.1042/BST20140226

1131 Jeffery CJ. 1999. Moonlighting proteins. *Trends Biochem Sci* **24**:8–11. doi:10.1016/s0968-

1132 0004(98)01335-8

1133 Joseph JA, Chakraborty D, Wales DJ. 2019. Energy Landscape for Fold-Switching in Regulatory

1134 Protein RfaH. *J Chem Theory Comput* **15**:731–742. doi:10.1021/acs.jctc.8b00912

1135 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates

1136 R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-
1137 Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M,
1138 Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW,
1139 Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with
1140 AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2
1141 Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, Artsimovitch I, Landick R, Darst
1142 SA. 2018. Structural Basis for Transcript Elongation Control by NusG Family Universal
1143 Regulators. *Cell* **173**:1650-1662.e14. doi:10.1016/j.cell.2018.05.017
1144 Kim AK, Porter LL. 2021. Functional and Regulatory Roles of Fold-Switching Proteins.
1145 *Structure* **29**:6–14. doi:10.1016/j.str.2020.10.006
1146 Kim C, Choi J, Lee SJ, Welsh WJ, Yoon S. 2009. NetCSSP: web application for predicting
1147 chameleon sequences and amyloid fibril formation. *Nucleic Acids Res* **37**:W469–W473.
1148 doi:10.1093/nar/gkp351
1149 Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. 2011. RNA polymerase and
1150 transcription elongation factor Spt4/5 complex structure. *Proc Natl Acad Sci USA* **108**:546–550.
1151 doi:10.1073/pnas.1013828108
1152 Knauer SH, Artsimovitch I, Rösch P. 2012. Transformer proteins. *Cell Cycle* **11**:4289–4290.
1153 doi:10.4161/cc.22468
1154 Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. 1996. AQUA and
1155 PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J*
1156 *Biomol NMR* **8**:477–486.
1157 Lella M, Mahalakshmi R. 2017. Metamorphic Proteins: Emergence of Dual Protein Folds from
1158 One Primary Sequence. *Biochemistry* **56**:2971–2984. doi:10.1021/acs.biochem.7b00375

1159 Lescop E, Schanda P, Brutscher B. 2007. A set of BEST triple-resonance experiments for time-
1160 optimized protein resonance assignment. *J Magn Reson* **187**:163–169.
1161 doi:10.1016/j.jmr.2007.04.002
1162 Li S, Xiong B, Xu Y, Lu T, Luo X, Luo C, Shen J, Chen K, Zheng M, Jiang H. 2014.
1163 Mechanism of the All-α to All-β Conformational Transition of RfaH-CTD: Molecular Dynamics
1164 Simulation and Markov State Model. *J Chem Theory Comput* **10**:2255–2264.
1165 doi:10.1021/ct5002279
1166 Makhatadze GI. 1999. Thermodynamics of Protein Interactions with Urea and Guanidinium
1167 Hydrochloride. *J Phys Chem B* **103**:4781–4785. doi:10.1021/jp990413q
1168 Makhatadze GI, Privalov PL. 1990. Heat Capacity of Proteins I. Partial Molar Heat Capacity of
1169 Individual Amino Acid Residues in Aqueous Solution: Hydration Effect. *J Mol Biol* **213**:375–
1170 384. doi:10.1016/S0022-2836(05)80197-4
1171 Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L,
1172 Horng J-C, Bona D, Miller EJ, Vallée-Bélisle A, Main ERG, Bemporad F, Qiu L, Teilum K, Vu
1173 N-D, Edwards AM, Ruczinski I, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y,
1174 Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE,
1175 Sosnick TR, Marqusee S, Davidson AR, Plaxco KW. 2005. Protein folding: Defining a
1176 "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins.
1177 *Protein Sci* **14**:602–616. doi:10.1110/ps.041205405
1178 Meyer O, Schlegel HG. 1983. Biology of Aerobic Carbon Monoxide-Oxidizing Bacteria. *Annu*
1179 *Rev Microbiol* **37**:277–310. doi:10.1146/annurev.mi.37.100183.001425
1180 Meyer PA, Li S, Zhang M, Yamada K, Takagi Y, Hartzog GA, Fu J. 2015. Structures and
1181 Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. *Mol Cell Biol*

35:3354–3369. doi:10.1128/MCB.00520-15

Möglich A. 2018. An Open-Source, Cross-Platform Resource for Nonlinear Least-Squares Curve Fitting. *J Chem Educ* **95**:2273–2278. doi:10.1021/acs.jchemed.8b00649

Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R. 2009. Two Structurally Independent Domains of *E. coli* NusG Create Regulatory Plasticity *via* Distinct Interactions with RNA Polymerase and Regulators. *J Mol Biol* **391**:341–358. doi:10.1016/j.jmb.2009.05.078

Murzin AG. 2008. BIOCHEMISTRY: Metamorphic Proteins. *Science* **320**:1725–1726. doi:10.1126/science.1158868

Pace CN, Shirley BA, Thomson JA, Creighton TE. 1990. Measuring the conformational stability of a proteinProtein Structure: A Practical Approach. pp. 311–330.

Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Nat Acad Sci USA* **94**:12366–12371. doi:10.1073/pnas.94.23.12366

Porter LL, Kim AK, Rimal S, Looger LL, Majumdar A, Mensh BD, Starich MR, Strub M-P. 2022. Many dissimilar NusG protein domains switch between α-helix and β-sheet folds. *Nat Commun* **13**:3802. doi:10.1038/s41467-022-31532-9

Porter LL, Looger LL. 2018. Extant fold-switching proteins are widespread. *Proc Natl Acad Sci USA* **115**:5968–5973. doi:10.1073/pnas.1800168115

Ramírez-Sarmiento CA, Noel JK, Valenzuela SL, Artsimovitch I. 2015. Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. *PLoS Comput Biol* **11**:e1004379. doi:10.1371/journal.pcbi.1004379

Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. 1998. TROSY in triple-resonance

experiments: New perspectives for sequential NMR assignment of large proteins. *Proc Nat Acad Sci* **95**:13585–13590. doi:10.1073/pnas.95.23.13585

Santoro MM, Bolen DW. 1988. Unfolding Free Energy Changes Determined by the Linear Extrapolation Method. 1. Unfolding of Phenylmethanesulfonyl α-Chymotrypsin Using Different Denaturants. *Biochemistry* **27**:8063–8068. doi:10.1021/bi00421a014

Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* **34**:93–158.

Schanda P, Van Melckebeke H, Brutscher B. 2006. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J Am Chem Soc* **128**:9042–9043. doi:10.1021/ja062025p

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. 2003. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**:65–73.

Seifi B, Wallin S. 2021. The C-terminal domain of transcription factor RfaH: Folding, fold switching and energy landscape. *Biopolymers* e23420. doi:10.1002/bip.23420

Sharma D, Rajarathnam K. 2000. ^{13}C NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular NMR* **18**:165–171. doi:10.1023/A:1008398416292

Shi D, Svetlov D, Abagyan R, Artsimovitch I. 2017. Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Res* **45**:8835–8843. doi:10.1093/nar/gkx523

Strauß M, Schweimer K, Burmann BM, Richter A, Güttler S, Wöhrl BM, Rösch P. 2016. The two domains of *Mycobacterium tuberculosis* NusG protein are dynamically independent. *J Biomol Struct Dyn* **34**:352–361. doi:10.1080/07391102.2015.1031700

Tomar SK, Knauer SH, Nandymazumdar M, Rösch P, Artsimovitch I. 2013. Interdomain

1228 contacts control folding of transcription factor RfaH. *Nucleic Acids Res* **41**:10077–10085.
1229 doi:10.1093/nar/gkt779
1230 Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A,
1231 Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M,
1232 Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S,
1233 Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper
1234 J, Hassabis D. 2021. Highly accurate protein structure prediction for the human proteome.
1235 *Nature* **596**:590–596. doi:10.1038/s41586-021-03828-1
1236 Tyler RC, Murray NJ, Peterson FC, Volkman BF. 2011. Native-state interconversion of a
1237 metamorphic protein requires global unfolding. *Biochemistry* **50**:7077–7079.
1238 doi:10.1021/bi200750k
1239 Vallurupalli P, Bouvignies G, Kay LE. 2012. Studying "Invisible" Excited Protein States in Slow
1240 Exchange with a Major State Conformation. *J Am Chem Soc* **134**:8148–8161.
1241 doi:10.1021/ja3001419
1242 van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M,
1243 Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P,
1244 Uversky VN, Wright PE, Babu MM. 2014. Classification of intrinsically disordered regions and
1245 proteins. *Chem Rev* **114**:6589–6631. doi:10.1021/cr400525m
1246 Vassylyeva MN, Svetlov V, Klyuyev S, Devedjiev YD, Artsimovitch I, Vassylyev DG. 2006.
1247 Crystallization and preliminary crystallographic analysis of the transcriptional regulator RfaH
1248 from *Escherichia coli* and its complex with ops DNA. *Acta Crystallogr F Struct Biol Cryst*
1249 *Commun* **62**:1027–1030. doi:10.1107/S174430910603658X
1250 Venters RA, Farmer II BT, Fierke CA, Spicer LD. 1996. Characterizing the Use of

1251 Perdeuteration in NMR Studies of Large Proteins: ^{13}C, ^{15}N and ^{1}H Assignments of Human
1252 Carbonic Anhydrase II. *J Mol Biol* **264**:1101–1116. doi:10.1006/jmbi.1996.0699
1253 Viguera AR, Martinez JC, Filimonov VV, Mateo PL, Serrano L. 1994. Thermodynamic and
1254 Kinetic Analysis of the SH3 Domain of Spectrin Shows a Two-State Folding Transition.
1255 *Biochemistry* **33**:2142–2150. doi:10.1021/bi00174a022
1256 Werner F. 2012. A Nexus for Gene Expression-Molecular Mechanisms of Spt5 and NusG in the
1257 Three Domains of Life. *J Mol Biol* **417**:13–27. doi:10.1016/j.jmb.2012.01.031
1258 Wishart DS, Sykes BD. 1994. [12] Chemical shifts as a tool for structure determinationMethods
1259 in Enzymology. Elsevier. pp. 363–392. doi:10.1016/S0076-6879(94)39014-2
1260 Xiong L, Liu Z. 2015. Molecular dynamics study on folding and allostery in RfaH. *Proteins*
1261 **83**:1582–1592. doi:10.1002/prot.24839
1262 Xun S, Jiang F, Wu Y-D. 2016. Intrinsically disordered regions stabilize the helical form of the
1263 C-terminal domain of RfaH: A molecular dynamics study. *Bioorg Med Chem* **24**:4970–4977.
1264 doi:10.1016/j.bmc.2016.08.012
1265 Zamora-Carreras H, Maestro B, Sanz JM, Jiménez MA. 2020. Turncoat Polypeptides: We Adapt
1266 to Our Environment. *Chem Bio Chem* **21**:432–441. doi:10.1002/cbic.201900446
1267 Zuber PK, Hahn L, Reinl A, Schweimer K, Knauer SH, Gottesman ME, Rösch P, Wöhrl BM.
1268 2018. Structure and nucleic acid binding properties of KOW domains 4 and 6-7 of human
1269 transcription elongation factor DSIF. *Sci Rep* **8**:11660. doi:10.1038/s41598-018-30042-3
1270 Zuber PK, Schweimer K, Rösch P, Artsimovitch I, Knauer SH. 2019. Reversible fold-switching
1271 controls the functional cycle of the antitermination factor RfaH. *Nat Commun* **10**:702.
1272 doi:10.1038/s41467-019-08567-6
1273

1274 **Table 1.** Solution structure statistics for *Vc*RfaH-KOW.

Experimental derived restraints		
Distance restraints		
	NOEs unique (total)	630 (734)
	intraresidual	59
	sequential	187
	medium range	89
	long range	295
	hydrogen bonds	2 · 18
Dihedral restraints		76
Restraint violation		
Average distance restraint violation (Å)		0.002584 ± 0.000700
Maximum distance restraint violation (Å)		0.12
Average dihedral restraint violation (°)		0.0654 ± 0.0265
Maximum dihedral restraint violation (°)		0.71
Deviation from ideal geometry		
Bond length (Å)		0.000544 ± 0.000039
Bond angle (Å)		0.1096 ± 0.0056
Coordinate precision *,†		
Backbone heavy atoms (Å)		0.32
All heavy atoms (Å)		0.90
Ramachandran plot statistics ‡ (%)		91.8 / 7.9 / 0.2 / 0.1

1275 * The precision of the coordinates is defined as the average atomic root mean square difference between the accepted simulated annealing
1276 structures and the corresponding mean structure calculated for the given sequence region.

1277 † calculated for residues 116 - 165

1278 ‡ Ramachandran plot statistics are determined by PROCHECK and noted by most favored/ additionally allowed/generously
1279 allowed/disallowed.

1280
1281

1282 **Table 2.** Selected thermodynamic parameters of the six KOW domains. The values were derived

1283 from thermal denaturations monitored by DSC and CD spectroscopy. Standard deviations result

1284 from data fitting.

1285

Parameter	*Ec*NusG-KOW	*Mt*NusG-KOW	*Mj*Spt5-KOW	hSpt5-KOW5	*Ec*RfaH-KOW	*Vc*RfaH-KOW
T_m (°C) pH 7 / pH 4						
CD	- § / -	76.6 ± 0.874 / -	- †/ - †	- / 60.5 ± 0.771	50.3 ± 0.388 / -	65.2 ± 1.78 / -
DSC	87.0 ± 0.0485 /-	77.0 ± 0.0885 /-	111 ± 0.0326 / -	- / 58.0 ± 0.162	47.3 ± 0.143 / -	70.2 ± 0.379 / -
ΔH_u (T_m) (kJ/mol) pH 7 / pH 4						
CD	- § / -	193 ± 11.3 / -	- †/ - †	- / 140 ± 12.4	121 ± 5.15 / -	162 ± 2.91 / -
DSC	222 ± 0.339 / -	192 ± 0.417 / -	293 ± 0.345 / -	- / 117 ± 0.735	129 ± 0.432 / -	169 ± 1.56 / -
ΔC_p (T_m) (kJ/(K mol)) pH 7 / pH 4	0.800 / -	0.346 / -	- §/ -	- / 2.27	2.18 / -	0.148 / -

1286

1287 † No denaturation.

1288 § Data was not fitted due to the lack of the baseline of the unfolded state.

1289

1290 **Table 3.** Thermodynamic parameters of the six KOW domains. The values were derived from
1291 chemical denaturations monitored by CD spectroscopy as well as fluorescence spectroscopy where
1292 indicated. Standard deviations result from data fitting.

Parameter	*Ec*NusG-KOW	*Mt*NusG-KOW	*Mj*Spt5-KOW	hSpt5-KOW5	*Ec*RfaH-KOW	*Vc*RfaH-KOW
$\Delta G_u(H_2O)$ (25 °C) (kJ/mol)						
urea, pH 4	19.8 ± 2.21	22.4 ± 3.46	- ‡	6.24 ± 4.42	- (native state aggregation)	10.8 ± 1.66 (10.8 ± 0.90)*
urea, pH 7	27.7 ± 4.21	26.4 ± 6.16	- ‡	14.3 ± 2.90	three-state	14.0 ± 1.74 (13.9 ± 0.61)*
GdmCl, pH 7	11.7 ± 2.07	15.7 ± 3.99	45.4 ± 4.83	7.37 ± 3.16	three-state	2.87 ± 4.92 (2.84 ± 6.55)*
m (25 °C) (kJ/ (mol M)) ‖						
urea, pH 4	2.51 ± 0.453	4.18 ± 0.660	- ‡	3.25 ± 0.857	- (native state aggregation)	2.91 ± 0.396 (2.98 ± 0.22)*
urea, pH 7	3.84 ± 0.681	5.71 ± 1.32	- ‡	3.83 ± 0.820	three-state	2.98 ± 0.388 (3.13 ± 0.14)*
GdmCl, pH 7	5.22 ± 0.809	8.26 ± 1.87	9.02 ± 0.984	4.95 ± 1.31	three-state	7.71 ± 3.68 (7.86 ± 4.29)*
$[Denat]_{1/2}$ (25 °C) (M)						
urea, pH 4	7.89	5.36	>10 ‡	1.92	- (native state aggregation)	3.71 (3.62)*
urea, pH 7	7.21	4.62	>10 ‡	3.73	~2.25 / ~ 4.25	4.70 (4.44)*
GdmCl, pH 7	2.24	1.90	5.03	1.49	~0.6 / ~1.3	0.37 (0.36)*

1293

1294 * Values were determined by fluorescence-based unfolding experiments

1295 ‡ No denaturation possible.

1296 ‖ The *m*-value is a measure of the broadness of the transition and correlates with the difference in the accessible surface area between N and U, and
1297 the transition midpoint.

1298

Table 4. Quantification of H-bond strengths from LR-HNCO NMR experiments for all KOW domains.

		*Ec*NusG-KOW				*Mt*NusG-KOW				*Mj*Spt5-KOW			
H-bond #	β-sheet	Donor	Acceptor	$\lvert^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert^{h3}J_{NC'}\rvert$ (Hz)	Donor	Acceptor	$\lvert^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert^{h3}J_{NC'}\rvert$ (Hz)	Donor	Acceptor	$\lvert^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert^{h3}J_{NC'}\rvert$ (Hz)
1	β1-β2	131	148	0.69	0.0077	188	205	0.61	0.0098	92	109	0.70	0.013
2	β1-β2	148	132	0.72	0.0083	205	189	0.69	0.0094	109	93	0.79	0.0088
3	β1-β2	134	146	0.67	0.0081	191	203	0.56	0.011	95	107	0.77	0.012
4	β1-β2	146	134	0.62	0.0096	203	191	0.62	0.0094	107	95	0.67	0.0095
5	β1-β2	136	144	0.65	0.0073	193	201	0.65	0.0080	97	105	0.68	0.048
6	β1-β2	143	136	0.65	0.0088	200	193	0.76	0.013	104	97	0.82	0.012
7	β2-β3	147	161	0.37	0.0105	204	218	0.31	0.015	108	122	0.54	0.010
8	β2-β3	161	147	Peak overlap		218	204	0.69	0.011	122	108	0.65	0.010
9	β2-β3	149	159	0.67	0.012	206	216	0.53	0.013	110	120	0.57	0.030
10	β2-β3	159	150	0.50	0.0086	216	207	0.45	0.012	120	111	0.60	0.010
11	β2-β3	152	157	0.46	0.019	209	214	Peak overlap		113	118	-	-
12	β3-β4	158	173	0.78	0.0077	215	230	0.83	0.0088	119	134	No HNCO peak	
13	β3-β4	173	158	0.64	0.010	230	215	0.62	0.011	134	119	0.62	0.010
14	β3-β4	160	171	0.73	0.0056	217	228	0.75	0.0089	121	132	0.88	0.012
15	β3-β4	171	161	Peak overlap		228	217	0.73	0.0062	132	121	0.47	0.014
16	β3-β4	162	169	0.51	0.0074	219	226	0.50	0.0090	123	130	No H-bond distance	
17	β3-β4	169	162	0.60	0.0066	226	219	0.61	0.0091	No equivalent		-	-
18	β3-β4	167	164	-	-	224	221	0.20	0.019	No equivalent		-	-
19	β5-β1	137	177	0.46	0.017	194	234	0.52	0.025	98	138	No HNCO peak	
20	β5-β1	179	135	Peak overlap		236	192	0.47	0.016	140	96	Peak overlap	
21	β5-β1	135	179	0.73	0.0060	192	236	Peak overlap		96	140	Peak overlap	
22	β5-β1	181	133	-	-	238	190	-	-	142	94	0.46	0.023
23	β5-β1	No equivalent		-	-	No equivalent		-	-	143	94	0.27	0.022
24	β5-β1	No equivalent		-	-	No equivalent		-	-	94	143	0.57	0.010

		hSpt5-KOW5				*Ec*RfaH-KOW				*Vc*RfaH-KOW			
H-bond #	β-sheet	Donor	Acceptor	$\lvert {}^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'}\rvert$ (Hz)	Donor	Acceptor	$\lvert {}^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'}\rvert$ (Hz)	Donor	Acceptor	$\lvert {}^{h3}J_{NC'}\rvert$ (Hz)	σ $\lvert {}^{h3}J_{NC'}\rvert$ (Hz)
1	β1-β2	707	724	0.60	0.0074	113	130	0.76	0.020	116	133	0.87	0.015
2	β1-β2	724	708	Peak overlap		130	114	0.53	0.051	133	117	0.59	0.024
3	β1-β2	710	722	0.70	0.0077	116	128	0.65	0.027	119	131	Peak overlap	
4	β1-β2	722	710	0.50	0.019	128	116	Peak overlap		131	119	0.46	0.020
5	β1-β2	712	720	No HNCO peak		118	126	0.53	0.056	121	129	0.57	0.014
6	β1-β2	719	713	Peak overlap		125	118	0.66	0.026	128	121	0.62	0.017
7	β2-β3	723	735	H-bond peak present. but too weak to quantify		129	142	0.41	0.029	132	145	0.42	0.019
8	β2-β3	735	723	Peak overlap		142	129	0.70	0.019	145	132	0.69	0.027
9	β2-β3	725	734	0.71	0.010	131	140	0.48	0.032	134	143	0.61	0.036
10	β2-β3	733	726	-	-	140	132	0.96	0.021	143	135	1.0	0.011
11	β2-β3	728	731	0.61	0.012	134	138	-	-	137	141	-	-
12	β3-β4	732	745	0.59	0.009	139	154	0.60	0.034	142	157	Peak overlap	
13	β3-β4	745	732	0.62	0.019	154	139	No HNCO peak		157	142	0.68	0.015
14	β3-β4	734	743	0.69	0.039	141	152	0.65	0.049	144	155	0.58	0.019
15	β3-β4	743	734	0.49	0.029	152	141	-	-	155	144	0.72	0.021
16	β3-β4	736	741	0.63	0.033	143	150	0.75	0.033	146	153	0.49	0.024
17	β3-β4	741	736	No H-Bond orientation		150	143	0.47	0.052	153	146	0.48	0.015
18	β3-β4	No equivalent		-	-	148	145	No H-bond orientation		151	148	No HNCO peak	
19	β5-β1	713	749	No HNCO peak		119	158	Peak overlap		122	161	-	-
20	β5-β1	751	711	-	-	160	117	0.51	0.036	163	120	0.57	0.037
21	β5-β1	711	751	0.68	0.023	117	160	0.68	0.015	120	163	0.53	0.016
22	β5-β1	753	709	-	-	162	115	-	-	165	118	-	-
23	β5-β1	No equivalent		-	-	No equivalent		-	-	No equivalent		-	-
24	β5-β1	No equivalent		-	-	No equivalent		-	-	No equivalent		-	-

1303

1304 **Table 5.** Exchange parameters derived from global fitting of the CEST experiments to a two-

1305 state exchange model.

Parameter	hSpt5-KOW5	*Vc*RfaH-KOW	*Ec*RfaH-KOW
p_A (%)	99.15 ± 0.02	99.57 ± 0.01	94.47 ± 0.46
p_B (%)	0.85 ± 0.02	0.43 ± 0.01	5.53 ± 0.46
k_{AB} (s^{-1})	0.76 ± 0.03	0.32 ± 0.02	0.82 ± 0.10
k_{BA} (s^{-1})	88.62 ± 3.12	74.24 ± 3.17	13.98 ± 1.24
k_{ex} (s^{-1})	89.38 ± 3.15	74.57 ± 3.18	14.80 ± 1.31
τ_A (s)	1.31 ± 0.05	3.08 ± 0.15	1.22 ± 0.15
τ_B (ms)	11.28 ± 0.40	13.47 ± 0.57	71.52 ± 6.33
ΔG (kJ/mol)	11.81 ± 0.05	13.48 ± 0.07	7.18 ± 0.21

1306

1307

1308

1309 **Table 6.** Primers used for cloning.

Primer	Sequence (5’ → 3’)
Fw-*Vc*RfaH-KOW	CAT GCC ATG GGA GAG CAA TTG AAG CAT GCC AC
Rv-*Vc*RfaH-KOW	CGC GGA TCC TTA GGT GAC TTC CCA ATC GG
Fw-hSpt5-KOW5	CAT GCC ATG GGC CGG AGG GAC AAC GAA CTC ATC GG
Rv-hSpt5-KOW5	TAG AAT TCT CAG CCC ACC GTG GTG AGC CGC TG
Fw-*Mj*Spt5-KOW	AT GCC ATG GGT AAG AAA ATC ATT GAA AAT ATT GAG AAA GG
Rv-*Mj*Spt5-KOW	CGG AAT TCT TAA TCT TTA TGC TTT GAA ACT ATT TTA AC

1310
1311
1312
1313

1314 **Table 7.** Plasmids.

Plasmid	description	Source
pVS13	*rfaH* from *V. cholera* in pTYB1	I. Artsimovitch
pHC301	*rfaH* from *V. cholera* in pIA238 (a pET28 derivative) (Artsimovitch and Landick, 2002)	(Carter et al., 2004)
pETGb1a-*Vc*RfaH-KOW	$rfaH^{103\text{-}165}$ from *V. cholera* in pETGb1a	This work
pETGb1a-hSpt5-KOW5	human $spt5^{699\text{-}754}$ in pETGb1a	This work
pETGb1a-*Mj*Spt5-KOW	$spt5^{583\text{-}147}$ from *M. janaschii* in pETGb1a	This work
pOTB7_huSUPT5H	cDNA plasmid containing human *spt5*	(Zuber et al., 2018)
pGEX-2TK_*Mj*Spt5-KOW	$spt5^{583\text{-}147}$ from *M. janaschii* in pGEX-2TK	(Hirtreiter et al., 2010)
pETGb1a-*Ec*NusG-KOW	$nusG^{123\text{-}181}$ from *E. coli* in pETGb1a	(Burmann et al., 2010)
pET101d-*Mt*NusG-KOW	$nusG^{178\text{-}238}$ from *M. tuberculosis* in pET101d	(Strauß et al., 2016)
pETGb1a-*Ec*RfaH-KOW	$rfaH^{101\text{-}162}$ from *E. coli* in pETGb1a	(Burmann et al., 2012)

1315

1316

1317 **Source data**

1318 **Figure 2 – Source data 1.** Data for thermal denaturation experiments for all KOW domains.

1319

1320 **Figure 3 – Source data 1.** Data for chemical denaturation experiments for all KOW domains.

1321

1322 **Figure 5 – Source data 1.** CEST fits for *Ec*RfaH-KOW, *Vc*RfaH-KOW and hSpt5-KOW5.

1323 **Figure 5 – Source data 2.** Experimentally determined chemical shift values of urea-denatured
1324 *Ec*RfaH-KOW and predicted random coil chemical shift values of *Vc*RfaH-KOW and hSpt5-
1325 KOW5.

1326

1327 **Figure 6 – Source data 1.** ANS binding by *Ec*RfaH-KOW during urea-based denaturation.

1328

1329

1330