

European Biophysics Journal

The TFE-induced transient native-like structure of intrinsically disordered σ 704 domain of Escherichia coli RNA polymerase.

--Manuscript Draft--

Manuscript Number:	EBJO-D-14-00068R1
Full Title:	The TFE-induced transient native-like structure of intrinsically disordered σ 704 domain of Escherichia coli RNA polymerase.
Article Type:	Original Paper
Keywords:	Intrinsically disordered proteins TFE-induced folding NMR 15N relaxation sigma-70 HLHTH motif
Corresponding Author:	Jarosław Poznański, PhD Institute of Biochemistry and Biophysics PAS Warszawa, POLAND
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Institute of Biochemistry and Biophysics PAS
Corresponding Author's Secondary Institution:	
First Author:	Piotr Kaczka
First Author Secondary Information:	
Order of Authors:	Piotr Kaczka Maria Winiewska Igor Zhukov Bozenna Rempola Krystyna Bolewska Tomasz Lozninski Andrzej Ejchart Anna Poznanska, PhD Kazimierz L Wierzchowski, Prof. Jarosław Poznański, PhD
Order of Authors Secondary Information:	
Abstract:	<p>The transient folding of the domain 4 of a E. coli RNA polymerase σ70 subunit (rECσ704) induced by an increasing concentration of 2,2,2-trifluoroethanol (TFE) in an aqueous solution was monitored by means of CD and heteronuclear NMR spectroscopy. NMR data, collected at a 30% TFE, allowed the estimation of the population of a locally-folded rECσ704 structure (CSI descriptors), and of local backbone dynamics (15N relaxation). The spontaneous organization of the helical regions of the initially-unfolded protein into a TFE-induced 3D-structure was revealed from structural constraints deduced from 15N- and 13C-edited NOESY spectra. In accordance with all the applied criteria, three highly populated α-helical regions, separated by much more flexible fragments, form a transient HLHTH motif resembling these found in PDB structures resolved for homologous proteins. All the data taken together demonstrate that TFE induces a transient native-like structure in the intrinsically disordered protein.</p>

Response to Reviewers:

see attachment

Reviewer #1:

The manuscript titled to "The TFE-induced transient native-like structure of intrinsically disordered σ 704 domain of Escherichia coli RNA polymerase." reports the authors' finding that the σ 704 domain of E. coli RNA polymerase (rEC σ 704) adopts partially-folded α -helical structures in 30% TFE. By employing CD and NMR spectroscopic techniques, the authors show that 30% TFE maximizes the native-like transient conformation of rEC σ 704, and at this condition, they were able to characterize dynamic behaviors of this protein and obtain structural models in solution. The determined structural models are comparable with recent X-ray crystallographic structures of E. coli RNA polymerase in its holo-state (Murakami 2013).

The authors reported in their previous publications (e.g. Poznanski et al. 2003, and Kaczka et al. 2009) that rEC σ 704 contains a significant amount of helical structures in its unfolded state or in the condition containing 10% TFE. The current manuscript follows up these previous studies by focusing on the structural and dynamic properties of rEC σ 704 in 30% TFE. The work is of interest in that it provides evidence that the conformation of rEC σ 704 at more physiological condition could be similar to the one observed here with 30% TFE, and in that it presents the plausible solution structural models of rEC σ 704. However, the manuscript writing should be improved significantly, correcting numerous errors.

A few other issues that need to be addressed before publication:

1. It appears that the figures S1 and S2 provide evidence suggesting that the conformation observed with 30% TFE is similar to the one observed at more physiological pH. However, detailed discussion regarding this observation is lacking.
 - A) For example, would it be possible to compare the α -helical contents at two conditions (i.e. 30% TFE vs. 100 mM NDSB at neutral pH)?

J.P. discussed and appropriate table inserted in supplementary materials.

s.s. element	200-260 nm		205-260 nm		210-260 nm		consensus	
	TFE	NDSB	TFE	NDSB	TFE	NDSB	TFE	NDSB
Helix	35.3%	34.2%	37.1%	39.2%	35.5%	36.3%	36.0 (1.0)%	36.6 (2.5)%
Antiparallel	8.0%	8.6%	7.1%	6.7%	7.6%	7.5%	7.6 (0.5)%	7.6 (1.0)%
Parallel	8.1%	7.9%	8.0%	7.5%	8.2%	8.0%	8.1 (0.1)%	7.8 (0.3)%
Beta-Turn	16.5%	17.0%	16.1%	15.8%	16.4%	16.3%	16.3 (0.2)%	16.4 (0.6)%
Rndm. Coil	30.5%	28.9%	30.7%	29.2%	31.7%	31.2%	31.0 (0.6)%	29.8 (1.3)%
Total Sum	98.4%	96.5%	99.0%	98.5%	99.4%	99.3%	98.9 (0.5)%	98.1 (1.4)%

B) What could be the cause for pH-dependent spectral changes of rEC σ 704 (particularly at higher pH)?

J.P. explained in the text as follows

This phenomenon should be directly addressed to the dissociation of Asp and Glu side-chain carboxyl groups at pH above 5. Dissociation of Asp and Glu side-chains enables formation of numerous intra- and intermolecular salt bridges that compete with hydrophobic interactions. This destroys a hydrophobic core of the transiently-folded protein, and the exposition of hydrophobic residues towards solvent results in intensive aggregation.

C) Also, please note that information regarding the sample concentration for CD is lacking (except for figure S2B).

J.P. explained: *All the spectra were recorded for ~1.5 μ M protein solution under a nitrogen atmosphere, using a 10 mm path-length quartz cell. Exact protein concentration was determined for each sample basing on UV absorption at 274 nm ($\epsilon = 1400 M^{-1} \cdot cm^{-1}$).*

2. According to the references cited for interpretation of R1xR2 product (Kneller et al. 2002 and Jaremko et al. 2014), any contribution from Rex would result in increasing R1xR2 product. However, the paragraph at the lines 28-37 of page 8 claimed that the decrease in R1xR2 product is caused by exchange process. Could the authors give more detailed explanation for this discrepancy?

J.P. This was reinterpreted in the terms of R2/R1, which seems to be more appropriate for unfolded proteins. The "exchange" above was an error; it should be that low R1*R2 indicates residues with fast motion

3. The authors provided the RMSD difference between their structural models and the structures of *T. maritima* TM σ A4. However, they did not quantify how different their structural models are from the crystal structure of *E. coli* RNAP holoenzyme, which was actually compared with the determined structures at figure 4.

J.P. corrected: New figure 4, now with both reference structures . RMSD to both structures is below 2.0Å.

4. Along with other minor errors that should be fixed, please check the following issues:
 - a. The caption for figure 4 states that the panel C shows the superposition of the solution structure with the published crystal structure of *E. coli* rEC σ 704. But, the first paragraph of the discussion section says that "In order to check to what extent the TFE-induced solution structure of rEC σ 704 resembles that of the homologous *Thermotoga maritima* TM σ A4 (Lambert et al. 2004), the most stable regions formed by the H1 and H3 helices of these two structures were compared (Figure 4)." This might confuse the reader which structures are indeed overlapped at figure 4.
 - b. At the 10th page line 36-37, the authors says "(see dis@30 vs. dis@20, dis@20 vs. dis@0, and ord@30 vs. ord@20 in Table III)". But, in table 3, there is no 'dis@20' or 'ord@20'.
 - c. At the table 2, the reviewer believes that the unit for τ_f is 'ps', not 'ns'. Similar errors exist at the table 3 (the unit for τ_f) and at the figure 3 (the units for τ_s and τ_f).
 - d. Figure 3 and 4 do not have labeling for their panels. And, the caption for figure 3 does not explain what the black bar above each gray square represents.
 - e. A reference error at the caption of the figure S3.

J.P. corrected.

Reviewer #2:

The authors submit a manuscript closely following their previous study published in Proteins [Kaczka et al, Proteins 78 (2010) 754-768]. Both papers demonstrate formation of a native-like structure of an intrinsically disordered protein (IDP) due to the presence of trifluoroethanol (TFE), a fact that is not surprising, but worthy of biophysical characterization. The only substantial difference between the papers is the increased concentration of TFE to 30 % in the current study (only 10% TFE was used in the previous paper). Therefore, the authors should provide a strong argument that the mentioned difference provides a qualitatively new insight if they want to present their data as a novel and original scientific result. I am afraid that the current version of the manuscript fails to provide such an argument (i) because an explicit comparison of both papers is missing, (ii) NMR relaxation analysis is not presented in a clear format, and (iii) the manuscript contains typos/inconsistencies and errors that obscure the results. As the quality of the experimental data and computational protocols seems to meet the standards, I believe that the study can be published, but only if the authors make a major revision of the manuscript and if they succeed to prove that the new results carry a qualitatively new "added value" compared to their previous report.

Major issues:

1. Analysis of the ^{15}N relaxation data is not described well and important bits of information are missing: It is not described how the Rex contribution at different magnetic fields is treated in the analysis (proportional to B^2 ?). References to important statistical procedures are completely missing (F-test, Anderson-Darling test) and the procedures are not described in the text. As these issues are treated properly in the previous paper [Kaczka 2010], where they are described in a clear and detailed manner both in Results and Methods, I recommend to follow the style of the previous article. The approach to the ^{15}N relaxation analysis differs substantially from the previous paper but it is not described why. The Clore's extended model-free formalism was found inappropriate in the previous study but is used in the current manuscript. The reduced spectral density mapping was performed in the previous study but is missing in the current manuscript. The authors should include the spectral density mapping and clearly explain reasons of using different approaches in the analysis.

J.P. We have reanalyzed all data, starting from manual correction of peak location in the spectra through spectral density mapping up to more appropriate statistical analyses. Now, we follow the methods used in Kaczka paper, so the results could be easily compared. The usage of Clore's extended model-free formalism was found applicable only for the His-tag and linker residues, for which an overall correlation time was fixed at 11.5ns (median estimated for HLHTH motif). But usage of overall correlation time for so weakly folded protein is debatable. Consequently we remove this extension, and followed the previously used approach.

2. The purpose of including the $R_1^*R_2$ analysis is not clear. The method has been proposed to distinguish effects of Rex and motional anisotropy in relatively large, well unfolded proteins with sufficiently long correlation times ($\tau \gg \omega$). Under such circumstances, $R_1^*R_2$ is almost independent of τ , being approximately proportional to S^4/ω^2 . In disordered proteins/protein regions, the condition $\tau \gg \omega$ is clearly not fulfilled; reference to Kneller et al on p. 8, l. 25 is not relevant in this context. This issue should be discussed in a clear way (do the authors consider only data for relatively ordered central domain?). It should be explained why the variations of $R_1^*R_2$ in the well-ordered region are much larger than those reported in the literature [Kneller et al. JACS 2002, Jaremko et al. J. Struct. Biol. 2014]. Interpretation of the $R_1^*R_2$ analysis is confusing (and inconsistent with the model-free analysis of the same data!). Exchange INCREASES R_2 (and consequently $R_1^*R_2$), low R_2 and NOE (and $R_1^*R_2$, p.8 l.32) shows fast internal dynamics, not exchange (Rex). The highest Rex values are not identified in the same region by the model-free analysis (Fig. 3E), but for residues exhibiting highest $R_1^*R_2$. Either exchange (Rex) is confused with fast dynamics (S^2 , τ) in the manuscript, or a wrong conclusion is made.

J.P. We must agree that the usage of $R_1^*R_2$ was not supported. We move back to the analysis of field-dependence of $J(0)$, and also presented R_2/R_1 as an alternative approach. Some of above disagreements came from extreme instability of fitting procedure, results of which strongly depended on initial values of S^2 , T_{local} and T_{e} . The previously used three parameter model (Kaczka in Proteins) was found reasonable, but even its extension for Rex contribution makes results "flexible". So we decide to mention for which residues Rex may contribute to transverse relaxation.

3. The Rex values (provided by the model-free analysis) are plotted in Fig. 3E, but not discussed in detail. The authors should make clear which Rex values they consider significant and why their conclusion differs from the previous paper [Kaczka 2010], where "the exchange phenomena were judged negligible for all residues of the protein". Does the protein behaves differently on the us-ms time scale in 30% TFE, or did the higher field

provide higher sensitivity that allowed the authors to observe Rex values that were below the detection limit in the previous study?

J.P. The appropriate test (proposed in 1852) has been applied, and four residues were identified as these experiencing slow motions. Location of two agrees with turn region in HLHTH motif, while two others are most probably involved in folding-unfolding equilibrium

4. The average order parameter (S^2) for 30% TFE in Table II does not agree with the values plotted in Fig. 3A. This is an extremely important issue! Does the higher TFE concentration increase the backbone rigidity dramatically (Fig. 3A) in spite of the only moderately higher helical propensity (Fig. 1a), or is the S^2 increase also marginal (from 0.32 to 0.36, Tab. 2)? Also, the τ_f average for 30% TFE in Tab. 2 seems to be lower than the values plotted in Figure As mentioned above, the effects of higher TFE concentration should be directly compared to the results of the previous study [Kaczka 2010] in Discussion.

J.P. All these are corrected. The increase in S^2 is significant, as tested by non-parametric K-W test. All the analyses were done again. The relaxation data do agree with those presented in the figures. Statistical methods used enable comparison of non-Gaussian distributions (Table3)

Minor issues:

Description of the model-free analysis in MATERIALS AND METHODS is somewhat difficult to follow (Eqs. 2A-2C). The Rex terms are mentioned in text but do not appear in equations. **J.P. Corrected**

This section can be improved by listing all used model-free models explicitly, like e.g. in Mandel et al., J. Mol. Biol. 246 (1995) 144-163.

J.P. Only the simplest model is used now, application of more advanced generally failed. The only method is to apply a set of correlation times for T_{local} attributed a priori to selected protein regions, as it was originally proposed by Buck and al for Lysozyme in TFE/water binary solvent (J. Mol Biol. 1996).

Including the Clore's extended model, the notation used by the authors is a bit confusing because the symbol τ_s is used for a different correlation time than in the original paper [Clore et al, JACS 112 (1990) 4991-4993].

J.P. Now they are t_{local} (against t_m) and t_e (as originally)

It would be interesting to see how the "full extended model" (Eq. 2 in [Clore et al, JACS 112 (1990) 4991-4993]) performs on the relatively large set of experimental data.

J.P. This model converges only when we have restricted T_{local} values. Otherwise the strong correlation between S_s and T_{local} makes results enormously biased.

The prediction of the helical content by the authors is convincing, but it would be interesting to compare it to results of other software (e.g. **d2d** by M. Vendruscolo).

J.P. Done, included in the text and agrees with other CSI analyses.

1. p. 8, l. 42 "chemical data" should be "chemical shift data"

J.P. corrected

2. p.10, l.9 "reaching maximum of its variance" - were higher TFE concentrations tested?

J.P. Up to 35%. Now presented in Supplementary Figure S3.

3. p.10, l.36 "dis@20" - not in Table III !

J.P. It should be dis@10. Now rewritten.

4. It should be mentioned explicitly that the cumulative tau plot contains only values obtained at 30% TFE.

J.P. Now distributions of S^2 , T_{local} and t_e are shown for all TFE concentrations. And discussed.

5. p. 17, l. 12 "Horizontal bars represent estimates the experimental error" should be "Vertical..."

J.P. corrected.

6. Fig. 3 "[ps]" should be "[ns]". Also, the labels should be placed at the axes as typical. The horizontal axis (showing residue #) should be labeled as well.

J.P. Corrected.

7. Why are the gray strips identifying helical regions different in Fig. 2 and Fig. 3?

J.P. strips were put manually, now carefully updated.

8. Check for typos throughout the manuscript, especially in the Figure legends ("stripes", "aorich")

J.P. Checked several times.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Piotr Kaczka^{1,a}, *Maria Winiewska*^{1,a}, *Igor Zhukov*¹, *Bożenna Rempola*¹, *Krystyna Bolewska*¹,
*Tomasz Łoziński*¹, *Andrzej Ejchart*¹, *Anna Poznańska*², *Kazimierz L. Wierzchowski*¹, *Jarosław*
Poznański^{1,*}

The TFE-induced transient native-like structure of intrinsically disordered σ^{70}_4 domain of *Escherichia coli* RNA polymerase.

1. Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a, 02-106 Warszawa, Poland
 2. Centre for Monitoring and Analyses of Population Health Status, National, Institute of Public Health - National Institute of Hygiene, Warszawa, Poland
- a. these authors contributed equally.
- * jarek@ibb. waw. pl; +48 22 592 5783

ABSTRACT

The transient folding of the domain 4 of a *E. coli* RNA polymerase σ^{70} subunit (rEC σ^{70}_4) induced by an increasing concentration of 2,2,2-trifluoroethanol (TFE) in an aqueous solution was monitored by means of CD and heteronuclear NMR spectroscopy. NMR data, collected at a 30% TFE, allowed the estimation of the population of a locally-folded rEC σ^{70}_4 structure (CSI descriptors), and of local backbone dynamics (¹⁵N relaxation). The spontaneous organization of the helical regions of the initially-unfolded protein into a TFE-induced 3D-structure was revealed from structural constraints deduced from ¹⁵N- and ¹³C-edited NOESY spectra. In accordance with all the applied criteria, three highly populated α -helical regions, separated by much more flexible fragments, form a transient HLHTH motif resembling these found in PDB structures resolved for homologous proteins. All the data taken together demonstrate that TFE induces a transient native-like structure in the intrinsically disordered protein.

Keywords: Intrinsically disordered proteins; TFE-induced folding; NMR; 15N relaxation; sigma-70; HLHTH motif

INTRODUCTION

1 Intrinsically disordered proteins (IDPs) are involved in the essential processes of the living cell (1). Their
2 sequences are generally enriched in polar and charged residues, but deficient of hydrophobic groups (2-4). Free IDPs in
3 a solution are highly flexible, and do not display stable a secondary/tertiary structure (5, 6). Under physiological
4 conditions IDPs are sampling a large conformational space, and some of conformers, called a transient secondary
5 structure, may resemble structured elements (7). In order to access the relationship between the structural dynamics of
6 IDPs and the function, a continuum of thermodynamically accessible states have to be characterized in terms of both
7 conformational free energy, and inter-conversion timescales. Hence, a global and/or local flexibility provides the
8 structural plasticity of IDPs, facilitating specific interactions with various partners (8), albeit the thermodynamics of
9 partner-induced IDP folding must be linked directly to the intrinsic conformational properties (2-4, 9). IDPs, among
10 other, are prominently represented in proteins associated with mRNA processing (10, 11), apoptosis (12), signal
11 transduction (13, 14) or transcription regulation (6, 7, 15-17).

12 In bacteria, two forms of RNA polymerase (RNAP), called holoenzyme (composed of $\alpha 2\beta\beta'\omega\sigma$ subunits of a
13 total weight of ~ 450 kDa), and core (composed of $\alpha 2\beta\beta'\omega$ subunits; ~ 400 kDa), are capable of transcribing DNA, but
14 only the holoenzyme initiates transcription solely at promoter sites. The identification of σ -factor as a dissociable RNA
15 polymerase subunit (18) suggested that RNA polymerase may recruit other σ -factors to switch on particular regulons
16 (19). There are seven different σ subunits identified in *Escherichia coli* (20), each binding to the same core RNAP
17 enzyme, and, thereby, directing the transcription of a specific ensemble of genes. The actual model of bacterial
18 transcription assumes a competition of sigma subunits for the limited amount of the core RNAP that it can be bound to
19 (20-22). In case of vegetative cell growth, σ^{70} predominates at a concentration exceeding that of the core RNAP (21,
20 23), and displaying the highest affinity of all sigma proteins to the core RNAP (20) contributes in the transcription of
21 70% of the operons (24).

22 In a solution, free σ^{70} -subunit does not bind to promoter DNA due to a direct intramolecular interactions
23 between N-terminal segment (region 1.1) and DNA-recognizing motifs located in the C-terminal part of the molecule
24 (regions 3.2 and 4.2) (25). The primary sigma factor of *Escherichia coli* RNA polymerase ($EC\sigma^{70}$) is a 613-residue
25 protein of 70 kDa, belonging to the large family of sigma factors composed of four strongly conserved domains (24).
26 Its regions (2.4 and 4.2) specifically recognize the promoter sequences of expressed genes, located 10 and 35 bp
27 upstream from the transcription initiation site, respectively. Some parts of the 4.2 region have also been identified as
28 targets for transcription-activating or anti-sigma factors (26-28). Free σ^{70} in a solution must be regarded as carrying
29 intrinsically disordered segments, as indicated by solution NMR studies on the segmental isotopic labeling of the
30 isolated region 4.2 of $EC\sigma^{70}_4$ (25); this conclusion seems to be supported by the absence of a crystal structure of the free
31 form of sigma subunit, with the only exception of the solution structure of *Thermotoga maritime* σ^A factor (TM σ^A)
32 (27).

33 At the moment, only two solution structures of $EC\sigma^{70}_4$ in complex with T4 AsiA (record 1TLH) (27) and *E.*
34 *coli* regulator Rsd (2P7V) (28) are accessible in the Protein Data Bank. There is the known crystal structure of $EC\sigma^{70}_4$
35 in complex with a DNA-promoter fragment (3T72) (29), and the structure of whole σ^{70} in RNAP holoenzyme of *E. coli*
36 (4IGC) (30), as well as its complexes with small-mass ligands (4KN7, 4KN4, 4KMU and 4JK2, 4JK1) (31, 32). There
37 are also the structures of homologous domains from thermophilic bacteria (27, 33, 34). However, there is still no
38 structural data concerning free $EC\sigma^{70}$ in solution.

1 For our part, we have demonstrated that the population of secondary structure elements was substantially
2 increased for rEC σ^{70}_4 in the presence of TFE (35). This generally agrees with the idea that TFE induces the closure of
3 individual hydrogen bonds (36), and thus propagates the context-dependent formation of native-like secondary
4 structures in polypeptides (37). Subsequently, we have monitored the TFE-induced conformational changes of rEC σ^{70}_4
5 with the aid of NMR spectroscopy (38). The drift of resonance signals in the ^1H - ^{15}N HSQC spectra indicated that in a
6 mixed water-TFE solution a non-specific uniform build-up of helical structures takes place for a TFE content smaller
7 than 10% (v/v). However, for a TFE content exceeding 10%, non-monotonic changes were observed for numerous
8 residues located either in the linker between domains: 3 and 4, or in the regions separating helical structures in the
9 putative HLHTH motif (helix-loop-helix-turn-helix), namely G564-Y571 (L) and D581-T583 (T) (see PDB record
10 4igc). It supports the stating that at higher TFE concentrations, rEC σ^{70}_4 would tend to fold in a transient native-like 3D
11 structure found in the crystal structure of the *E. coli* RNAP whole holoenzyme (30). Yet, does EC σ^{70}_4 in a solution
12 resemble a sigma factor from *E. coli* RNAP much more than the homologous domains of *Thermus aquaticus* (34) and
13 *Thermus thermophilus* (39), or the solution structure of *Thermotoga maritime* (27), and that of a distant one, σ^{54} from
14 *Axulfex aeolicus* (2ahq) (40)? In this work, on the basis of further CD and NMR studies we show a low-resolution
15 model of *E. coli* σ^{70}_4 fold induced by a 30% TFE, and we analyze internal motions of the protein backbone with the use
16 of ^{15}N relaxation.
17
18
19
20
21
22
23
24

25 MATERIALS AND METHODS

26
27 **Materials.** The preparation of the recombinant EC σ^{70}_4 fragment, rEC σ^{70}_4 , (86 C-terminal residues of the *E. coli* σ^{70}
28 preceded by 21-aa segment carrying His₆-Tag) was described previously (41). The protein was found to be chemically
29 stable for a prolonged time both in aqueous acidic, and in mixed aqueous/TFE solutions.
30

31 **CD Spectroscopy.** Circular dichroism spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a
32 Peltier thermostatic cell holder. All the spectra were recorded for $\sim 1.5\mu\text{M}$ protein solution under a nitrogen atmosphere,
33 using a 10 mm path-length quartz cell. Exact protein concentration was determined for each sample basing on UV
34 absorption at 274 nm ($\epsilon = 1400 \text{ M}^{-1}\cdot\text{cm}^{-1}$). Each CD spectrum was measured three times at 25°C in the range of 195-270
35 nm. The contribution of the secondary structure elements were estimated with the aid of the CDNN program (42).
36
37
38

39 **NMR Spectroscopy.** 2.5 mM protein samples were prepared in a 30% (v/v) TFE/water binary system (10% D₂O) at
40 pH 4.6. All rEC σ^{70}_4 NMR spectra were collected at 298K either on Varian Unity Plus 500 or Varian VNMRs 800
41 spectrometers, processed with the aid of NMRPipe (43) and analyzed, using SPARKY (44). Prior to the Fourier
42 transformation, resolutions in indirect dimensions were increased by $\pi/3$ shifted squared sine-bell weight function,
43 followed by zero-filling. The zero-order baseline correction was applied in all dimensions. ^{13}C and amide ^1H and ^{15}N
44 resonances were assigned, using a combination of HNCBCA (45), CBCA(CO)NH (46) and C(CO)NH (47)
45 experiments. Additional HNHA (48) and HAHB(CO)NH (49) spectra enabled the assignment of the H α and H β
46 signals, while most of the remaining ^1H aliphatic side-chain resonances were assigned, using H(C)(CO)NH spectrum
47 (50). Carbonyl ^{13}C resonances were assigned using the combination of HNCO (51) and HN(CA)CO (52) spectra. To
48 overcome a problem with strong resonances overlapping, sequence-specific assignments were additionally confirmed
49 by the inspection of sequential contacts identified in ^{15}N -edited 3D-NOESY spectra. The assignments have been
50 deposited in BMRB (Entry 15975).
51
52
53
54
55
56

57 **Chemical shift analysis.** Secondary chemical shifts, $\Delta\delta(i)$, were calculated for $^{13}\text{C}\alpha$, ^{13}CO and $^1\text{H}\alpha$ nuclei according to
58 the reference values for random-coil chemical shifts corrected for sequence-dependent contributions (53). The
59
60
61
62
63
64
65

population of helically folded regions in $rEC\sigma_4^{70}$ was then evaluated, according to the concept of the chemical shift index (CSI) (54). The helical population of individual residues, $p_\alpha(i)$, was estimated as the ratio of $\Delta\delta(i)/\Delta\delta_o(i)$, where the reference values of secondary chemical shifts, $\Delta\delta_o(i)$, were assumed to 2.8, 2.3 and -0.4 ppm for $^{13}C\alpha$, ^{13}CO and $^1H\alpha$, respectively (54, 55). In view of the consistency of the $\Delta\delta(i)$ distribution pattern for $^{13}C\alpha$, ^{13}CO and $^1H\alpha$, the consensus local propensities towards a helical structure were expressed as the geometric average of the three descriptors referred to above (38). The population of helical structures were additionally estimated on $\delta 2D$ server (56), algorithm of which is based on the larger set of chemical shift data ($^1H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO , N and HN) (57). The protein backbone ϕ and ψ angles, and the square of the order parameter for the backbone amides, S^2 , were additionally estimated with the aid of TALOS+ (58).

^{15}N relaxation data. Relaxation parameters were measured at 298K for a uniformly ^{15}N -labelled protein sample dissolved in a mixed aqueous solution with 30% TFE content (pH 4.6) on Varian UnityPlus 500, Varian VNMRS 600 and Varian VNMRS 800 spectrometers at 11.4, 14.1 and 18.7 T, respectively. The pulse sequences used for the determination of the ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates were identical for all the three spectrometers, being analogous to these originally reported (59). For R_2 measurements, a Carr-Parcell-Meiboom-Gill (CPMG) 180° pulse train with the refocusing delay of 650 μs was used during evolution (60). Delays between proton π pulses, used for the suppression of cross-correlation effects between 1H and ^{15}N nuclei (61), were 5 and 10 ms in R_1 and R_2 measurements, respectively. The recycle delay was kept as long as 2.5 s. 1024 and 128 complex data points in time domains were collected in the hypercomplex mode. ^{15}N decoupling during the acquisition was performed as 3.2 kHz GARP pulse scheme (62). The relaxation rates were estimated, using ten delays for R_1 (0.01, 0.09, 0.17, 0.29, 0.41, 0.55, 0.69, 0.85, 1.01 and 1.25 s), and eight delays for R_2 (0.01, 0.03, 0.05, 0.09, 0.13, 0.17, 0.21 and 0.25 s). $\{^1H\}$ - ^{15}N heteronuclear *NOEs* were measured according to the dynamic progressive saturation (63, 64), using the standard pulse sequence included in the BioPack (Varian Inc., PaloAlto, USA) software. To overcome problems with peaks overlapping, the *NOE* intensities were assumed as cross-peak's heights. The R_1 and R_2 rates were then estimated by fitting an exponential decay curve in the form $I(t) = I_0 \exp(-R_i \cdot t)$ implemented in SPARKY, where I is the signal intensity and t is the evolution time for appropriate magnetization. Severe signal overlapping limited the quantitative interpretation of relaxation parameters to only 84 out of 101 assigned backbone amide ^{15}N nuclei.

Analysis of the ^{15}N relaxation data. The ^{15}N relaxation data (R_1 , R_2 , 1H - ^{15}N *NOE*) were interpreted according to the reduced spectral density mapping, via sampling spectral density function, $J(\omega)$, at low (0), intermediate (ω_N) and high (ω_H) frequencies (65).

$$R_1 = \left(\frac{d^2}{4} \right) \cdot [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N)$$

$$R_2 = \left(\frac{d^2}{8} \right) \cdot [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \left(\frac{c^2}{6} \right) \cdot [4J(0) + 3J(\omega_N)] + R_{ex}$$

$$NOE = 1 + \left(\frac{d^2}{4R_1} \right) \cdot \left(\frac{\gamma_H}{\gamma_N} \right) \cdot [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

$$where : c = \frac{\omega_N \cdot \Delta\sigma}{\sqrt{3}}, d = \left(\frac{\mu_0 h}{8\pi^2} \right) \left(\gamma_N \gamma_H \right) \left\langle \frac{1}{r_{NH}^3} \right\rangle, \Delta\sigma = -160 ppm, r_{NH} = 1.02 \text{ \AA}, R_{ex} = a \cdot B_0^2$$

(1)

where $J(\omega)$ is the spectral density of molecular motion at a given angular frequency, and the additional term R_{ex} , which scales with the square of magnetic field, stands for the contribution of micro- to millisecond motions to R_2 (66, 67). The above system of equation can be resolved against $J(\omega)$ (65) giving:

$$\begin{aligned}
 J(0.87 \cdot \omega_H) &\cong (0.8/d^2) \cdot (\gamma_N / \gamma_H) \cdot (NOE - 1) \cdot R_1 \\
 J(\omega_N) &= (4R_1 - 7d^2 \cdot J(0.87 \cdot \omega_H)) / (3d^2 + 4c^2) \\
 J(0) &= (6 \cdot R_2 - (9d^2/4 + 3c^2) \cdot J(\omega_N) - 39d^2/4 \cdot J(0.87 \cdot \omega_H)) / (3d^2 + 4c^2)
 \end{aligned} \tag{2}$$

The residues with possible contribution of conformational and/or chemical exchange to the transverse ^{15}N relaxation were identified from the field-dependence of $R_2 - 1/2R_1$ (Equation 3), as these for which estimated R_{ex} were identified, according to Peirce's criterion (68) as outliers in the overall distribution of R_{ex} .

$$R_2 - \frac{R_1}{2} = \frac{d^2}{4} [2J(0) + 3J(\omega_H)] + \frac{2c^2}{3} J(0) + R_{ex} \cong \underbrace{\frac{d^2}{2} J(0)}_{a_0} + \underbrace{\frac{2c^2}{3} J(0) + R_{ex}}_{a_2 \cdot B_0^2} \tag{3}$$

Since the studied protein was known to be highly unfolded (35, 38), the standard Lipari-Szabo model-free approach (69, 70), which attributes a global correlation time (τ_m) to the overall rotational diffusion, was found inappropriate (71, 72). Hence, an alternative approach of the direct fitting of the spectral density function (Equation 4) to the $J(\omega)$ data sampled at 0, ω_N and $0.87 \cdot \omega_H$ was used.

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \cdot \tau_{local}}{1 + (\omega \cdot \tau_{local})^2} + \frac{(1 - S^2) \cdot \tau'}{1 + (\omega \cdot \tau')^2} \right] \tag{4}$$

where S^2 is the square of the residue-specific generalized order parameter for slow motions, $1/\tau' = 1/\tau_{local} + 1/\tau_e$, and τ_{local} and τ_e are the effective, residue-specific correlation times for slow ($>1\text{ns}$) and fast ($<100\text{ps}$) motions, respectively.

Structural calculations. NOE-derived distance restraints were extracted from the ^{15}N -edited 3DNOESY-HSQC (73) (mixing time: 150 ms) and ^{13}C -edited 3D-NOESY-HSQC (74) (mixing times: 90 ms and 250 ms) experiments. The cross-peaks in 3D-NOESY spectra were manually picked and assigned with the SPARKY program (44), and their heights were then converted into distance restraints with the aid of the CALIBA procedure included in CYANA 2.1 (75). Additional constrains for backbone torsion angles were adopted from TALOS+ (58). Initial structural calculations were made with the simulated annealing (SA) protocol implemented in XPLOR 3 (76, 77). The structure obtained with homology modeling procedure was used as a template. The SA protocol started with 6ns dynamics at 1000K, with dihedral angle restraints and NOE restraints scaled to $5 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ and $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$, respectively, and the van der Waals' radii scaled down by a factor of 0.75. After that, the system was slowly cooled to 100K during 3ns, with the gradually increased terms for the repulsive van der Waals' term (from 0.003 to $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^4$) and for dihedral angles ($200 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^2$). Each SA cycle finished with 1ns low-temperature evolution, followed by 1000 steps of Powell energy minimization. This protocol was repeated 500 times, and a cluster of 11 lowest pseudo-energy structures was selected for further analyses. The final refinement was performed, using additional constraints for experimental $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$ and $^1\text{H}\alpha$ chemical shifts, using chemical shift protocol (78, 79). Finally, the structures were then inspected with PROCHECK-NMR (80), and all of them were subjected to further 5 ns molecular dynamic simulations in a water box, using Yasara2 force field implemented in Yasara Structure package (81). To prevent the momentary unfolding of the protein, and to mimic the helix-promoting propensities of a 30% TFE solution, the medium-range ($i, i+3$)

constraints and angular restraints, imported directly from the SA protocol, were applied during the first 2 ns of simulations.

Statistical tests. The relaxation parameters were analyzed in terms of their distributions, and the consistency of this data with normal distribution was initially tested with the aid of the Anderson-Darling goodness-of-fit test (82). Since some of the distributions significantly differ from the Gaussian one, the relaxation parameters obtained at different conditions (i.e. TFE concentration) or for different subsets of residues (e.g. His₆-tag vs. HLHTH motif) were compared using the appropriate nonparametric Mann-Whitney U-test (83). All tests were done with the aid of Statistica package (84).

Computational methods. All the data analyses and presentations were performed using GnuPlot 4.6 (85).

RESULTS

The HLHTH fold of rEC σ ⁷⁰₄ in a solution. The protein has been previously found soluble solely in a low-pH solution (35). Our previous studies demonstrated that the solubility of the protein substantially decreases above pH 4.5, which is far away from the theoretical pI of 9.1±0.3, estimated as the average of nine different methods (86). This phenomenon should be directly addressed to the dissociation of Asp and Glu side-chain carboxyl groups at pH above 5. Dissociation of Asp and Glu side-chains enables formation of numerous intra- and intermolecular salt bridges that compete with hydrophobic interactions (35). This destroys a hydrophobic core of the transiently-folded protein, and the exposition of hydrophobic residues towards solvent results in intensive aggregation.

In order to stabilize the protein in a neutral solution, a number of solvent systems preventing the formation of salt bridges have been tested. Thus, 200 mM Na₂SO₄ and 1.8 M MgSO₄ were found to improve minutely protein solubility at neutral pH, additionally stabilizing helical forms. However, limited protein solubility and required salt concentration precluded the application of any NMR techniques. The high concentration of arginine (~1 M) also improved rEC σ ⁷⁰₄ solubility (87), but such a high concentration precluded even the CD monitoring of a putative Arg-induced secondary structure formation. Substantial progress was achieved with the use of zwitterionic Non-Detergent SulfoBetaine stabilizers (NDSB). NDSB195 does not affect rEC σ ⁷⁰₄ at low-pH (Supplementary Figure S1A), albeit its 100 mM solution efficiently screens electrostatic interactions between dissociated side chains of Asp/Glu and these of Arg/Lys, and the protein solubility approached 50 μ M at pH 7.7. Moreover, in the presence of 100 mM NDSB195, an increase in pH of rEC σ ⁷⁰₄ solution results in the detectable increase of the estimated contribution of helical forms, which is clearly evidenced in CD spectra recorded at pH ranging from 2.6 to 7.5 (Supplementary Figure S2A), while the further increase of pH unfolds the protein. An optimal range of solution pH at which a protein is properly folded is a common protein property (88, 89). All CD spectra, analyzed together under the assumption of two-state equilibrium, show that, even at neutral pH, the protein is only partially folded. The maximal contribution of a putative folded state (approximately 50%) was achieved at the physiological pH of 7.5 (see the insert in Supplementary Figure S2A). It should be stressed that 100mM NDSB195 does not preserve protein structure for rEC σ ⁷⁰₄ concentration exceeding 2 μ M (see Supplementary Figure S2B for details), so this solvent system could not be used for heteronuclear NMR studies. Altogether, it clearly indicates that the previously proposed TFE-water binary solvent system is the most suitable for the NMR studies on rEC σ ⁷⁰₄ structural preferences. The 30% TFE concentration was found optimal, and the further addition of TFE did not result in visible changes of protein structure (Supplementary Figure S3). It should be mentioned that the CD spectrum of rEC σ ⁷⁰₄ recorded in the presence of NDSB195 at pH 7.5 (Supplementary Figure S1A) resembles that recorded in TFE-water binary solvent (Supplementary Figure S3). This impression is strongly

supported by the deconvolution of CD spectra against contribution of secondary structure elements, populations of which were estimated with the aid of CDNN program (see Supplementary Table S1).

NMR resonance assignment and protein secondary structure preferences derived from CSI data in a 30% (v/v)

TFE solution. The resonance assignments for rEC σ_4^{70} at 30% TFE (cf. Methods) were deposited in a BMRB database under accession number 15975. The consistent pattern of the downfield shifts of $^{13}\text{C}\alpha$ and ^{13}CO resonances, accompanying the upfield shift of the $^1\text{H}\alpha$, was observed for numerous residues. The values of the score function, $\tilde{p}_\alpha(i)$ (see (38) for details) identify significantly populated helical patterns, located in the regions: 3.2 (L532-A542; H0), 4.1 (L551- F563; H1), and 4.2 (L573-L599; H2, H3). Another less populated fragment, S602-S609, is located in the C-terminal part of rEC σ_4^{70} (Figure 1A). The three clearly distinguished maxima coincide with the location of helical regions in the structures of homologous σ subunits (27-30, 33), while the last one (S602- S609) could only be found in σ^A_4 of *T. maritime* (27) and in *E. coli* σ_4^{70} in complex with an Rsd regulator (28), and in the recently solved structure of *E. coli* RNAP holoenzyme (30). The same trends, including both location and population of helical structures, are also clearly visible, when they are obtained with the aid of $\delta 2\text{D}$ server (Figure 1B). The average population of helical structures in region 4.2 reaches 0.6 for a 30% (v/v) TFE solution, while at 10% (v/v) TFE it approached only 0.4 (38), respectively 0.33 and 0.37 when estimated using $\delta 2\text{d}$. The average population of helical structures for the whole protein was estimated to be 0.25 (38) and 0.38 for 10 and 30% TFE solutions, and respectively 0.25 and 0.26, when calculated using $\delta 2\text{D}$ approach. These values are close to 0.35 estimated from CD spectrum (see Supplementary table S1), and substantially higher than 0.11 determined for the protein at low pH (35); 0.05 determined using $\delta 2\text{D}$. This clearly indicates a considerable increase in the propensity of rEC σ_4^{70} backbone towards the α -helical fold in proportion to the increasing TFE concentration in an aqueous solution.

A comparison of the location of helical regions in the TFE-induced structure of rEC σ_4^{70} with that found for its homologous domain of known structures (see grey vertical strips in Figure 1, also highlighted in further figures), indicates that the C-terminal fragment of the H1 helix is considerably frayed, while H2 and H3, organized into the HTH motif, are in a similar location. The inspection of the estimated distribution of helical conformations along the protein sequence shows that the H3 helix is the most stable one, and it definitively does not propagate behind L598. The short helix H2 became significantly stabilized only at the highest TFE concentration (30%), and the location of the loop region that separates H1 and H2 at a 30% TFE solution is similar to those seen in the crystal structures of the homologous proteins (27-30, 33). It is worth noting that, according to the $\delta 2\text{D}$ analysis (Figure 1B), the residues located in the turn region of HTH motif (D581-T583), which are in helical conformation at 10% TFE, became non-helical at higher TFE concentration. This perfectly agrees with the break of monotonic drift in HSQC spectra upon addition of TFE, previously observed for these residues (38). Altogether proves that 30% TFE does induce formation of transient HLHHTH structure, which is absent, when TFE concentration is lower than 15% (38). The global analysis of chemical shift patterns performed with TALOS+ (58) also confirmed the presence of four helical regions in rEC σ_4^{70} (Figure 1C.). The values of ϕ and ψ backbone torsion angles were predicted as 'good' for 66 out of 103 analyzed residues. Among them, 32 are located in the putatively helical regions corresponding to HLHHTH motif. The conformation for the 20 residues located in the N-terminal region of the protein was consistently attributed as dynamic. The next 17 residues, nine of which are located within HLHHTH motif, were excluded from the further analysis since their backbone conformation was predicted as uncertain. Interestingly, the three N-terminal residues of the helix H1 (L551, T552, A553) were also predicted in non-helical conformation, what, according to the reported 'good' quality of the prediction, indicates that a TFE-induced helix H1 may be for rEC σ_4^{70} shorter than these found in the known

1 structures of homologous proteins. Similarly, residues Y571, T572 and G577 located at both of the termini of H2 are
2 also predicted in non-helical conformation, what, together with CSI data (Figure 1A, B), proves that in $\text{rEC}\sigma^{70}_4$ H2 is
3 significantly shorter than these from homologous proteins. (27-30, 33)

4 **The internal motions of $\text{rEC}\sigma^{70}_4$ in a 30% TFE solution derived from ^{15}N relaxation data.** The R_1 and R_2
5 relaxation rates, and $\{^1\text{H}\}$ - ^{15}N NOEs, were determined for $\text{rEC}\sigma^{70}_4$ in a 30% TFE solution of at 11.4, 14.1, and 18.7 T
6 (see Supplementary Figure S4). As expected for substantially unfolded proteins, most of the $\{^1\text{H}\}$ - ^{15}N heteronuclear
7 NOEs do not exceed 0.5, while the values expected for the limited internal mobility of N-H vectors in the folded
8 proteins of a comparable size are of the order of 0.7 (90). However, the residues located in the putative HLHTH region
9 display visibly increased NOE values (Supplementary Figure S4C). Moreover, the variations in transverse relaxation
10 rates along the protein sequence are consistent with $\{^1\text{H}\}$ - ^{15}N NOEs profile, clearly identifying two regions that differ
11 significantly in their conformational flexibility. Elevated R_2 values determined for residues R534-L598 (varying in the
12 range of 15-45 s^{-1}) strictly coincide with the location of regions displaying propensities towards a TFE-induced helical
13 structure (Figure 1). It is worth noting that residues: S539, L540, T552, M561, R562, Y571, T572, L573, R586, I587
14 and R588, most of which is located in the putative HLHTH motif, display the substantially decreased rates of backbone
15 ^{15}N transverse relaxation, what indicates their increased flexibility. The ratio R_2/R_1 is commonly used for identification
16 of residues that experience motion of a timescale that differs from average molecular tumbling. Thus decreased values
17 of R_2/R_1 , which reflect the variation in R_2 , together with negative values of heteronuclear $\{^1\text{H}\}$ - ^{15}N NOEs, thus again
18 identify residues experiencing increased flexibility (see Figure 2A). And high R_2/R_1 values point to residues with
19 possible contributions of chemical exchange to the transverse relaxation. They may be also identified as these, which
20 display substantial field-dependence of $R_2^{-1/2}\cdot R_1$. Both methods pointed to D581, R584, Q589 and K593 (Figures 2A,
21 B), which were also identified as outliers in distribution of R_{ex} (Figure 2C). Two of them are located in the turn region
22 of HLHTF motif, while the two other (Q589, K593) are located just in the middle of H3.

23 **Internal $\text{rEC}\sigma^{70}_4$ dynamics deduced from the reduced spectral density mapping.** This analysis of ^{15}N relaxation
24 data does not required any assumption about the nature of molecular motion (65). The distributions of $J(\omega)$ along the
25 protein sequence are presented in Supplementary Figure S5A-C. The $J(0)$, which is most sensitive for slow motions on
26 the nanosecond time-scale (Supplementary Figure S5C), follows the trend observed for R_2 (Supplementary Figure
27 S4C). The residues T537, E538, T544 T545, K557, D581-R584, E591, K593 and A594, for which $J(0)$ substantially
28 varies with the magnetic field, may undergo slow conformational exchange processes on the μs -ms time-scale,
29 contribution of which to the rate of transverse relaxation (R_2) scales with the square of magnetic field. The extremely
30 low $J(0)$ values observed for residues located in N-terminal part of the protein of His₆-tag and linker fragment
31 preceding Leu 528 (average of 0.24 ns/rad), together with visibly high correlation times for fast motions, $J(\omega_{\text{H}})$
32 ($>20\text{ps/rad}$), show that this N-terminal part of the protein is almost unfolded. The same trend is observed for the C-
33 terminal part of the protein (i.e. for residues succeeding HLHTH motif), albeit visibly higher $J(0)$ (average of 0.9
34 ns/rad) and slightly lower $J(\omega_{\text{H}})$ suggest that motion of residues located in the C-terminal part of the protein is
35 somehow restricted. The central part of the $\text{rEC}\sigma^{70}_4$ display substantially higher values of $J(0)$ and smaller $J(\omega_{\text{H}})$. This
36 is indicative for the folded proteins, however the values of $J(\omega_{\text{H}})$ higher than 7.5 ps/rad are indicative for residues
37 experiencing significant internal flexibility (91). Interestingly, a few residues from HLHTH motif are substantially
38 much more flexible than the others. All of them, R562, Y571, T572 and F573 are located in the loop region of HLHTH
39 motif. In should be thus concluded that this loop region undergoes fast conformational sampling. Finally, as expected,
40 $J(\omega_{\text{N}})$ strongly varies with the strength of magnetic field.

Internal rEC σ_4^{70} dynamics estimated according to model-free approach. The simplest relaxation model describing contribution of slow and fast internal motions together with order parameter for the slow motion was used. In the case of poorly folded proteins, a correlation time for slow motions cannot be attributed to the correlation time characterizing overall rotation of the protein, but must be regarded as residue-dependent parameter. This makes challenging the problem of fitting $J(\omega)$ to the relaxation data. In general, only the simplest model (equation 4) was found reproducibly fitted upon several trial runs, while convergence all others strongly depended on starting values.

The distribution of the square of the generalized order parameter (S^2) along the protein sequence (Figure 3A) demonstrates that the mobility of all residues located within the HLHTH motif is definitively much more restricted. It should be emphasized that the trend in S^2 variation, estimated from chemical shift data (a thick broken line in Figure 3A), follows that obtained directly from the ^{15}N relaxation data. However, the local extrema do not strictly coincide with the expected location of helical regions indicated by CSI analysis (see Figure 1A) and the predicted backbone angles (see Figure 1B), identified in the structures of homologous proteins. S^2 values for residues located in HLHTH region (Table 1) are visibly smaller than those usually found for properly folded proteins ($S^2 > 0.9$), however exceed the value of 0.6 observed for TFE-induced helix of calmodulin in 35% TFE aqueous solution (90).

The distribution of local correlation times for slow motions (τ_{local} ; cf. Figure 3B) remains in a complete agreement with the conclusion derived from the data presented in Figures 1 and 2. Again, the four regions of rEC σ_4^{70} (L532-D546, R554-M561, E574-D581 and R584-R599), characterized by the elevated values of τ_{local} , are clearly visible. A considerable simultaneous decrease in S^2 and τ_{local} values observed for some residues located between D566 and L573 additionally supports the increased flexibility of this loop region. The median of 11.6 ns/rad (see Table 1) is visibly higher than that expected for the proteins of a similar size, what may result from increased viscosity of TFE-water mixed solvents (92). The possible contribution of slow conformational motions identified for residues T537, E538, T544 T545, K557, D581-R584, E591, K593 and A594 may also resulted in overestimation of the τ_{local} values.

The distribution of local correlation times for fast motions, τ_e (Figure 3C) is strongly diverged, and the values estimated for HLHTH region are close to 30 ps.

The low-resolution structure of rEC σ_4^{70} in a 30% TFE aqueous solution from the NOESY data. 1237 NOE contacts were assigned in ^{13}C -, and ^{15}N -*edited* 3D-NOESY spectra recorded at a 30% (v/v) TFE. This includes 778 intraresidual, 330 sequential, 106 medium-range ($1 \leq |i - j| \leq 4$) and 23 long-range ($|i - j| > 4$) ones. The number of medium-range contacts (see Table 2) significantly exceeds that previously identified at 10% (v/v) TFE (106 vs. 15, respectively). However, most importantly, 23 long-range structural contacts between the regions that display a high population of secondary structure elements have been unequivocally assigned (see Supplementary Table S2 and Supplementary Figure S6). These contacts occur mainly between helical regions (e. g. H2-H3), but some involving loop regions (e. g. H1-L) can also be observed (see Table 2). Altogether, it clearly evidences the formation of, at least transient, a three-dimensional **HLHTH** structure. Such a small number of long-range contacts precluded any reasonable ensemble-based interpretation of NMR data (93), and only representative low-resolution structures that generally agree with CSI and NOE data.

As shown in Table 2 and Supplementary Table S2, the assignment of NOESY medium- and long-range cross-peaks, together with the relaxation data (Figures 2-3, S4, S5) and CSI analysis (Figure 1), unequivocally indicate that rEC σ_4^{70} contains four helical regions, separated by two loops and one short turn, similarly as it has been found in the solution structure of homologous domains. The location of these helical regions was further constrained in the

1 structural modeling to access the agreement between the TFE-induced transient fold of rEC σ^70_4 with NOESY-derived
2 distance constraints (see Supplementary Table S3 for short statistics).

3 The ensemble of the resulting structures is presented in Figure 4A. The arrangement of the H1 and H3 remains
4 the most stable (see Figure 4B), providing a kind of framework around which the more flexible LHT motif (i.e. loop
5 and the short H2 region, followed by the turn) is organized in the form resembling these of TM σ^A_4 (Figure 4C) and
6 recently solved for EC σ^70 the whole RNAP holoenzyme (Figure 4D) (30). The average RMSD calculated for backbone
7 heavy atoms for the H1 and H3 helical regions within the cluster of the 11 lowest-(pseudo)energy structures was
8 reduced to 2.1Å. The increased mobility of the LHT region may also be related to its physiological function (see
9 Discussion for details).

10 **The MD study of the structural stability of a TFE-induced HLHTH fold of rEC σ^70_4 .** The conformational screening
11 was performed by eleven 5 ns MD runs, and each of them started from one of the 11 lowest-(pseudo)energy structures.
12 Most of these trajectories preserved the initial protein fold, but during one of them a visible reorganization of the
13 HLHTH motif, involving the simultaneous breaking of the H1 helix at R554-E555 and of the H3 helix at E591-A592
14 was observed. Both locations agree with the increased contribution of the chemical exchange process to ^{15}N relaxation
15 rates found for residues and E591-R593 (Figure 2).

22 DISCUSSION

23 According to our earlier studies (35, 38) and all the data presented in this work, rEC σ^70_4 in an aqueous solution
24 must be regarded an intrinsically disordered, but capable of transient folding in the HLHTH motif in the presence of the
25 30% of TFE. In order to check to what extent the TFE-induced solution structure of rEC σ^70_4 resembles that of the
26 homologous proteins, the most stable regions formed by the H1 and H3 helices of these two structures were compared
27 (Figure 4C, D). The backbone architecture of these two helices identified by us in NMR-derived structure of rEC σ^70_4 is
28 close to those of TM σ^A_4 in solution and EC σ^70_4 in crystal structure of RNAP complex (31), but the RMSD difference
29 for the backbone atoms between the modeled solution structure and each of these two structures exceeds 10 Å.
30 However, this difference is smaller than 2 Å, when calculated only for H1 and H3, confirming the high structural
31 homology of these regions (see Figure 4 C, D). It has to be stressed that the sequences of all known reference σ^70_4
32 structures differ noticeably in their loop regions of the canonical HLHTH DNA-recognizing motif (I565-Y571 in *E.*
33 *coli*). This loop is involved in the proper orientation of the recognition helix, interacting with the DNA major groove,
34 and is also directly involved in an interaction with promoters (34, 94).

35 **The internal dynamics of the TFE-induced fold of rEC σ^70_4 .** The Anderson-Darling goodness-of-fit test
36 demonstrated that distribution of S^2 , τ_{local} and τ_e estimated for rEC σ^70_4 are generally not Gaussian. The resulting
37 medians accompanied by quartiles are presented in Table 1. For comparison corresponding mean values and standard
38 deviations are also presented. All the distribution were compared with the aid of nonparametric Mann-Whitney U-test.

39 The global effect of TFE concentration on ^{15}N relaxation is clearly visible (Figure 5). S^2 increases with the
40 content of this agent (Table 1), reaching the maximum at 30% TFE (v/v). Four helical regions indicated by the CSI
41 analysis (see Figure 1 for backbone angles estimated with TALOS+) agree with the distribution of relaxation
42 parameters along a protein sequence (Figure 3). It could be thus concluded that the extent of a transient fold of the
43 protein increases upon TFE titration. According to the nonparametric Mann-Whitney U-test (MW-test) the distribution
44 of S^2 differ significantly ($p < 10^{-6}$) for the all three protein regions, and the differences are statistically significant for all
45 conditions tested (low-pH, 10% TFE, 30% TFE, see Table 1 details and Table 3 for exact p-values for comparisons)

1 The medians for effective local correlation time for fast motions, τ_e , calculated for the HLHTH region at 0, 10
2 and 30% (v/v) TFE concentrations equal to 48ps, 31ps, and 28ps, respectively, thus confirming a TFE-induced
3 restriction of fast motions within the HLHTH motif of $rEC\sigma_{70}^4$. The distribution of slow motions (τ_{local} ; Figure 3B)
4 remains in a full agreement with the results obtained for S^2 and τ_e . Again, the four regions of $rEC\sigma_{70}^4$ (L532-D546,
5 A549-R560, V576- R582 and E585-R599) can be distinguished by the elevated τ_{local} . The medians for τ_{local} are: 4.8 ns,
6 9.5 ns, and 11.6 ns for 0%, 10% and 30% (v/v) TFE concentrations, respectively.
7

8 Majority of these distributions differ significantly (see Table 3), proving that TFE addition does induce, at
9 least transient, protein structure, as indicated by elevated S^2 and restricted fast motions. A considerable simultaneous
10 decrease in S^2 and τ_{local} values, observed for residues D566-L573, identify a flexible structure located at the same
11 protein sequence where the putative loop region of HLHTH motif .
12

13 The above data clearly evidence that the addition of TFE to a protein solution stabilizes the HLHTH motif
14 much more than that of the other region of the protein, as proved by the direct comparison of HLHTF motif with
15 flanking residues (others in Table 3). Generally, the distributions of S^2 and both correlation times estimated for residues
16 located in the HLHTH motif significantly differ from those calculated separately for the peripheral residues located
17 outside of this motif. However, it should be noted that all the TFE-related conformational changes generally lead to an
18 increase in the population of the ordered structure, even for the residues outside of the HLHTH motif. This confirms
19 that TFE acts as a non-specific structure-inducer, stabilizing the local secondary structures of folded proteins (95-97),
20 including α -helices (98), β -turns (99) and β -hairpins (100). The thermodynamic effect of TFE should be connected with
21 weakened non-local hydrophobic interactions, accompanied by slightly enhanced local ones, thus propagating context-
22 dependent formation of a native-like secondary structure in polypeptides (101). This reflected in an increase of the
23 order parameter, S^2 , accompanied by a moderate decrease of the local correlation time for fast motions, τ_e . The
24 structure-inducing effect observed for $rEC\sigma_{70}^4$ is definitively not uniform, but follows sequence-dependent propensities
25 of the backbone already visible in a TFE-free low-pH protein solution (35).
26

27 All the results referred to above clearly demonstrate that IDPs, even in the absence of physiological partners,
28 may obtain their functional folds. And these states can be sampled with a variety of commonly used stabilizing agents
29 like $MgSO_4$, arginine, proline, sucrose, sarcosine, glycerol, NDSBs, TMAO or even by TFE, which also may act as a
30 stabilizer (95).
31

32
33
34
35
36
37
38
39
40
41
42
43 **ACKNOWLEDGMENTS.** This work was partially supported by the grant CRP/08/11 founded by the International
44 Centre for Genetic Engineering and Biotechnology, Trieste, Italy.
45

46 47 **REFERENCES.**

- 48 1. Pancsa, R., and P. Tompa. 2012. Structural Disorder in Eukaryotes. *PLoS ONE* 7:e34687.
- 49 2. Uversky, V. N. 2002. What does it mean to be natively unfolded? *European Journal of Biochemistry* 269:2-12.
- 50 3. Uversky, V. N. 2011. Intrinsically disordered proteins from A to Z. *The international journal of biochemistry &*
51 *cell biology* 43:1090-1103.
- 52 4. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear
53 sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the*
54 *United States of America* 110:13392-13397.
- 55 5. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nature Reviews*
56 *Molecular Cell Biology* 6:197-208.
- 57 6. Tompa, P. 2002. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* 27:527-533.
- 58 7. Tompa, P. 2005. The interplay between structure and function in intrinsically unstructured proteins. *Febs Letters*
59 *579:3346-3354.*
60
61
62
63
64
65

8. Wright, P. E., and H. J. Dyson. 2009. Linking folding and binding. *Current Opinion in Structural Biology* 19:31-38.
9. Dima, R. I., and D. Thirumalai. 2004. Asymmetry in the shapes of folded and denatured states of proteins. *Journal of Physical Chemistry B* 108:6564-6570.
10. Fu, X. D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA* 1:663-680.
11. Buljan, M., G. Chalancon, S. Eustermann, G. P. Wagner, M. Fuxreiter, A. Bateman, and M. M. Babu. 2012. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Molecular cell* 46:871-883.
12. Rautureau, G. J. P., C. L. Day, and M. G. Hinds. 2010. Intrinsically Disordered Proteins in Bcl-2 Regulated Apoptosis. *International Journal of Molecular Sciences* 11:1808-1824.
13. Galea, C. A., Y. Wang, S. G. Sivakolundu, and R. W. Kriwacki. 2008. Regulation of Cell Division by Intrinsically Unstructured Proteins: Intrinsic Flexibility, Modularity, and Signaling Conduits†. *Biochemistry* 47:7598-7609.
14. Moncoq, K., I. Broutin, V. Larue, D. Perdereau, K. Cailliau, E. Browaeys-Poly, A. F. Burnol, and A. Ducruix. 2003. The PIR domain of Grb14 is an intrinsically unstructured protein: implication in insulin signaling. *Febs Letters* 554:240-246.
15. Liu, J. G., N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. 2006. Intrinsic disorder in transcription factors. *Biochemistry* 45:6873-6888.
16. Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. *Nature Genetics* 31:69-73.
17. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* 293:321-331.
18. Burgess, R. R., A. A. Travers, J. J. Dunn, and E. K. F. Bautz. 1969. Factor stimulating transcription by RNA polymerase. *Nature* 221:43-46.
19. Losick, R., and J. Pero. 1981. Cascades of sigma factors. *Cell* 25:582-584.
20. Maeda, H., N. Fujita, and A. Ishihama. 2000. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research* 28:3497-3503.
21. Grigorova, I. L., N. J. Phleger, V. K. Mutalik, and C. A. Gross. 2006. Insights into transcriptional regulation and If competition from an equilibrium model of RNA polymerase binding to DNA. *Proceedings of the National Academy of Sciences of the United States of America* 103:5332-5337.
22. Nyström, T. 2004. Growth versus maintenance: A trade-off dictated by RNA polymerase availability and sigma factor competition? *Molecular Microbiology* 54:855-862.
23. Jishage, M., A. Iwata, S. Ueda, and A. Ishihama. 1996. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular levels of four species of sigma subunit under various growth conditions. *Journal of Bacteriology* 178:5447-5451.
24. Raffaele, M., E. I. Kanin, J. Vogt, R. R. Burgess, and A. Z. Ansari. 2005. Holoenzyme switching and stochastic release of sigma factors from RNA polymerase in vivo. *Molecular Cell* 20:357-366.
25. Camarero, J. A., A. Shekhtman, E. A. Campbell, M. Chlenov, T. M. Gruber, D. A. Bryant, S. A. Darst, D. Cowburn, and T. W. Muir. 2002. Autoregulation of a bacterial σ factor explored by using segmental isotopic labeling and NMR. *Proceedings of the National Academy of Sciences of the United States of America* 99:8536-8541.
26. Typas, A., and R. Hengge. 2006. Role of the spacer between the -35 and -10 regions in Ifs promoter selectivity in *Escherichia coli*. *Molecular Microbiology* 59:1037-1051.
27. Lambert, L. J., Y. Wei, V. Schirf, B. Demeler, and M. H. Werner. 2004. T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J* 23:2952-2962.
28. Patikoglou, G. A., L. F. Westblade, E. A. Campbell, V. Lamour, W. J. Lane, and S. A. Darst. 2007. Crystal structure of the *Escherichia coli* regulator of sigma70, Rsd, in complex with sigma70 domain 4. *J Mol Biol* 372:649-659.
29. Blanco, A. G., A. Canals, J. Bernues, M. Sola, and M. Coll. 2011. The structure of a transcription activation subcomplex reveals how sigma(70) is recruited to PhoB promoters. *EMBO J* 30:3776-3785.
30. Murakami, K. S. 2013. X-ray Crystal Structure of *Escherichia coli* RNA Polymerase sigma(70) Holoenzyme. *Journal of Biological Chemistry* 288:9126-9134.
31. Molodtsov, V., I. N. Nawarathne, N. T. Scharf, P. D. Kirchhoff, H. D. H. Showalter, G. A. Garcia, and K. S. Murakami. 2013. X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem* 56:4758-4763.
32. Mechold, U., K. Potrykus, H. Murphy, K. S. Murakami, and M. Cashel. 2013. Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. *Nucleic Acids Res* 41:6175-6189.
33. Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz, and S. A. Darst. 2002. Structure of the bacterial RNA polymerase promoter specificity If subunit. *Molecular Cell* 9:527-539.
34. Campbell, E. A., O. Muzzin, M. Chlenov, J. L. Sun, C. A. Olson, O. Weinman, M. L. Trester-Zedlitz, and S. A. Darst. 2002. Structure of the bacterial RNA polymerase promoter specificity σ subunit. *Molecular Cell* 9:527-539.

35. Poznanski, J., K. Bolewska, I. Zhukov, and K. L. Wierzchowski. 2003. Characterization of the low pH solution structure and dynamics of the region 4 of Escherichia coli RNA polymerase sigma(70) subunit. *Biochemistry* 42:13438-13448.
36. Jaravine, V. A., A. T. Alexandrescu, and S. Grzesiek. 2001. Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide. *Protein Science* 10:943-950.
37. Lawrence, J. R., and W. C. Johnson. 2002. Lifson-Roig nucleation for α -helices in trifluoroethanol: Context has a strong effect on the helical propensity of amino acids. *Biophysical Chemistry* 101-102:375-385.
38. Kaczka, P., A. Polkowska-Nowakowska, K. Bolewska, I. Zhukov, J. Poznanski, and K. L. Wierzchowski. 2010. Backbone dynamics of TFE-induced native-like fold of region 4 of Escherichia coli RNA polymerase sigma(70) subunit. *Proteins-Structure Function and Bioinformatics* 78:754-768.
39. Vassilyev, D. G., S. I. Sekine, O. Laptenko, J. Lee, M. N. Vassilyeva, S. Borukhov, and S. Yokoyama. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417:712-719.
40. Doucleff, M., L. T. Malak, J. G. Pelton, and D. E. Wemmer. 2005. The C-terminal RpoN domain of σ 54 forms an unpredicted helix-turn-helix motif similar to domains of σ 70. *Journal of Biological Chemistry* 280:41530-41536.
41. Poznanski, J., I. Zhukov, K. Bolewska, and K. L. Wierzchowski. 2001. Sequence-specific H-1, N-15, and C-13 resonance assignments for the whole region 4 of Escherichia coli RNA polymerase sigma(70) subunit. *Journal of Biomolecular Nmr* 20:181-182.
42. Gerald, B. 1997. CDNN. Institut für Biotechnologie, Martin-Luther-Universität, Halle-Wittenberg, Germany.
43. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6:277-293.
44. Goddard, T. D., and D. G. Kneller. 2002. SPARKY 3. University of California, San Francisco. <http://www.cgl.ucsf.edu/home/sparky>.
45. Wittekind, M., and L. Mueller. 1993. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance, Series B* 101:201-205.
46. Grzesiek, S., and A. Bax. 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* 114:6291-6293.
47. Gardner, K. H., R. Konrat, M. K. Rosen, and L. E. Kay. 1996. An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated ¹⁵N,¹³C-labeled proteins. *Journal of Biomolecular NMR* 8:351-356.
48. Kuboniwa, H., S. Grzesiek, F. Delaglio, and A. Bax. 1994. Measurement of H N-H α J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *Journal of Biomolecular NMR* 4:871-878.
49. Wang, A. C., P. J. Lodi, J. Qin, G. W. Vuister, A. M. Gronenborn, and G. M. Clore. 1994. An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins. *Journal of Magnetic Resonance, Series B* 105:196-198.
50. Grzesiek, S., J. Anglister, and A. Bax. 1993. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in ¹³C/¹⁵N-Enriched Proteins by Isotropic Mixing of ¹³C Magnetization. *Journal of Magnetic Resonance, Series B* 101:114-119.
51. Grzesiek, S., and A. Bax. 1992. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance* 96:432-440.
52. Clubb, R. T., V. Thanabal, and G. Wagner. 1992. A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13 Chemical-Shifts in N-15-C-13-Labeled Proteins. *Journal of Magnetic Resonance* 97:213-217.
53. Schwarzing, S., G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. 2001. Sequence-dependent correction of random coil NMR chemical shifts. *Journal of the American Chemical Society* 123:2970-2978.
54. Wishart, D. S., and B. D. Sykes. 1994. The ¹³C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *Journal of Biomolecular NMR* 4:171-180.
55. Schwarzing, S., G. J. A. Kroon, T. R. Foss, P. E. Wright, and H. J. Dyson. 2000. Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. *Journal of Biomolecular Nmr* 18:43-48.
56. δ 2D server; Aviable: <http://www.vendruscolo.ch.cam.ac.uk/d2D/>; Jobs submitted: July 2014.
57. Camilloni, C., A. De Simone, W. F. Vranken, and M. Vendruscolo. 2012. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* 51:2224-2231.
58. Shen, Y., F. Delaglio, G. Cornilescu, and A. Bax. 2009. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* 44:213-223.
59. Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. 1994. Backbone dynamics of a free and a phosphopeptide-complexed src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 33:5984-6003.
60. Meiboom, S., and D. Gill. 1958. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments* 29:688-691.

- 1 61. Kay, L. E., L. K. Nicholson, F. Delaglio, A. Bax, and D. A. Torchia. 1992. Pulse sequences for removal of the
2 effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the
3 measurement of heteronuclear T1 and T2 values in proteins. *J. Biomol. NMR* 97:359–375.
- 4 62. Shaka, A. J. 1985. Computer-optimized decoupling scheme for wideband applications and low-level operation.
5 *J. Magn. Reson.* 64:547–552.
- 6 63. Kowalewski, J. 1978. Determination of NOE factors using the dynamic overhauser enhancement technique
7 combined with a nonlinear least-squares-fitting procedure. *J. Magn. Reson.* 31:165–169.
- 8 64. Zhukov, I., and A. Eijchart. 1999. Factors improving the accuracy of determination of ¹⁵N relaxation parameters in
9 proteins. *Acta Biochimica Polonica* 46:665-671.
- 10 65. Farrow, N. A., O. W. Zhang, A. Szabo, D. A. Torchia, and L. E. Kay. 1995. Spectral density function mapping
11 using ¹⁵N relaxation data exclusively. *Journal of Biomolecular Nmr* 6:153-162.
- 12 66. Korzhnev, D. M., M. Billeter, A. S. Arseniev, and V. Y. Orekhov. 2001. NMR studies of Brownian tumbling and
13 internal motions in proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* 38:197-266.
- 14 67. Stone, M. J., W. J. Fairbrother, A. G. Palmer III, J. Reizer, M. H. Saier Jr, and P. E. Wright. 1992. Backbone
15 dynamics of the *Bacillus subtilis* glucose permease IIA domain determined from ¹⁵N NMR relaxation
16 measurements. *Biochemistry* 31:4394-4406.
- 17 68. Peirce, B. 1852. Criterion for the Rejection of Doubtful Observations. *Astronomical Journal* 45:161-163.
- 18 69. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation
19 in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society* 104:4559-4570.
- 20 70. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation
21 in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* 104:4546-4559.
- 22 71. Alexandrescu, A. T., and D. Shortle. 1994. Backbone dynamics of a highly disordered 131 residue fragment of
23 staphylococcal nuclease. *Journal of Molecular Biology* 242:527-546.
- 24 72. Brutscher, B., R. Brüschweiler, and R. R. Ernst. 1997. Backbone dynamics and structural characterization of the
25 partially folded A state of ubiquitin by ¹H, ¹³C, and ¹⁵N nuclear magnetic resonance spectroscopy. *Biochemistry*
26 36:13043-13053.
- 27 73. Talluri, S., and G. Wagner. 1996. An optimized 3D NOESY-HSQC. *Journal of Magnetic Resonance Series B*
28 112:200-205.
- 29 74. Muhandiram, D. R., N. A. Farrow, G. Y. Xu, S. H. Smallcombe, and L. E. Kay. 1993. A gradient C-13 NOESY-
30 HSQC experiment for recording NOESY spectra of C-13-labeled proteins dissolved in H₂O. *Journal of Magnetic*
31 *Resonance Series B* 102:317-321.
- 32 75. Guntert, P. 2004. Automated NMR structure calculation with CYANA. *Methods in molecular biology* (Clifton,
33 N.J.) 278:353-378.
- 34 76. Schwieters, C. D., J. J. Kuszewski, N. Tjandra, and G. M. Clore. 2003. The Xplor-NIH NMR molecular structure
35 determination package. *Journal of Magnetic Resonance* 160:65-73.
- 36 77. Brunger, A. T. 1992. XPLOR. A system for X-ray Crystallography and NMR, Yale University Press, New Haven.
- 37 78. Kuszewski, J., A. M. Gronenborn, and G. M. Clore. 1995. The impact of direct refinement against proton
38 chemical-shifts on protein-structure determination by NMR. *Journal of Magnetic Resonance Series B* 107:293-
39 297.
- 40 79. Kuszewski, J., J. Qin, A. M. Gronenborn, and G. M. Clore. 1995. The impact of direct refinement against C-
41 ¹³(alpha) and C-¹³(beta) chemical-shifts on protein-structure determination by NMR. *Journal of Magnetic*
42 *Resonance Series B* 106:92-96.
- 43 80. Laskowski, R. A., J. A. C. Rullmann, M. W. MacArthur, R. Kaptein, and J. M. Thornton. 1996. AQUA and
44 PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of*
45 *Biomolecular Nmr* 8:477-486.
- 46 81. Krieger, E., G. Koraimann, and G. Vriend. 2002. Increasing the precision of comparative models with YASARA
47 NOVA - a self-parameterizing force field. *Proteins-Structure Function and Genetics* 47:393-402.
- 48 82. Anderson, T. W., and D. A. Darling. 1952. Asymptotic theory of certain goodness of fit criteria based on stochastic
49 processes. *Annals of Mathematical Statistics* 23:193-212.
- 50 83. Mann, H. B., and D. R. Whitney. 1947. On a test of whether one of two random variables is stochastically larger
51 than the other. *Annals of Mathematical Statistics* 18:50-60.
- 52 84. 2011. STATISTICA (data analysis software system) version 10. StatSoft, Inc.
- 53 85. Williams, T., and C. Kelley. 2007. <http://www.gnuplot.info>.
- 54 86. Kozlowski, L. 2007-2012. Isoelectric Point Calculator.
- 55 87. Buchner, J., I. Pastan, and U. Brinkmann. 1992. A method for increasing the yield of properly folded recombinant
56 fusion proteins - single-chain immunotoxins from renaturation of bacterial inclusion-bodies *Analytical*
57 *Biochemistry* 205:263-270.
- 58 88. Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133-7155.
- 59 89. Copeland, R. A. 2000. Experimental Measures of Enzyme Activity. In *Enzymes*. John Wiley & Sons, Inc. 188-
60 265.

- 1 90. Brokx, R. D., R. M. Scheek, A. M. Weljie, and H. J. Vogel. 2004. Backbone dynamic properties of the central
2 linker region of calcium-calmodulin in 35% trifluoroethanol. *Journal of Structural Biology* 146:272-280.
- 3 91. Viles, J. H., D. Donne, G. Kroon, S. B. Prusiner, F. E. Cohen, H. J. Dyson, and P. E. Wright. 2001. Local structural
4 plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochemistry* 40:2743-2753.
- 5 92. Olive, F., S. K. Chaudhari, K. R. Patil, and A. Coronas. 1996. Viscosity of the binary systems containing
6 trifluoroethanol, water and tetraethylene glycol dimethyl ether. Prediction of the ternary viscosity from binary data.
7 *Canadian Journal of Chemical Engineering* 74:163-169.
- 8 93. Rezaei-Ghaleh, N., M. Blackledge, and M. Zweckstetter. 2012. Intrinsically Disordered Proteins: From Sequence
9 and Conformational Properties toward Drug Discovery. *Chembiochem* 13:930-950.
- 10 94. Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin, and S. A. Darst. 2002. Structural basis of transcription
11 initiation: An RNA polymerase holoenzyme-DNA complex. *Science* 296:1285-1290.
- 12 95. Luo, Y. Z., and R. L. Baldwin. 1998. Trifluoroethanol stabilizes the pH 4 folding intermediate of sperm whale
13 apomyoglobin. *Journal of Molecular Biology* 279:49-57.
- 14 96. Bhakuni, V. 1998. Alcohol-Induced Molten Globule Intermediates of Proteins: Are They Real Folding
15 Intermediates or Off Pathway Products? *357:274-284*.
- 16 97. Shiraki, K., K. Nishikawa, and Y. Goto. 1995. Trifluoroethanol-induced stabilization of the alpha-helical structure
17 of beta-lactoglobulin - implication for non-hierarchical protein-folding. *Journal of Molecular Biology* 245:180-
18 194.
- 19 98. Nelson, J. W., and N. R. Kallenbach. 1989. Persistence of the .alpha.-helix stop signal in the S-peptide in
20 trifluoroethanol solutions. *Biochemistry* 28:5256-5261.
- 21 99. Cann, J. R., R. E. London, C. J. Unkefer, R. J. Vavrek, and J. M. Stewart. 1987. CD-NMR study of the solution
22 conformation of bradykinin analogs containing alpha-aminoisobutyric-acid. *International Journal of Peptide and
23 Protein Research* 29:486-496.
- 24 100. Ramirez-Alvarado, M., L. Serrano, and F. J. Blanco. 1997. Conformational analysis of peptides corresponding to
25 all the secondary structure elements of protein L B1 domain: Secondary structure propensities are not conserved in
26 proteins with the same fold. *6:162-174*.
- 27 101. Thomas, P. D., and K. A. Dill. 1993. Local and nonlocal interactions in globular-proteins and mechanisms of
28 alcohol denaturation. *Protein Science* 2:2050-2065.

Table 1. Relaxation parameters determined for $rEC\sigma^{70}_4$ in an acidic solution, and for a 10% and a 30% TFE solution at pH 4.55, calculated separately for the residues located in N-terminal His₆-tag and succeeding linker, in HLHTH region and both regions flanking HLHTH motif (others). All the medians are accompanied by lower and upper quartiles, and corresponding mean values are accompanied by the standard deviations. Since some of the distributions deviated, according to the Anderson-Darling test, from the normal ones, they were compared using appropriate nonparametric tests (see Table 3 and methods for details).

conditions	n	S ²		τ_{local} [ns]		τ_e [ps]	
		Median	Mean	Median	Mean	Median	Mean
His ₆ -tag + linker (residues preceding Leu 528)							
low-pH	19	0.07 (0.05;0.09)	0.07 (\pm 0.03)	3.2 (2.3;4.3)	4.1 (\pm 2.8)	67 (61;83)	72 (\pm 15)
10% TFE	21	0.11 (0.10;0.12)	0.11 (\pm 0.04)	4.0 (3.3;4.6)	4.1 (\pm 1.0)	66 (58;72)	66 (\pm 12)
30% TFE	21	0.36 (0.30;0.45)	0.37 (\pm 0.10)	1.0 (0.8;2.6)	1.7 (\pm 1.1)	34 (16;75)	97 (\pm 143)
HLHTH (Ala 549-Arg 599)							
low-pH	46	0.17 (0.16;0.19)	0.18 (\pm 0.03)	4.8 (4.3; 5.2)	4.9 (\pm 0.8)	48 (45;52)	49 (\pm 5)
10% TFE	49	0.32 (0.29;0.34)	0.32 (\pm 0.03)	9.5 (8.6; 9.9)	9.3 (\pm 1.0)	31 (26;36)	31 (\pm 6)
30% TFE	47	0.65 (0.58;0.69)	0.61 (\pm 0.12)	11.6 (9.9;12.9)	10.6 (\pm 3.4)	28 (16;34)	26 (\pm 16)
others (L528-L548 and H600-D613)							
low-pH	27	0.12 (0.11;0.12)	0.11 (\pm 0.01)	3.4 (3.0;3.8)	3.5 (\pm 0.5)	57 (55;62)	59 (\pm 7)
10% TFE	30	0.23 (0.20;0.28)	0.23 (\pm 0.05)	6.8 (6.0;8.5)	7.0 (\pm 1.5)	46 (39;53)	45 (\pm 12)
30% TFE	28	0.45 (0.43;0.59)	0.49 (\pm 0.10)	5.5 (5.1;9.1)	6.8 (\pm 2.3)	33 (28;36)	32 (\pm 8)

Table 2. Long-range cross-peaks unequivocally assigned in ¹⁵N- and ¹³C-edited NOESY spectra recorded for $rEC\sigma^{70}_4$ at pH 4.55 in 30% (v/v) TFE solution.

	¹⁵ N-edited 3D-NOESY mixing time: 150 ms	¹³ C-edited 3D-NOESY mixing time: 90ms	¹³ C edited 3D-NOESY mixing time: 250ms
H1-L		I565H β -R560H β 2,3	
H1-H2			V576H β -M561H α , γ 3
H1-H3	E591HN-L559H α	I587H γ 2-V558H γ 1	I590H α , β -E555H α ,
H2-H3	R584HN-G577HN; T583HN- E575H α , β 1, β 2, γ 1, γ 2; T583HN-K578H α , β 1, γ , ϵ	I565H β -V576H β	I587H α -V576H γ 1, γ 2

Table 3. The comparison of the estimated distributions of relaxation parameters determined for $rEC\sigma^{70}_4$ in various conditions made with the aid of Mann-Whitney test. The IDs corresponds to the data presented in Table 1.

IDs	compared distributions					
	HLHTH vs. others			TFE 30% vs. TFE 10%		
conditions	low-pH	10% TFE	30% TFE	His ₆ -tag	HLHTH	others
n1,n2	46, 27	49, 30	47, 28	21, 21	47, 49	28, 30
parameter	p-values for comparison					
S ²	< 10 ⁻⁶	< 10 ⁻⁶	< 10 ⁻⁶	1.6·10 ⁻⁴	< 10 ⁻⁶	< 10 ⁻⁶
τ_{local} [ns]	< 10 ⁻⁶	< 10 ⁻⁶	< 10 ⁻⁶	0.13	< 10 ⁻⁶	0.44
τ_e [ps]	< 10 ⁻⁶	< 10 ⁻⁶	1.7·10 ⁻³	0.32	1.6·10 ⁻³	< 10 ⁻⁶

Figure Legends

1
2
3 **Figure 1.** The distribution of a TFE-induced secondary structure along the rEC σ_4^{70} sequence. The population of α -
4 helical conformation was estimated from CSI data (A) for an acidic aqueous solution (pH 2.8, white bars), 10%
5 (v/v) TFE (pH 4.5, grey bars) and 30% (v/v) TFE (pH 4.6, in black). In parallel (B), the backbone angles phi (ϕ),
6 and psi (ψ) were predicted with the aid of TALOS+ for rEC σ_4^{70} in a 30% TFE solution at pH 4.6. Grey vertical
7 strips mark helical regions found in the structure of σ_4^{70} homologues from thermophilic bacteria, and the lighter
8 strips mark additional helical regions found in *E. coli* σ_4^{70} in the RNAP complex. The size of the error bars
9 represents the standard deviation from the average of the dihedral angles. Grey and white markers point residues,
10 for which conformation estimated by TALOS+ was reported as either dynamic or uncertain, respectively. Two of the
11 latter, namely A594 and V606 were most likely wrongly predicted in non-helical conformation.

12
13 **Figure 2.** The distributions of R2/R1 ratio along protein sequence determined for rEC σ_4^{70} in a 30% TFE solution (A),
14 and chemical exchange contribution, R_{ex} , estimated at 11.4 T according to equation 3 (B). White, grey and black
15 markers in panel A identify the values determined at 11.4, 14.1 and 18.7T, respectively. Vertical bars represent
16 estimates of the experimental error. Grey strips mark locations of the helical segments of the HLHTH motif. Grey
17 circles mark outliers in cumulative distribution of R_{ex} (C).

18
19 **Figure 3.** The interpretation of rEC σ_4^{70} NMR relaxation according to Model-Free approach optimized for IDP proteins
20 (see equations 2A-C). Grey vertical strips mark helical regions present in HLHTH motif. Structured regions
21 stabilized by TFE exhibit the increased values of the order parameter S^2 , indicative of the build-up of at least
22 transient 3D-structure. The distributions of both the generalized order parameter, S^2 , and the corresponding slow-
23 motion correlation times, τ_{local} , indicate that the N-terminal 21-residue fragment is almost unfolded, and the mobility
24 of the residues located in the central HLHTH motif are generally much more restricted than those located outside of
25 this motif. The thick broken line follows the S^2 values estimated on the basis of the experimental chemical shifts of
26 ^{13}C , ^{15}N , $^{13}C\alpha$, $^{13}C\beta$ and $^1H\alpha$ resonances according to the RCI algorithm included in TALOS+. For comparison the
27 data determined previously for low-pH solution (gray dotted line) and 10%TFE (black line following diamonds) are
28 also presented.

29
30 **Figure 4.** The transient solution structure of rEC σ_4^{70} at a 30% (v/v) TFE concentration. The ribbon representation of the
31 ensemble of 11 lowest-energy structures superposed for the H1 and H3 helical regions (A), HLHTH motifs of the
32 latter (B) and superposition of the representative transient low-energy structure with the solution structure of *T.*
33 *maritima* σ^A (C) and just published crystal structure of *E. coli* σ_4^{70} (D) are presented.

34
35 **Figure 5.** The cumulative distribution of generalized order parameter S^2 (A) and correlation times for slow (B) and fast
36 (C) correlation times determined for EC σ_4^{70} at low-pH solution (open), and 10% (grey) and 30% (black) TFE
37 concentration.

Figure 1.

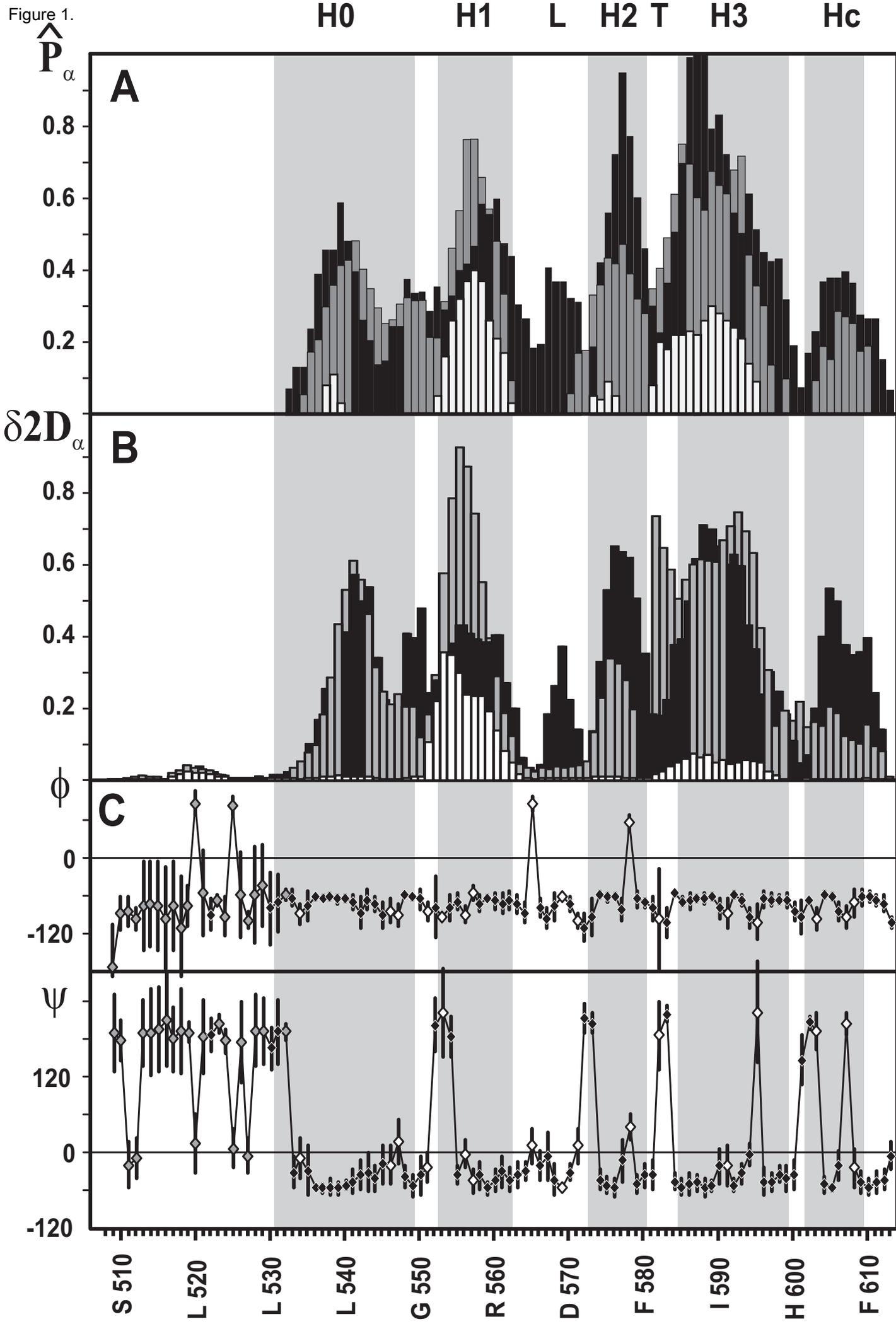


Figure 2.

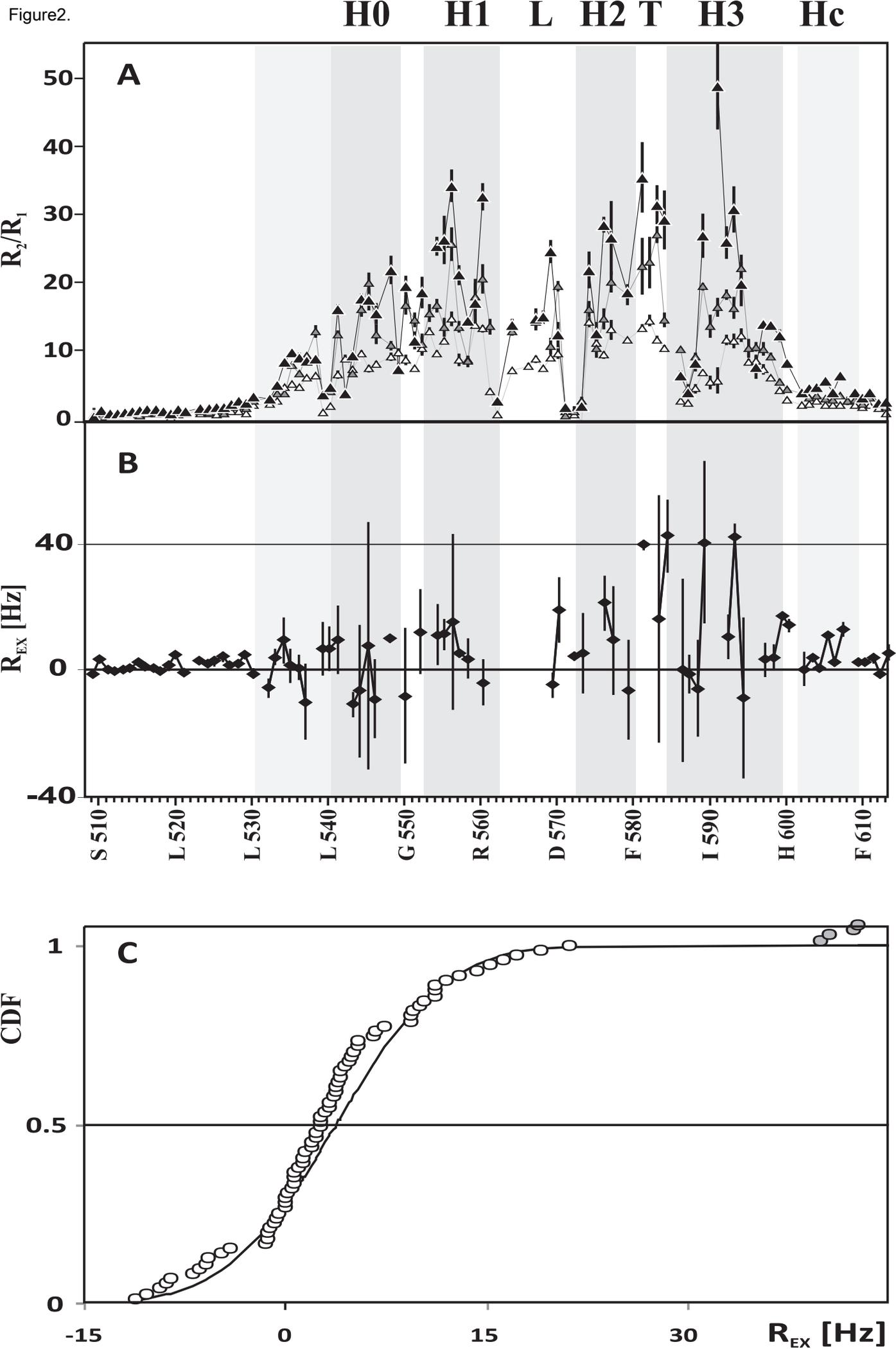


Figure 3.

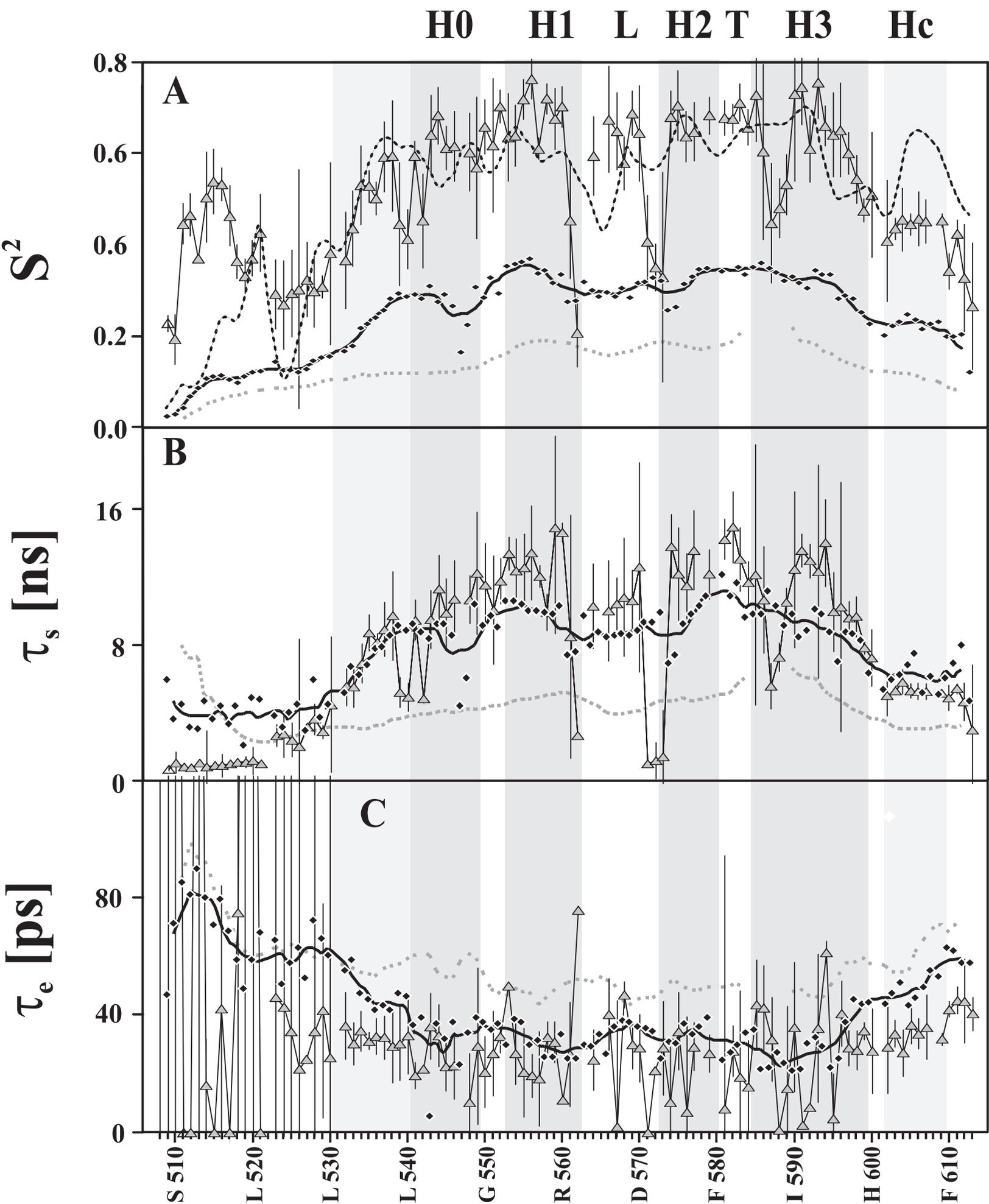


Figure 4.
[Click here to download high resolution image](#)

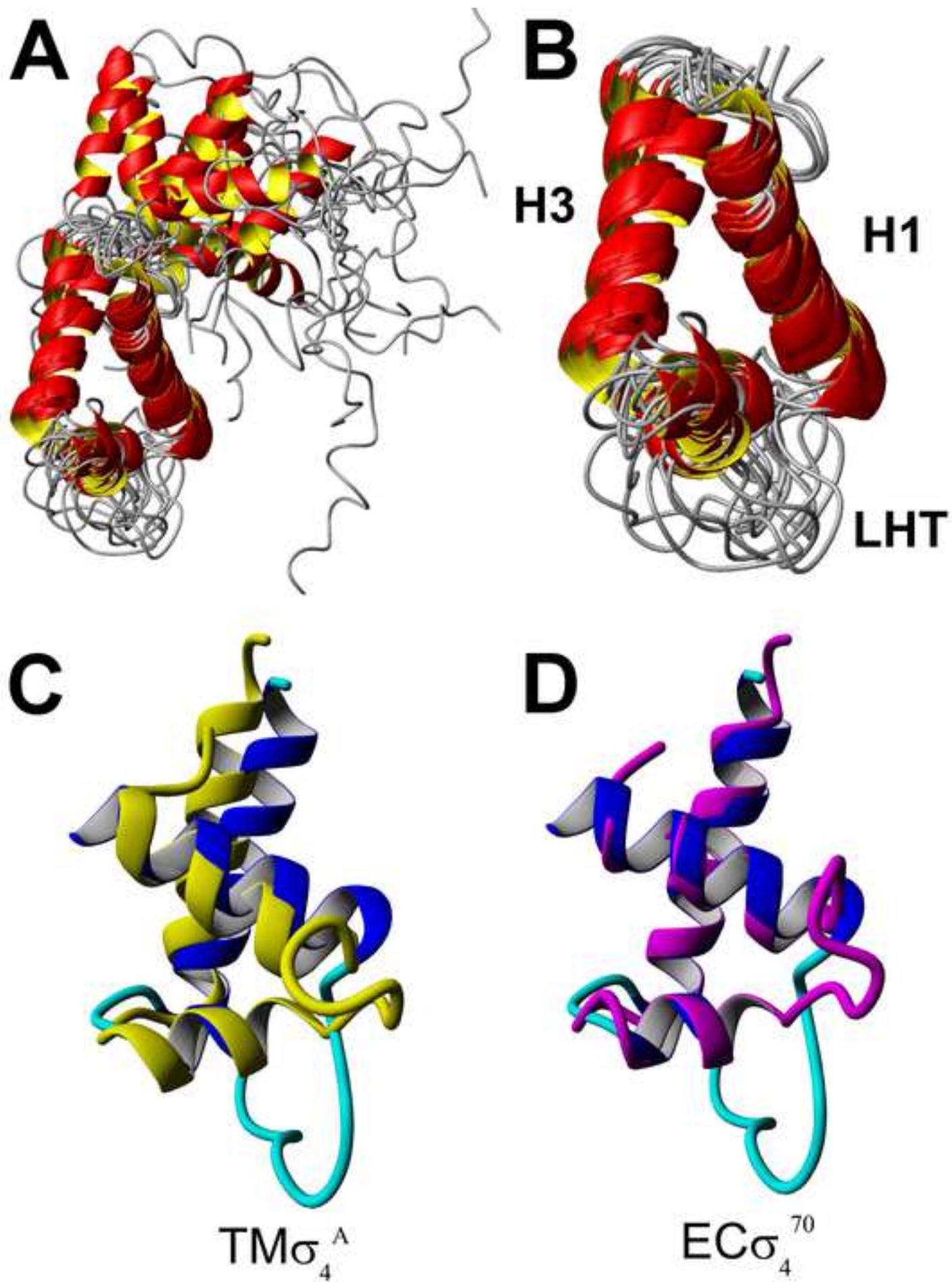
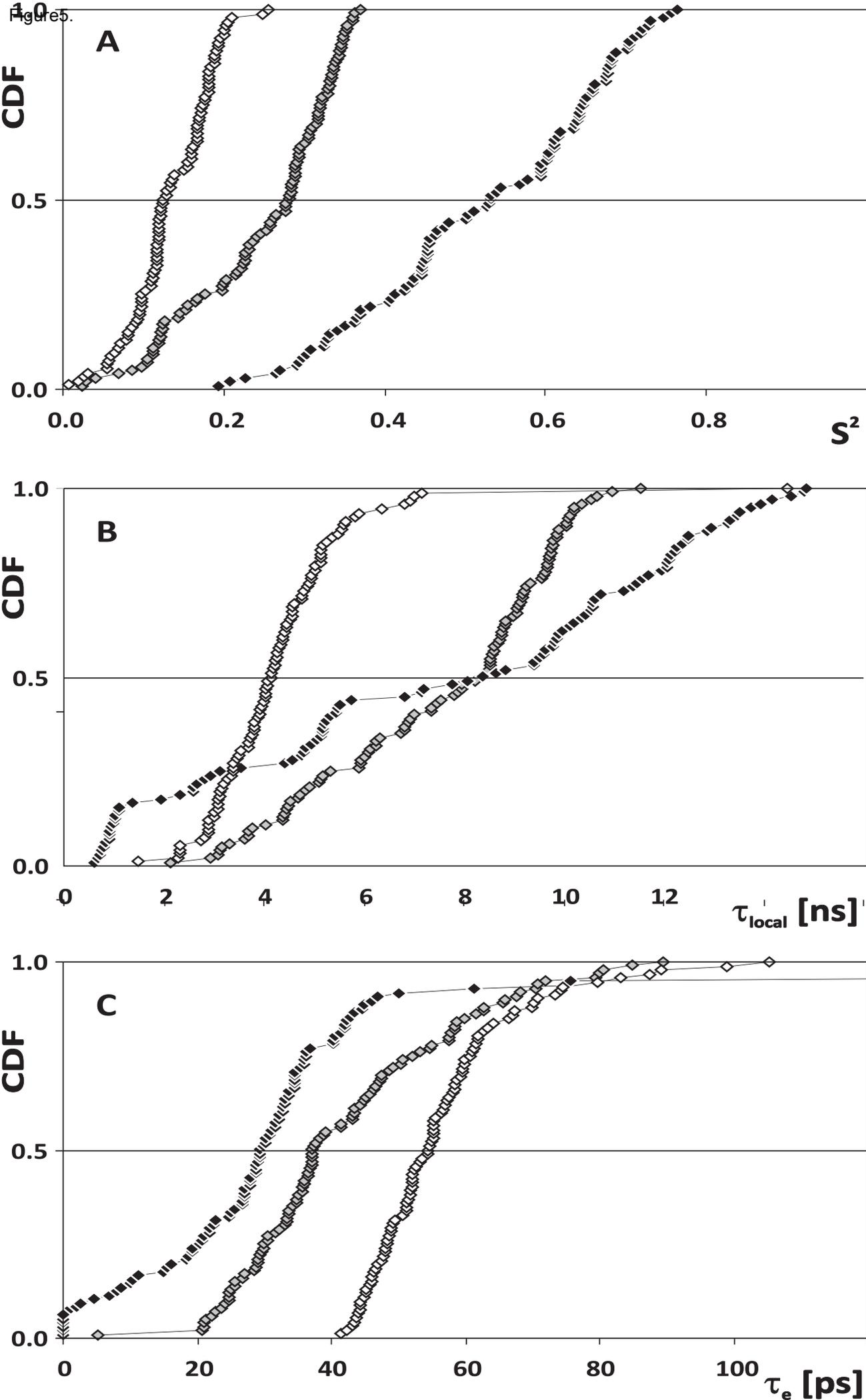


Figure 5.



Supplemental Files

[Click here to download Supplemental Files: Supplementary Figures_140731.pdf](#)