Sozańska *et al. Cell Communication and Signaling* (2025) 23:154
https://doi.org/10.1186/s12964-025-02154-7

Cell Communication and Signaling

RESEARCH Open Access

The molecular properties of the bHLH TCF4 protein as an intrinsically disordered hub transcription factor

Nikola Sozańska[1], Barbara P. Klepka[2], Anna Niedzwiecka[2], Lilia Zhukova[3], Michał Dadlez[3], Beata Greb-Markiewicz[1], Andrzej Ożyhar[1] and Aneta Tarczewska[1*]

Abstract

Background Transcription factor 4 (TCF4) is a member of the basic helix-loop-helix (bHLH) family of transcription factors that guides proper embryogenesis, particularly neurogenesis, myogenesis, heart development and hematopoiesis. The interaction of TCF4 with DNA is dependent on the presence of a conserved bHLH domain, particularly the presence of a basic (b) motif. Most mutations in the *Tcf4* gene are either associated with the development of serious nervous system disorders, such as Pitt-Hopkins syndrome or schizophrenia, or are lethal. Although TCF4 is essential for the proper development and function of the human body, there is a lack of fundamental knowledge about the structure of TCF4 since structural studies were previously limited exclusively to its bHLH.

Methods Recombinant full-length TCF4 was expressed in bacterial cells and purified using chromatographic techniques. To compare the properties of TCF4 in its *apo* and *holo* form, we determined the dissociation constant (K_D) of the TCF4:DNA complex using independent methods, including fluorescence polarization (FP), electrophoretic mobility shift assay (EMSA), and fluorescence correlation spectroscopy (FCS). Then we compared the properties of TCF4 in its *apo* and *holo* form in relation to the changes of the conformation of the polypeptide chain (hydrogen/deuterium exchange mass spectrometry; HDX-MS), hydrodynamic properties (e.g., sedimentation-velocity analytical ultracentrifugation; SV-AUC), and stability (thermal shift, circular dichroism; CD).

Results We demonstrate the molecular characteristics of TCF4, the dimer of which is one of the largest intrinsically disordered proteins (IDPs) described to date. According to our findings, the structure of TCF4 is extensively disordered. Only the bHLH domain exhibits a stable fold. Strikingly, Ephrussi-box (E-box) binding *via* the bHLH domain has no significant effect on the disordered nature of TCF4, but it does influence the dynamic of bHLH and stability of the protein.

Conclusions We suggest that bHLH plays the role of an anchor localizing TCF4 to specific gene sequences. The dual nature of the TCF4 structure and the fact that the intrinsically disordered regions (IDRs) represent most of the protein

*Correspondence:
Aneta Tarczewska
aneta.tarczewska@pwr.edu.pl

Full list of author information is available at the end of the article

sequence, suggest that TCF4 may act as a hub transcription factor regulating the expression of specific genes through the interaction of IDRs with gene-specific partners.

Keywords Transcription factor 4 (TCF4), Immunoglobulin transcription factor 2 (ITF2), SL3-3 enhancer factor 2 (SEF2), Basic helix-loop-helix (bHLH), Intrinsically disordered regions (IDRs), Intrinsically disordered proteins (IDPs), Hub proteins, Transcription factors (TFs), Ephrussi-box (E-box)

Background

Transcription factor 4 (TCF4), also known as immunoglobulin transcription factor 2 (ITF2) or SL3-3 enhancer factor 2 (SEF2), is a member of the class I bHLH (basic helix-loop-helix) family of transcription factors [1], called E proteins. Other known representatives of mammalian E proteins are TCF3 and TCF12 [1, 2]. TCF4 is expressed in several tissue types, but mainly in the nervous system, where the protein is essential for brain development, memory and cognition. Mutations in the human *Tcf4* gene (ID 6925), located on chromosome 18, have been associated with schizophrenia, autism spectrum disorder and Pitt-Hopkins syndrome [3]. *Tcf4* consists of 41 exons, 21 of which are alternative 5' coding exons, that results in the existence of many isoforms of TCF4 (A–R) with unique N-terminal sequences, generated by the alternative splicing [4].

TCF4 regulates gene transcription in a ligand-independent manner acting as a repressor [5–7] or activator [8–10] of transcription, depending on the heterodimerizing partner. Similarly to other members of class I bHLH proteins, TCF4 can interact and heterodimerize with a potent inducers of cell type specification, such achaetescute homolog 1 (ASCL1) or atonal bHLH transcription factor 1 (ATOH1), representing class II of bHLH TFs [11, 12]. The dimerization partners of TCF4 also include intellectual disability (ID) proteins from class V of the bHLH TFs. The interaction of TCF4 with the DNA-binding protein inhibitor ID-2 results in the formation of an inactive heterodimer that binds DNA but is unable to bind activators that are essential for the activation of transcription [13]. The transcriptional activity of class I bHLH proteins including TCF4 is also repressed after binding/heterodimerization with hairy and enhancer of split (HES) proteins from class VI of the bHLH family [14, 15].

The C-terminal bHLH domain interacting with negatively charged DNA, contains a highly conserved basic region composed of positively charged amino acid residues, followed by the HLH structure responsible for the formation of homo- and heterodimers with other bHLH TFs [16–18]. Upon dimerization, the bHLH domain forms a four-helix bundle with a hydrophobic core, where each monomer contacts an E-box DNA element (5'–CANNTG–3') half site [17–22] in the promoters and enhancers of TCF4-responsive genes, such as the μE5 heavy and κE2 light chain immunoglobulin enhancers [23], an enhancer in the murine leukemia virus SL3-3 genome [24], the rat tyrosine hydroxylase enhancer [25], and the human somatostatin receptor-2 promoter [5]. Mutations associated with a genetic disorder known as Pitt-Hopkins syndrome (PTHS), which is characterized by intellectual disability, distinct facial features, developmental delay, and autonomic dysfunction, mostly occur within the bHLH domain of TCF4 [6, 7, 26–29]. Interestingly, associations of TCF4 mutations occurring outside the bHLH domain with schizophrenia have been reported; common variants in human *Tcf4* were among the first genes to reach significance in genome-wide association studies of schizophrenia [30], and rare coding *Tcf4* variants outside of the bHLH domain have been identified in individual schizophrenia patients *via* deep sequencing [31, 32].

Despite the expansion of knowledge regarding the function of TCF4, the expression patterns of individual isoforms in tissues, and the consequences of mutations, the understanding of the structure and molecular properties of the TCF4 polypeptide chain remains limited. The C-terminal bHLH domain of TCF4 is the only structurally characterized fragment of TCF4 to date [21, 33]. The in silico prediction of the full-length TCF4 structure shown in our previous work suggested that the regions outside the bHLH domain in TCF4 might be disordered [34]. The disordered conformation confers many functional advantages. It is flexible and can quickly switch between different folds or remain unfolded. Disordered regions have also been shown to be susceptible to post-translational modifications. The activation domains of TFs are often disordered [35, 36], which significantly increases the range of possible interactions and explains the regulation of protein function. TCF4 exists in multiple isoforms containing different numbers of activation and repression domains, indicating predisposition of the protein to interact with a variety of different partners [4, 37, 38]. The object of our interest was isoform I^- of human TCF4, expressed predominantly in the brain [4]. The lack of fundamental knowledge about the structure of TCF4, prompted us to characterize the full length of the protein, including the region predicted as the IDR. Given that the pivotal function of TCF4 as a transcription factor is binding of DNA through the bHLH domain, we decided to determine whether the molecular properties of this hub protein can be altered upon E-box binding. Previous studies focused only on the DNA binding by TCF4 fragment encompassing bHLH domain [39–41].

In this study, we demonstrate for the first time the molecular characterization of full-length TCF4, highlighting the dual nature of the TCF4 structure. The ordered bHLH domain binds the E-box sequence and anchors the protein to a specific genomic location. The regions outside bHLH, which account for 90% of the sequence, are characterized by a state of complete disorder. This, in turn, results in an extended, flexible conformation. Consequently, TCF4 is capable of interacting with multiple partners, thereby enabling the protein to function as a hub transcription factor. Importantly, our observations revealed that DNA interaction does not affect the disordered nature of TCF4, but it cause the alteration of the stability of the TCF4 molecule.

Materials and methods

In silico analysis of the TCF4 primary sequence

In the present study, all the experiments were performed with the human TCF4 isoform I⁻ (UniProt ID P15884-16). It will be referred to as TCF4 in later sections of this article.

Analysis of the TCF4 sequence was performed using bioinformatics tools with default settings. The ProtParam tool (https://web.expasy.org/protparam/) was used to compute various physical and chemical parameters (e.g., molecular weight, theoretical pI, extinction coefficient). Disorder prediction was performed *via* the PONDR server (http://www.pondr.com) [42], IUpred3 (https://iupred.elte.hu/) [43–45] and NetSurfP-2.0 (https://services.healthtech.dtu.dk/service.php?NetSurfP-2.0) [46]. Protein backbone dynamics was calculated using DynaMine (http://dynamine.ibsquare.be) [47, 48]. Charge-hydropathy analysis was made using the PONDR server [42]. Secondary structure prediction was made by PSIPRED [49]. Example full atom conformations of the TCF4 dimer were generated by ColabFold [50] on the basis of the MMseqs2 homology search and AlphaFold 2.0 [51]. Protein structures were drawn using Discovery Studio 2024 (Dassault Systemes BIOVIA).

Preparation of the expression construct

The cDNA of TCF4 was synthesized *de novo* in GeneArt and optimized *via* a Gene Optimizer (Thermo Fisher Scientific) to obtain the highest expression level of the synthetic gene in *Escherichia coli*. The optimized sequence was used as a template for PCR, and the primers used for amplification were as follows: TCF4_F: GCGAACAGATTGGTGGCATGatgcaggatggtcatcatagc, TCF4_R: GTGCTCGAGTGCGGCCGC*CTA*catctgacccatatgattgctt. The obtained insert was subsequently inserted into the pET-SUMO vector (EMBL, Germany) [52] *via* restriction-free (RF) cloning [53, 54]. The sequence of the obtained expression plasmid was confirmed by sequencing.

Protein overexpression for in vitro analyses

E. coli ArcticExpress competent cells (Agilent Technologies) were transformed with pET-SUMO/TCF4 recombinant plasmids by heat shock and grown overnight at 37 °C on Luria Broth (LB) agar plates supplemented with kanamycin (50 µg/ml) and gentamicin (100 µg/ml). Single colonies were picked and grown in LB media (Invitrogen) supplemented with kanamycin (50 µg/ml) and gentamicin (100 µg/ml) at 37 °C overnight. The main culture in Terrific Broth (TB) media (Invitrogen) supplemented with 50 µg/ml kanamycin was inoculated to a final OD_{600} of 0.1. The culture was incubated at 37 °C and 182 rpm until the OD_{600} reached 0.6, after which the culture was cooled to 16 °C, and protein expression was induced with 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The expression continued for 18 h. The bacteria were harvested by centrifugation (5000 × g at 4 °C for 10 min), resuspended in buffer L (20 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 5% (v/v) glycerol, 2 mM TCEP, pH 7.5) and stored at -80 °C until purification.

Purification of TCF4

The frozen cell suspension was slowly thawed on ice and supplemented with 20 µg/ml DNase I (Merck), 20 µg/ml RNase A (Sigma Aldrich), 1 mM β-mercaptoethanol (Roth), 200 µg/ml phenyl methyl sulfonyl fluoride (PMSF) (Sigma Aldrich) and 50 µg/ml lysozyme (Sigma Aldrich). The cells were lysed *via* sonication. The lysates were then centrifuged at 18 000 × g at 4 °C for 1.5 h. The soluble fractions were purified using immobilized metal affinity chromatography (IMAC). The cell lysate was incubated for 1 h at 4 °C with Ni-NTA His•Bind® Resin (Novagen), which had been previously equilibrated with buffer L. Afterwards, the resin was washed with 10 column volumes (cv) of buffer L to remove unbound proteins, then washed with 5 cv of buffer L supplemented with 6 M urea and re-equilibrated with 10 cv of urea-free buffer L. HisTag-SUMO was removed from the fusion protein by cleavage on resin at 4 °C overnight, using 5 cv of buffer L containing 20 µg/ml SUMO hydrolase dtUD1 (EMBL, Germany) [52]. Contrary to expectations, TCF4 was not present in the flow-through fraction after cleavage of HisTag-SUMO. Instead, the protein was released from the resin using 5 cv of buffer E containing imidazole (20 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, 5% (v/v) glycerol, 2 mM TCEP, pH 7.5). The eluted fractions were pooled, concentrated to a total volume of 1 ml using the Amicon Ultracel-4 centrifugal filter units (Merck Millipore) with a cut-off limit of 30 kDa and injected into the Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences) equilibrated with buffer S (20 mM TrisHCl, 150 mM NaCl, 2 mM TCEP, pH 7.5). The column was operated at room temperature with a flow rate of 0.5 ml/min on an Äkta Avant system (GE Healthcare

Life Sciences). Purified TCF4 samples were stored at -80 °C in buffer S. The purity of the protein was estimated densitometrically using ImageJ [55] and was approximately 95% (Additional file 1: Fig. S1A). The molecular mass of recombinant TCF4 was determined *via* ESI MS at a Mass Spectrometry Laboratory (IBB PAS, Warsaw) (Additional file 1: Fig. S1B).

TCF4 labeling

TCF4 was labeled with Alexa Fluor 488 NHS Ester (Invitrogen). The protein in buffer P (10 mM phosphate, 150 mM NaCl, 2 mM TCEP) was concentrated to 50 µM and incubated with the label at a 1:20 ratio. The conjugation reaction was performed for 1 h at RT, and the pH was adjusted to 8.0. The unbound label was separated from the sample using a Superdex 200 Increase 10/300 column (GE Healthcare). This step also provided exchange of buffers to S. The calculated degree of labeling (DOL) was 1.2 on the basis of the following equation:

$$DOL = \frac{A_{max} \times \epsilon_{TCF4}}{(A_{280} - A_{max} \times CF) \times \epsilon_{max}}, \quad (1)$$

where A_{max} is the absorbance at λ_{max}, ε_{TCF4} is the molar extinction coefficient (in $M^{-1} \times cm^{-1}$) of the pure protein at 280 nm, ε_{max} is the molar extinction coefficient (in $M^{-1} \times cm^{-1}$) of the dye at its absorbance maximum, and CF is the correction factor.

SDS-PAGE

Protein samples were analyzed by SDS-PAGE using 4%/10% polyacrylamide gels developed in a Tris/Glicyne system [56]. Electrophoresis was performed at a constant current of 20 mA/1 mm gel. The Unstained Protein Molecular Weight Marker (Thermo Fisher Scientific) was used. After electrophoresis, the gels were stained with Coomassie Brilliant Blue R-250 [57].

dsDNA binding assays

To determine the dissociation constant (K_D), we assumed that double-stranded DNA is bound by the dimeric TCF4 (dTCF4) at a 1:1 stoichiometry on the basis of the known properties of the E-box-binding bHLH transcription factors [58]. The dTCF4:E-box K_D was determined by fluorescence polarization (FP), electrophoretic mobility shift assay (EMSA), and fluorescence correlation spectroscopy (FCS) based on the fluorescence of the 6-carboxy-fluorescein (FAM)-labeled double-stranded DNA (dsDNA) probe (FAM-CCGGT*CACGTG*TCCTA [21], where the E-box canonical sequence is in italics).

Fluorescence polarization

FP experiments were performed using a CLARIOstar Plus plate reader (BMG LABTECH) at RT. All measurements were performed in triplicate, using default settings and top optics. Prior to measurements, the dsDNA probe (40 nM) was incubated with increasing amounts of the dTCF4 protein (0–1.5 µM, calculated for the dimer) for 30 min in buffer S with the addition of BSA (final concentration of 0.5 mg/ml) to avoid nonspecific interactions.

Fluorescence correlation spectroscopy

Prior to the FCS experiments, the TCF4 protein solution was centrifuged to remove nonspecific aggregates. A FAM-labeled dsDNA probe at 100 nM was incubated with increasing concentrations of dTCF4 (0–2.5 µM per dimer) for at least 30 min. The FCS experiments were performed on a Zeiss LSM 780 with ConfoCor 3 in buffer S in droplets of 25 µl. A single measurement lasted 3 s and was repeated 50 times in series. The series were repeated in 3 independent droplets. The temperature inside the droplet (23 ± 1 °C) was checked after the FCS measurements using a certified calibrated microthermocouple. The structural parameter (*s*) was determined with the AF 488 probe in pure water ($D_{AF\ 488}$ = 435 $\mu m^2 s^{-1}$) [59] individually for each microscope slide previously passivated with BSA. The actual solution viscosity of the buffer and solutions at increasing protein concentrations was determined by comparison of the diffusion times for AF 488 in pure water and in the solution at the same instrument calibration. The experiments were performed at an excitation wavelength of 488 nm with a relative argon multiline laser power of 3%, MBS 488 nm, BP 495–555 nm and a damping factor of 10%. The resulting diffusion time (τ_{app}) of the dsDNA probe measured by FCS was the weighted average of the freely diffusing probe and the dTCF4-bound form. The raw measurements were manually curated to eliminate possible nonspecific oligomers or aggregates in the confocal volume. A one-component model of 3D diffusion was used, taking into account the triplet state lifetime of the probe. The FCS data were analyzed by global fitting *via* Zen2010 software (Zeiss) essentially as described previously [60], and the τ_{app} values were subsequently corrected for increasing solution viscosity (Additional file 1: Fig. S2, Table S1). The total experimental uncertainty of the FCS results was determined according to the propagation rules for small errors [61], taking into account the statistical dispersion of the FCS results and the uncertainty of other experimental values used for calculation of the results.

Electrophoretic mobility shift assay

EMSA was performed on the same set of samples used in the FP assays. An aliquot of 10 µl of the reaction mixed with glycerol (final concentration of 20%) and bromophenol blue (final concentration of 1%) was loaded onto 5% native polyacrylamide gels and run at 150 V for 150 min

in 0.5 × TBE buffer at 4 °C. Images were captured using the ChemiDoc MP Imaging System (Bio-Rad). DNA probe loss was estimated densitometrically from the captured image (Additional file 1: Fig. S3A) using ImageJ [55].

Numerical data analysis of dsDNA binding assays

OriginPro software (version 9.0) and GraphPad Prism (version 6.07) were used to calculate the mean values and standard deviations from the measurements and to perform nonlinear, least-squares regressions. Millipolarization (mP) from the FP experiments and the apparent diffusion time (τ_{app}) from the FCS experiments (denoted Y) are described by the following equation:

$$Y = \frac{(Y_{max} - Y_0) \times [cx]}{[dsDNA]} + Y_{0\prime} \quad (2)$$

where Y_{max} and Y_0 are the signals (mP or τ_{app}) of the probe in the complex with dTFC4 and in the free state, respectively, and [cx] is the equilibrium concentration of the dTCF4:E-box complex, expressed as a function of the total dTCF4 concentration, according to the following equation:

$$[cx] = \frac{1}{2}\left([dTCF4] + [dsDNA] + K_D - \sqrt{([dTCF4] - [dsDNA] + K_D)^2 + 4 \times [dsDNA] \times K_D} \right). \quad (3)$$

The theoretical curve (Eq. 2 + 3) with the free parameters K_D, Y_{max} and Y_0 was fitted to the experimental data points.

EMSA data were analyzed according to the modified Hill equation, which could compensate for deviations from ideal conditions [62], as follows:

$$dsDNA_{free} = \frac{A}{1 + \left(\frac{[dTCF4]}{K_D}\right)^n} + B, \quad (4)$$

where A and B are coefficients representing the upper and lower plateaus of the titration, respectively, and n is the Hill coefficient. Statistical analysis was performed on the basis of runs tests (*P*-value) and goodness of fit (R^2), as previously described [63]. The *P*-value for the fits was > 0.4 in each case, indicating that there are no statistically significant deviations from the applied models of dTCF: dsDNA binding.

We used the K_D value determined by the FP technique to calculate the amount of specific dsDNA (CCGGT-*CACGTG*TCCTA, where the E-box canonical sequence is in italics) used in other experiments to compare the properties of free TCF4 and DNA-bound TCF4. In each case, we assumed a molar excess of DNA such that the saturation of dTCF4 with the dsDNA probe was not less than 95%.

Hydrogen/deuterium exchange mass spectrometry

Hydrogen/deuterium exchange mass spectrometry (HDX-MS) was performed on free TCF4 and the TCF4:DNA complex. For measurements of TCF4 in complex with DNA, an E-box was used at twice the molar excess of dsDNA per dTCF4, and samples were incubated on ice for 30 min prior to the experiment. Prior to the HDX reactions, nondeuterated fractions of both variants served as sources for peptide lists. For this purpose, the LC-MS analysis was performed with all steps the same as described below for the HDX runs, but in this case, the D_2O used for exchange was replaced with H_2O. Peptide identification was performed using ProteinLynx Global Server software (Waters). For the HDX reaction, the starting stock protein concentration was 7 µM. HDX exchange incubations of both variants were performed at five time points: 10 s, 1 min, 5 min, 30 min, and 2.5 h in quadruplicate. 5 µl aliquots of protein stocks were added to 45 µl of deuteration buffer (20 mM Tris DCl, 150 NaCl, 5 mM TCEP, pD 7.45) at room temperature. The final protein concentration in the deuteration reactions was 0.7 µM. The H/D exchange reactions were quenched by transferring the exchange aliquots to precooled tubes (on ice) containing 10 µl of quenching buffer (2 M glycine in 99.99% D_2O, pD 2.3). After quenching, the samples were immediately frozen in liquid nitrogen and stored at -80 °C until mass spectrometry measurement. The samples were thawed directly prior to measurement and manually injected onto the nano ACQUITY UPLC system equipped with HDX-MS Manager (Waters). Proteins were digested on a 2.1 mm × 20 mm nepenthensin-2 (POROS™) (AffiPro) column for 1.5 min at 20 °C and eluted with 0.07% formic acid in water at a flow rate of 200 µl/min. The digested peptides were passed directly to the ACQUITY BEH C18 VanGuard precolumn, from which they were eluted onto the reversed-phase ACQUITY UPLC BEH C18 column (Waters) using a 10–35% gradient of acetonitrile in 0.01% of formic acid at a flow rate of 90 µl/min at 0.5 °C. The samples were measured on a SYNAPTG2 HDX-MS instrument (Waters) without IMS mode. The instrument parameters for MS detection were as follows: ESI—positive mode; capillary voltage—3 kV; sampling cone voltage—35 V; extraction cone voltage—3 V; source temperature—80 °C; desolvation temperature—175 °C; and desolvation gas flow rate—800 l/h.

Two control experiments were performed to determine the minimum and maximum H/D exchange levels. To obtain the minimum exchange of each peptide (M_{min}), 10 µl of a quench buffer was mixed with 45 µl of D_2O reaction buffer (20 mM Tris DCl, 150 NaCl, 5 mM TCEP,

pD 7.45) prior to the addition of 5 µl of protein stock mixture and analyzed by LCMS. To obtain the maximum H/D exchange level (M_{max}), the deuteration reaction was run for 24 h and then quenched with a quench buffer kept on ice. The control experiments were also performed in quadruplicate.

Peptide lists obtained from nondeuterated protein samples were used to analyze the exchange data using DynamX 3.0 software (Waters). The PLGS peptide list was filtered by minimum intensity criteria—3000 and minimal product per amino acid—0.3. All the raw files were processed and analyzed *via* DynamX 3.0 software. All MS assignments in Dynamix were inspected manually. The percentage of deuteration D[%] for all peptides was calculated in an Excel file from exported DynamX 3.0 data, based on the Eq. 5, which takes into account the minimal and maximal exchange of a given peptide:

$$D\,[\%] = \frac{M - M_{min}}{M_{max} - M_{min}} \times 100\%, \qquad (5)$$

where M is the centroid mass of a given peptide after deuterium uptake, M_{min} is the centroid mass of a peptide with minimal exchange and M_{max} is the centroid mass of a peptide with a maximal exchange.

The error bars for the fraction exchanged represent the SDs calculated from at least three independent experiments. The difference in the fraction exchanged (Δ fraction exchanged) was calculated by subtracting the fraction-exchanged values for peptides in the selected state from the values for the same peptides in the control state. Error bars for differences were calculated as the square root of the sum of the variances of the compared states. Student's t test for two independent samples with unequal variances and unequal sample sizes (also known as Welsh's t test) was performed to evaluate differences in fraction-exchanged values between the same peptides in two different states. The final data analysis and visualization steps were carried out using the in-house HaDeX software [64].

Native PAGE

Samples of free TCF4 and TCF4 with E-box (the final concentrations of dTCF4 and E-box were 2 µM and 20 µM, respectively) were mixed with glycerol (final concentration of 20%) and bromophenol blue (final concentration of 1%). Then, samples of 20 µl were loaded onto 5% native polyacrylamide gels, and run at 150 V for 40 min in 0.5 × TBE buffer at 4 °C. After electrophoresis, the gels (Additional file 1: Fig. S3B) were stained with Coomassie Brilliant Blue R-250 [57] and analyzed using Image Lab Software (Bio-Rad).

Analytical size exclusion chromatography

The analytical size exclusion chromatography (SEC) experiment was performed on an AKTA Avant FPLC system (GE Healthcare) with a Superdex 200 Increase 10/300 column (GE Healthcare). The column in buffer S was calibrated with a Gel Filtration Marker Kit (MW range 29–700 kDa) (Sigma-Aldrich) according to the manufacturer's protocol. A total of 75, 50, and 25 µg of TCF4 samples in 100 µl of S buffer were spun down for 10 min at 18,000 × g, injected onto the column at a flow rate of 0.5 ml/min and detected by absorbance measurements at wavelengths of 280 and 260 nm. An analogous set of samples at the same concentrations in the presence of E-box dsDNA (twice the molar excess of dsDNA to dTCF4) was then spun down for 10 min at 18,000 × g, injected onto the column at a flow rate of 0.5 ml/min and detected by absorbance measurements at 280 and 260 nm wavelengths. E-box dsDNA was also injected onto the column at a concentration of 10 µM. All the injections were performed in triplicate.

Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments were performed on a Beckman Coulter Proteome Lab XL-I ultracentrifuge (Beckman Coulter Inc.) using an AN-60Ti 4-hole rotor, a 12 mm path length, a charcoal-filled double sector Epon® centerpiece, and quartz windows. Scans were collected at 20 °C and a rotor speed of 50,000 rpm. TCF4 was analyzed at 0.25, 0.50, and 0.75 mg/ml (2.5, 5.0, and 7.5 µM dTCF4, respectively) without and with dsDNA at 5.0, 10.0, and 20.0 µM, respectively, in buffer S with a detection wavelength of 280 nm.

The time-corrected data were analyzed using the SEDFIT software (version 16.1) using the continuous $c(f/f_0)$ distribution with a confidence level of 0.95. This model was chosen because it allows a known molecular weight (MW) value to be fixed while treating the frictional coefficient ratio (f/f_0) [65] and the hydrodynamic radius (R_h) as free parameters to be determined from the numerical analysis along with their approximate errors. In a SEDFIT analysis, all these parameters can be calculated for individual entities separated by AUC. At least two such entities were present in each sample.

The input density (1.005 g/ml) and viscosity (0.010213 mPa × s) of the solution were calculated using SEDNTERP [66]. The partial specific volume, $\bar{\nu}$, of the protein (0.7131 ml/g) was determined from the sequence using SEDNTERP [67]. The $\bar{\nu}$ value for the dTCF4:dsDNA complex (0.697 ml/g) was calculated as the weighted average of the $\bar{\nu}$ values for the protein and the DNA [68], where the weights were the MWs of dTCF4 and dsDNA, assuming a known 1:1 complex stoichiometry. The initial f/f_0 resolution in the continuous $c(f/f_0)$ distribution

analysis was defined by an equidistant ladder of 12 steps from 1 to 3.5.

Further post-SEDFIT analysis focused on the peaks associated with dTCF4 or the dTCF4:dsDNA complex in the sedimentation profiles obtained for these samples. The first step was to select the ranges of the sedimentation coefficient, s, corresponding to these entities in their profiles (Additional file 1: Fig. S4A), to be used in further analysis. Next, we found the values of the shape factor f/f_0 [65] for dTCF4 and the complex as the values for which the center of the MW distribution generated by SEDFIT corresponds to the known molecular weight of the dimer (96.5 kDa) or the complex (106.4 kDa) (Additional file 1: Fig. S4B). The uncertainty of the f/f_0 value ($\Delta f/f_0$) is the difference between the best f/f_0 value and the nearest analyzed f/f_0 values. Then, for the best f/f_0 and two closest f/f_0 values, R_h was plotted as a function of MW (Additional file 1: Fig. S4C). The desired value of R_h is the one that intersects the graph of the nonlinear function $R_h(MW)$ for the best f/f_0, and ΔR_h is the difference between R_h and the value determined for the neighboring f/f_0.

This procedure for the continuous $c(f/f_0)$ distribution was repeated three times, with each iteration refining the analysis by progressively narrowing the f/f_0 range (2nd iteration: 1.7 to 3.5 with 20 steps; 3rd iteration: 2.27 to 2.45 and 2.18 to 2.36 for dTCF4 and the complex, respectively, with 6 steps).

Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded using a Jasco-815 spectropolarimeter (Jasco Inc.) equipped with a Peltier temperature controller (CDF-426 S/15). The spectra were collected in the spectral range of 195–300 nm at a scanning speed of 50 nm/min at 20 °C, D.I.T. of 2 s and a 1 nm bandwidth. The data step was 0.5 nm. The spectra were measured in 5 accumulations using 1 mm path length quartz cuvettes, and the concentration of dTCF4 was 0.5 mg/ml (5 μM). An E-box was applied at twice the molar excess of dsDNA over dTCF4, and the samples were incubated on ice for 30 min before measurement. All the spectra were corrected for the effect of buffer and converted to molar residual ellipticity units [69]. The mean residue weight (MRW) for TCF4 I^- is 106.65 Da.

Temperature-dependent denaturation was monitored by following the changes in ellipticity in the range of 195–260 nm by increasing the temperature from 20 to 95 °C and then decreasing it from 95 to 20 °C at a constant rate of 1 °C/min. The melting temperature was determined by fitting a sigmoidal (Boltzmann) curve to the points of the Θ_{MRW}-temperature relationship for a wavelength of 220 nm. The fitting was performed using OriginPro software (version 9.0).

Thermal shift assays

Thermal shift assays were performed using QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific). Each sample contained 5 × SYPRO® Orange Protein Gel Stain (Merck). Stock concentrations of protein, SYPRO® probe and E-box DNA were prepared in buffer S. The assays were performed for the samples containing dTCF4 (5 μM) with and without dsDNA, dsDNA alone, and SYPRO® Orange as a control sample. An E-box was applied at a 2-fold molar excess of dsDNA over dTCF4. All the samples were prepared in 3 replicates. The measurements were performed in a 96-well plate. The samples were incubated on ice for 30 min before measurement. The temperature was subsequently increased from 25 to 99 °C at a rate of 0.05 °C/s. The optical filter used was 470 ± 15 nm for excitation and 586 ± 10 for emission. The curves were then averaged for each species and corrected for the temperature exponential decay of the fluorescence intensity on the basis of the correction curve determined from the temperature fluorescence quenching of pure SYPRO® Orange. The melting point was then calculated from the first derivative of the corrected temperature-dependent fluorescence curve as the maximum absolute value of this derivative.

Results

In silico analyses suggest that TCF4 is mostly disordered

Our previous in silico predictions of the TCF4 structure suggests that regions outside the bHLH domain in TCF4 may be disordered. Here, we present the results of a detailed analysis to determine the propensity of TCF4 (I^-) to adopt an intrinsically disordered structure.

IDPs are characterized by a low overall hydrophobicity and a large net charge, so it is possible to categorize them using a charge-hydropathy plot [70]. The charge-hydropathy plot prepared for TCF4 (Fig. 1A) shows that this protein is located slightly below the boundary and is grouped with the ordered set of proteins. This may be due to the presence of a locally ordered structure — the well-studied, highly conserved bHLH domain. However, it should be noted that using only charge-hydropathy plot does not give certain results, and as presented in Fig. 1A there are many known examples of proteins documented as disordered whose location also suggests their ordered nature.

To investigate to what extent TCF4 could reveal features of IDPs, we used independent predictors of disorder: PONDR, IUpred3, PSIPRED, and NetSurfP-2.0. The results obtained with all the algorithms are consistent and indicate the presence of extensive disorderness, of about 80%. The remainder of the sequence consists mainly of a fragment containing the bHLH domain (residues 348–401) (Fig. 1B), and according to the NetSurfP-2.0 prediction results, it is involved in the formation of helices

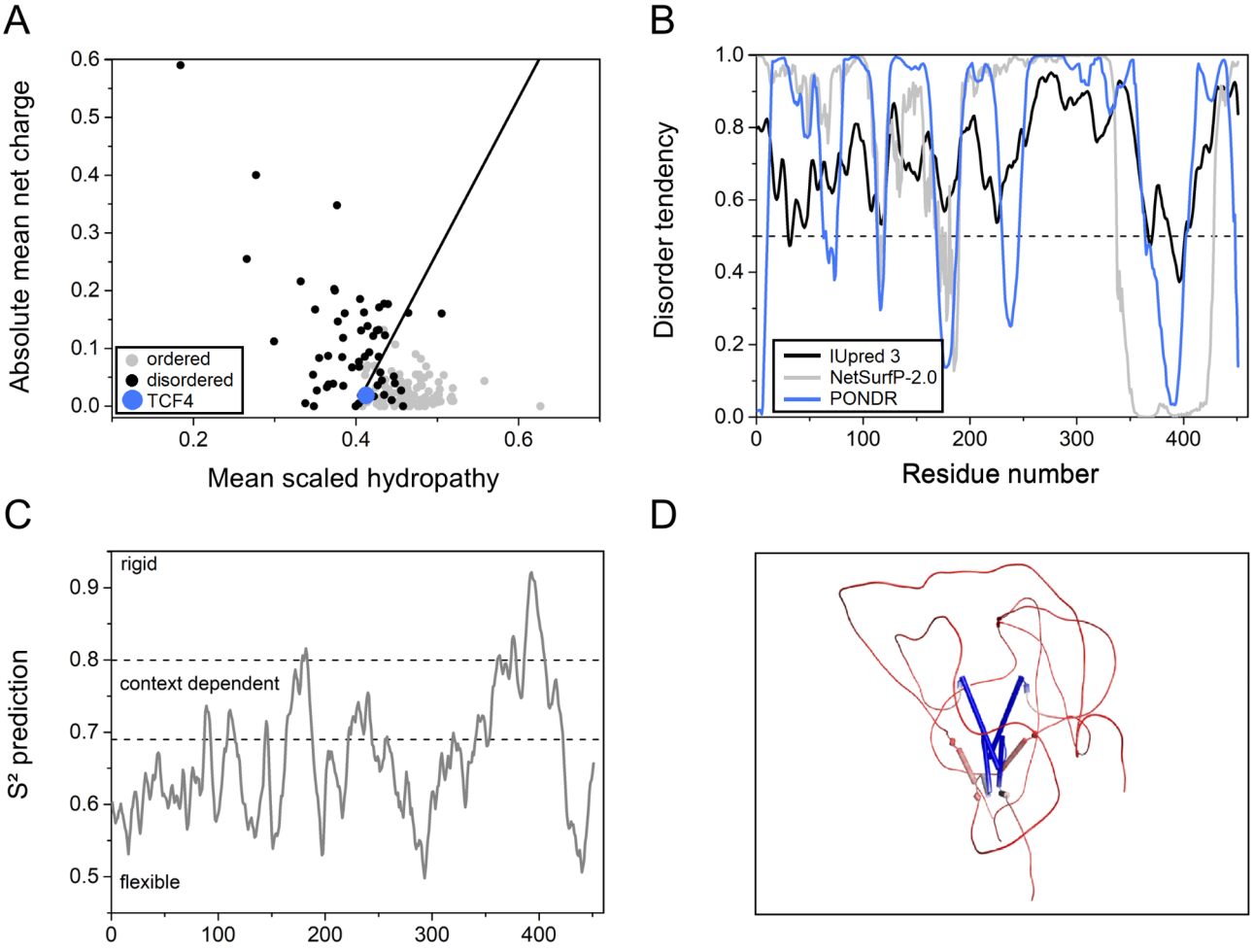

Fig. 1 In silico analyses of TCF4. (**A**) Charge-hydropathy plot of fully ordered proteins (gray dots), fully disordered proteins (black dots) and TCF4 (blue dot). The black line represents the boundary between the ordered and disordered proteins. (**B**) Prediction of the degree of the disorder in the TCF4 sequence calculated from the primary structure. The results using 3 independent predictors are shown: IUpred 3 (black line), NetSurfP-2.0 (gray line), and PONDR (blue line). A score above 0.5 indicates a high probability of disorder. (**C**) Prediction of protein backbone dynamics using DynaMine. An S^2 value higher than 0.8 indicates high rigidity of the protein backbone, whereas an S^2 value lower than 0.69 indicates high flexibility, which is typical for disordered segments. Values between 0.69 and 0.8 are characteristic of the context-dependent structural organization of polypeptide chains. (**D**) Example TCF4 dimer conformation generated by AlphaFold 2.0 for visualization purposes [71]. A schematic protein representation colored according to the average isotropic displacement, from low (blue) to high values (red); α-helices are shown as cylinders

(Additional file 1: Fig. S5A). To predict the dynamics of the protein backbone, we used DynaMine, which defines the presence of flexible regions [47, 48]. The more flexible the region is, the greater the probability of the existence of the disordered structures [47, 48]. The TCF4 sequence was mostly identified as flexible/disordered, as indicated by values less than 0.69. Context-dependent (values between 0.69 and 0.8) and rigid/ordered regions (values above 0.8) mainly cover the bHLH domain (Fig. 1C). Considering the crystal structures of the TCF4 bHLH domain obtained for both the DNA-free and DNA-bound forms of TCF4 [21], we assumed that the protein exists as a dimer and used a dimeric sequence to simulate its conformation. The results obtained using ColabFold [50] based on the MMseqs2 homology search and AlphaFold 2.0 [51] (Fig. 1D) suggest the possibility of the existence of an additional α-helical fragment in the region of 167–183 of the long N-terminal IDR, which is consistent with the PSIPRED and NetSurfP-2.0 predictions of the helical propensity for the 166–187 fragment (Additional file 1: Fig. S5B) as well as with a decreased disorder tendency (Fig. 1B) and increased rigidity (Fig. 1C) for this sequence region.

The results of the in silico analyses suggested that TCF4 might be mainly disordered, but as mentioned earlier, the experimental results do not always match the predictions. For this reason, we decided to conduct in vitro experiments to analyze the molecular properties of TCF4. Since TCF4 is a DNA-binding protein, we decided to perform

experiments on both DNA-free TCF4 and TCF4 in complex with a specific E-box sequence.

dTCF4 binds the canonical E-box sequence in vitro

In this study we focused on the differences between TCF4 structures in the free state and in the DNA-bound state. Therefore, it was especially important to obtain pure TCF4 that did not contain DNA from bacterial cells. During the purification process, we found that to achieve this goal, it was necessary to use a buffer containing 6 M urea for washing the IMAC column with the bound 6xHis-SUMO-TCF4 fusion protein. By omitting this step, the protein preparation was significantly contaminated with bacterial nucleic acids (Additional file 1: Fig. S6).

To test the DNA-binding ability of TCF4 and further compare it with the data for isolated bHLH domain [21], we used a double-stranded fluorescently labeled DNA probe that contained the E-box canonical sequence. First, we determined the dissociation constant (K_D) *via* a fluorescence polarization (FP) assay and electrophoretic mobility shift assay (EMSA). To determine the K_D, we assumed that dsDNA is bound by the dimeric protein (dTCF4) at a 1:1 stoichiometry on the basis of the known properties of the E-box-binding bHLH transcription factors [58].

The results of the FP binding assay are shown in Fig. 2A. The K_D value obtained using the FP-based method was determined to be 0.12 ± 0.03 μM. The same set of samples was used to perform EMSA (Fig. 2B). After the EMSA results were acquired, the relationship between the amount of unbound fluorescent probe and the concentration of dTCF4 was determined densitometrically. Numerical analysis of the relative integrated intensity of the bands revealed that the K_D value obtained from EMSA was 0.327 ± 0.014 μM.

As there were some differences in the results obtained using FP and EMSA, we decided to use a third technique, FCS. The autocorrelation curves show a systematic shift toward longer lag times in the presence of higher amounts of dTCF4 (Fig. 3A). On the basis of the FCS curves, we calculated the apparent diffusion times (τ_{app}) of the probe at different protein concentrations. The analysis of the increasing τ_{app} value as a function of the dTCF4 concentration yielded K_D = 0.33 ± 0.11 μM (Fig. 3B). The diffusion time of the dTCF4:E-box complex determined from FCS was τ_{max} = 390 ± 30 μs, which corresponds to a hydrodynamic radius R_h = 68 ± 6 Å.

The results of the binding assays for the dTCF4:E-box complex formation are presented in Table 1. All the approaches used showed that the affinity of dTCF4 for the E-box-containing dsDNA probe is submicromolar. The K_D value was quantitatively consistent between the EMSA and FCS measurements. The FP results may even indicate a slightly higher binding affinity within the complex, but the goodness of fit, R^2, of the fit is worse due to the limited number of points at lower concentrations. Notably, however, the K_D value for the interaction of the E-box with the TCF4 bHLH domain was shown to be

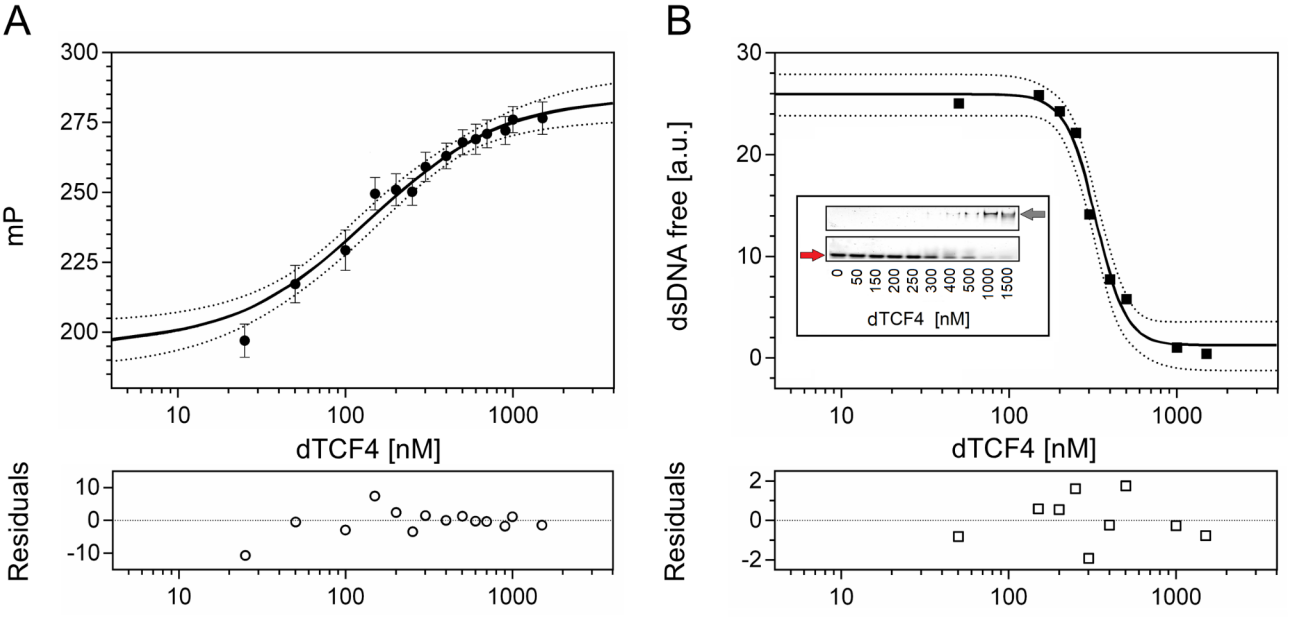

Fig. 2 Determination of the dTCF4:E-box K_D using the FP assay and EMSA. (**A**) Data from the FP assay. The black circles are the average millipolarization (mP) values from three measurements, shown with the standard deviation; the binding curve was fitted to the experimental points according to Eq. 2 and Eq. 3 (see the Materials and methods section). (**B**) Loss of the free fluorescent dsDNA probe determined from the intensity of the bands on the polyacrylamide gel after EMSA (inset); the red and gray arrows show the free and dTCF-bound dsDNA probes, respectively; the binding curve was fitted according to Eq. 4. The accuracy of the binding assays was assessed on the basis of the 95% CI (dotted lines) and fitting residuals (bottom panels)

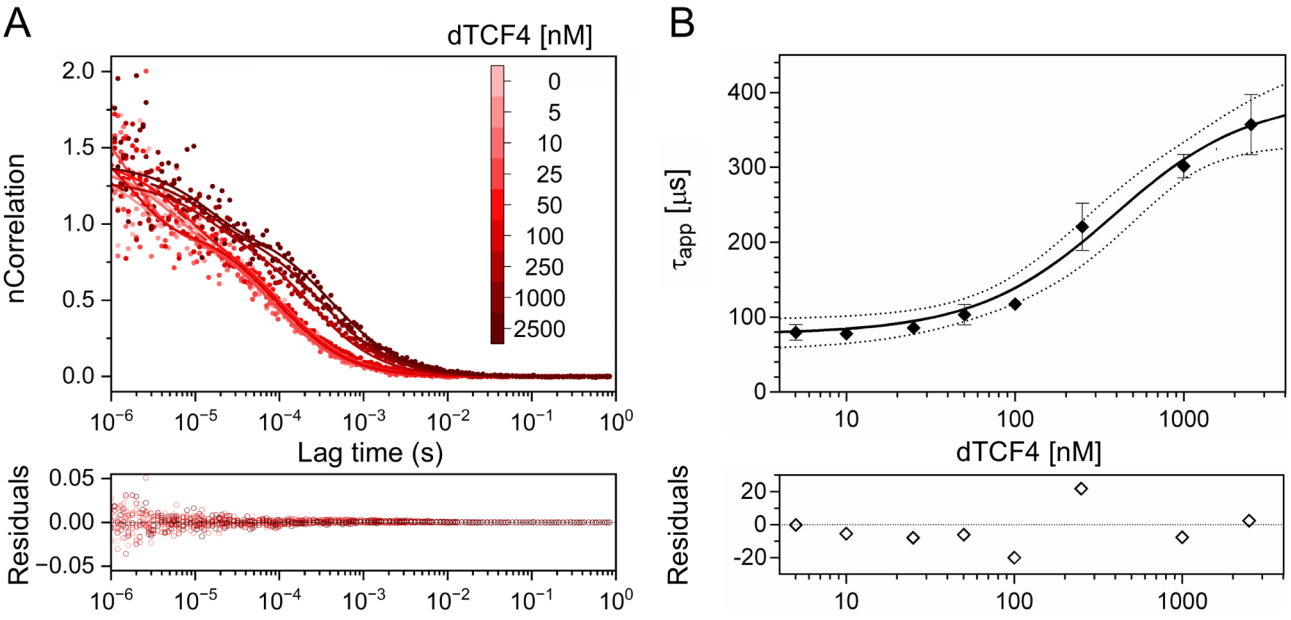

Fig. 3 Determination of the dTCF4:E-box K_D using FCS. (**A**) Normalized FCS data (dots) and autocorrelation curves (solid lines) for the FAM-labeled dsDNA probe at 100 nM titrated with increasing concentrations of dTCF4, marked by red intensity (upper panel), together with the corresponding nonnormalized FCS fitting residuals (bottom panel). (**B**) The binding curve was fitted to the experimental points according to Eq. 2 and Eq. 3 (see the Materials and methods section); the accuracy of the binding assay was assessed on the basis of the 95% CI (dotted lines) and the fitted residuals (bottom panel)

Table 1 Dissociation constants (K_D) with their 95% confidence intervals and goodness of fit (R^2) for dTCF4 binding to a dsDNA probe containing an E-box

	K_D (µM)*	R^2
FP	0.12 ± 0.03	0.973
EMSA	0.327 ± 0.014	0.989
FCS	0.33 ± 0.11	0.983

*K_D calculated vs. TCF4 monomer concentration are: 0.26 ± 0.05, 0.65 ± 0.03, and 0.76 ± 0.23 µM from FP, EMSA, and FCS, respectively

sensitive to ionic strength and the presence of a crowding agent [21]. Considering that each of the binding assays was performed under different viscosity, crowding, and temperature conditions, we can conclude that the K_D values we obtained for the dTCF4:dsDNA complex by the three methods are semiquantitatively concordant, ranging from 0.2 to 0.3 µM.

TCF4 is almost entirely disordered except for the canonical bHLH domain, which binds the E-box sequence

To estimate the secondary structure of TCF4 and evaluate how it changes after DNA binding, we performed UV CD spectroscopy measurements of TCF4 in its free and DNA-bound form. The recorded spectra were characteristic for disordered proteins with some fraction of helical structures, and differed only slightly between *apo* and *holo* forms of TCF4, suggesting only a transition of the basic region of the bHLH domain from disordered to ordered toward the α-helical fold after DNA binding (Additional file 1: Fig. S7). The results obtained with CD indicate, in line with the crystal structures [21], that there is no significant effect of E-box DNA binding on the overall protein secondary structure.

To directly map the IDRs in TCF4, we utilized hydrogen/deuterium exchange mass spectrometry (HDX-MS). In this technique, the rate of chemical exchange between deuterons in solution and hydrogens on the amide groups of a protein is determined by mass spectrometry. Amide groups deeply buried in the core of a protein or those involved in intramolecular bond formation are characterized by a very low rate of hydrogen exchange. In contrast, amides from the dynamic disordered fragments exchange hydrogens very rapidly. This difference can be used to identify IDRs in a given protein [72, 73]. The results of HDX-MS analysis for free TCF4 after 10 s and 150 min are shown in Fig. 4A. (For full results after all the times analyzed, see Additional file 2). The entire sequence, except for the region covering the bHLH domain (residues 360–405) is characterized by efficient, almost total H/D exchange after 10 s of the reaction. After 150 min, the fractional deuterium uptake of peptides containing the bHLH domain increased, indicating the dynamics of this protein fragment.

The next step was to perform an analogous experiment for TCF4 in complex with the dsDNA probe to localize the changes that occur after DNA binding. Compared with the free form of TCF4, we found that TCF4 bound to dsDNA contained an additional region of slower exchange (less efficient exchange after 10 s). These differences were observed for the peptides involving amino acid residues 352–367 (ANNARERLRVRDINEA) and

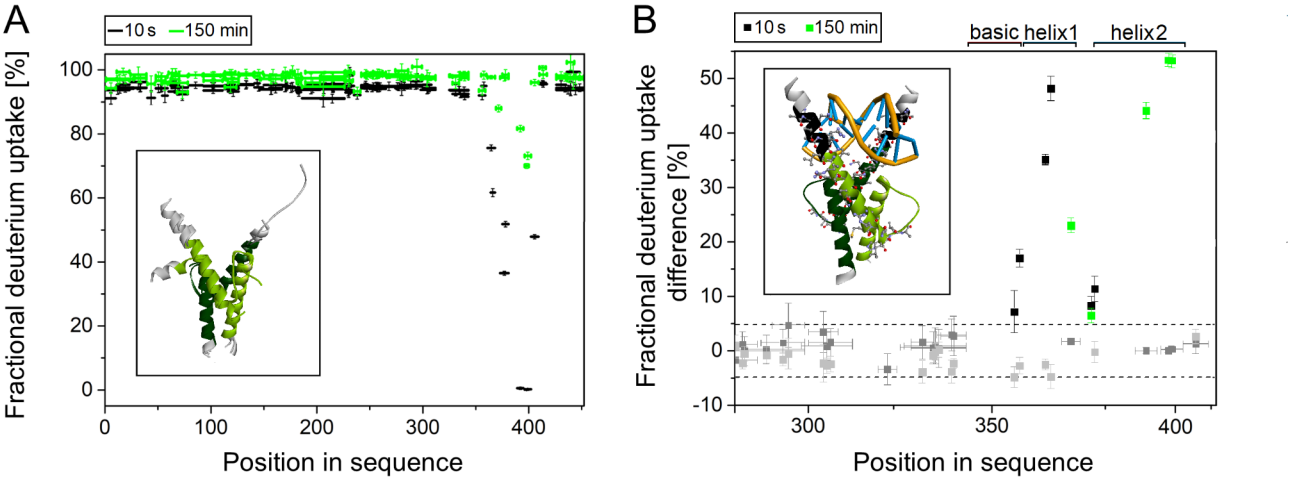

Fig. 4 HDX-MS analysis of TCF4. (**A**) Fractional deuterium uptake of free dTCF4 peptides. The horizontal bars correspond to individual peptides generated by the digestion of TCF4 with nepenthensin-2. The vertical bars indicate the measurement error. The black color represents the results of the measurements after 10 s, and the green color represents the results after 150 min. Inset: Conformations of the *apo* dTCF4 bHLH domain (pdb 6OD3, 6OD4 [21]), with the residues protected from the immediate HDX (360–405) marked light green for one chain and dark green for the other chain of TCF4. (**B**) Fractional deuterium uptake difference between the free and dsDNA-bound forms of dTCF4. The gray squares located within the dashed lines correspond to peptides for which the fractional deuterium uptake difference between the two dTCF4 forms is insignificant (within the 98% CI). The black and green squares above the upper dashed line correspond to peptides of TCF4 that are significantly protected upon the DNA binding. The black color represents the results after 10 s of HDX, and the green color represents the results after 150 min. Inset: dTCF4 bHLH domain bound to the E-box (pdb 6OD4 [21]), with the residues more protected in the complex (352–367 and 372–380) shown as balls and sticks; the residues protected in the complex but exposed in the *apo* form (352–359) are marked in black. Light green corresponds to one chain and dark green to the other chain of TCF4

Table 2 Peptides that differ in fractional deuterium uptake depending on DNA binding by TCF4. A 98% confidence interval was used for the Student's t test

	Sequence	Position	10 s	1 min	5 min	30 min	150 min
basic + helix 1	ANNARER	352–358	Yes	No	No	No	No
	NARERL	354–359	Yes	No	No	No	No
	RVRDINEA	360–367	Yes	Yes	Yes	No	No
	DINEA	363–367	Yes	Yes	Yes	No	No
	FKELGR	368–373	No	Yes	Yes	Yes	Yes
	GRMVQLHLK	372–380	Yes	No	No	No	Yes
	MVQLHLK	374–380	Yes	No	No	No	No
helix 2	LLILHQA	388–394	No	No	Yes	Yes	Yes
	VAVIL	395–399	No	No	No	Yes	Yes
	VAVILSL	395–401	No	No	No	Yes	Yes

372–380 (GRMVQLHLK) (Fig. 4B), and their sequences are listed in Table 2. This finding indicates that the residues 352–359 (ANNARERL) preceding the canonical HLH domain are protected from the immediate HDX while the protein is bound to a specific DNA sequence.

After binding of the DNA, the results of the HDX-MS experiment revealed a slower exchange in the DNA-binding region of TCF4 (Fig. 4B inset, black residues), which could be expected for the complex on the basis of the crystal structures [21]. In contrast, a surprising result was obtained for the second region of TCF4, which is located at the C-terminus of the first α-helix and at the beginning of the loop of the bHLH domain (residues 372–380). The region is not directly engaged in the interaction with the E-box, however, it was observed to be HDX-protected upon the complex formation after 10 s of HDX. Importantly, H378 of the helix 1 of TCF4 has been previously shown to interact with E402 located on the C-terminus of helix 2 of other chain of bHLH, both in *apo* and *holo* forms of TCF4 (pdb 6OD4 [21]),. Analysis of HDX results at different time points (Additional file 2) showed that both the DNA-binding basic region and the C-terminus of the helix 1 of HLH exhibit high dynamics/exposure to solvent, both in *apo* and *holo* forms. The residues initially (after 10 s) protected undergo practically complete HDX after 30 min. In contrast, helix 2 seems to be more stable as it does not exhibit complete HDX, even after 150 min (Additional file 2). What is more, it is way more protected in DNA-bound state even if it has been shown, that only K387 from the N-terminus of helix 2 may interact with

DNA (pdb 6OD4 [21]),. The differences in HDX-protection in this region are visible after longer period of HDX, suggesting that in both forms of TCF4 helix 2 is protected from exposition to solvent, way more in DNA-bound state. Taken together, this data shows that binding of DNA not only change the conformation/secondary structure of the DNA-binding region (b), but also, more importantly, leads to the global changes in the arrangement of helices forming HLH domain.

TCF4 appears exclusively as a dimer regardless of the presence of an E-box

To comprehensively analyze the impact of DNA on TCF4, we decided to investigate the effect of DNA binding on the hydrodynamic properties of TCF4. Our results presented above revealed a significant degree of TCF4 disorder. Given that TCF4 is a relatively large IDP, we were aware that in contrast to globular proteins, MW of TCF4 cannot be calculated from its R_h. Considering the crystals of the dimeric bHLH domain obtained for both the DNA-free and DNA-bound forms of TCF4 [21], we assumed that full-length TCF4 occurs in dimeric form both in the *apo* and *holo* states. We carried out a series of experiments (native PAGE, SEC, and SV-AUC) to test the hydrodynamic properties of the molecules in their free and DNA-bound states. The first experiment we performed was native electrophoresis, the results of which are shown in Fig. 5, inset. The resulting image after DNA binding differs slightly from that of free TCF4. The band moved faster in the polyacrylamide gel, probably because the TCF4 complex gained an additional negative charge carried by the DNA. In both types of samples, we observed single bands demonstrating single populations of molecules with similar properties under native PAGE conditions.

Next, we performed an analytical SEC of free and DNA-bound TCF4. In this technique molecules are separated on the basis of their hydrodynamic properties. The free form of TCF4 showed a single major peak (Fig. 5, black line). Chromatography of the TCF4 sample with DNA revealed two single peaks, the first with an elution volume approximating the major peak coming from the free form of TCF4 and the second indicating the presence of significantly smaller molecules (Fig. 5, red line). By performing a control free DNA chromatography experiment, we classified the second peak as a signal from unbound DNA (Fig. 5, gray line). TCF4 elution profiles indicate the existence of a population of TCF4 molecules with similar properties in each of the samples analyzed, regardless of the presence of DNA. R_h determined from SEC for *apo* and *holo* forms of TCF4 were 64 ± 6 Å and 66 ± 4 Å, respectively. Based on these findings, we can conclude that the hydrodynamic properties

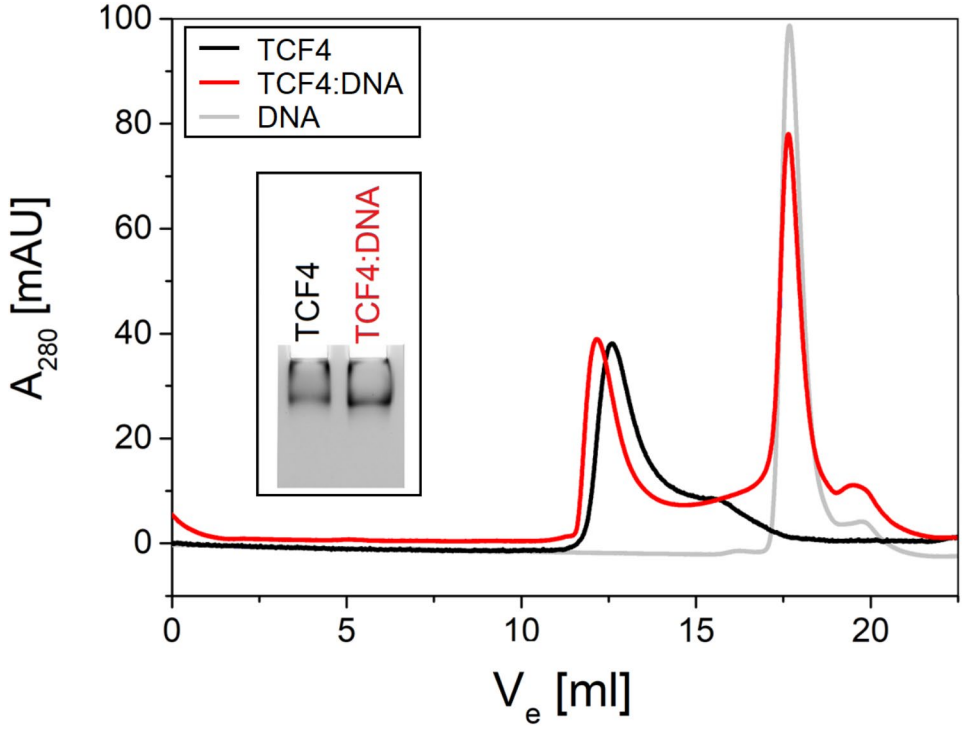

Fig. 5 Determination of TCF4 population homogeneity using native PAGE and SEC. Overlaid representative chromatograms at 280 nm from SEC of free TCF4 (black), DNA-bound TCF4 (red), and free DNA (gray). The absorbance values for the free form of TCF4 were multiplied by a factor of 4.5, for simplified readability of the graph. Inset: Native PAGE of free TCF4 and TCF4 bound with DNA performed on 5% polyacrylamide gel

of the TCF4:DNA complex remain apparently unchanged compared to the *apo* form of TCF4.

To further test the conformational changes in TCF4 induced by DNA binding, we performed an SV-AUC experiment. Representative sedimentation profiles are presented in Fig. 6A and B. The protein was analyzed at three different concentrations in the absence or presence of the ligand in the form of E-box DNA. In this analysis, we focused on determining the changes in the overall shape of dimeric TCF4 in relation to its *apo* and *holo* forms. We applied the continuous model of c(f/f_0) distribution and determined the R_h values of both forms of dTCF4. To do so, we first determined the sedimentation coefficient (s) distribution for dTCF4 and the dTCF4:dsDNA complex (Fig. 6C and D). The best agreement of the experimental data with the s distribution simulated by SEDFIT was obtained for the shape factor values of $f/f_0 = 2.41$ and 2.37 for *apo* dTCF4 and the dTCF4:dsDNA complex, respectively (Fig. 6C and D, Additional file 1: Fig. S8). For the known molecular weights of the *apo* and *holo* forms of dTCF4 (96.5 kDa and 106.4 kDa, respectively), we determined their hydrodynamic radii as $R_h = 72.7 \pm 1.1$ and 67.5 ± 1.1 Å, respectively (Fig. 6C and D, Additional file 1: Fig. S8). The results obtained using SV-AUC for the *holo* form is in good agreement with the R_h value obtained both from the FCS and the SEC experiments. However, there seems to be some discrepancy for the *apo* form of TCF4.

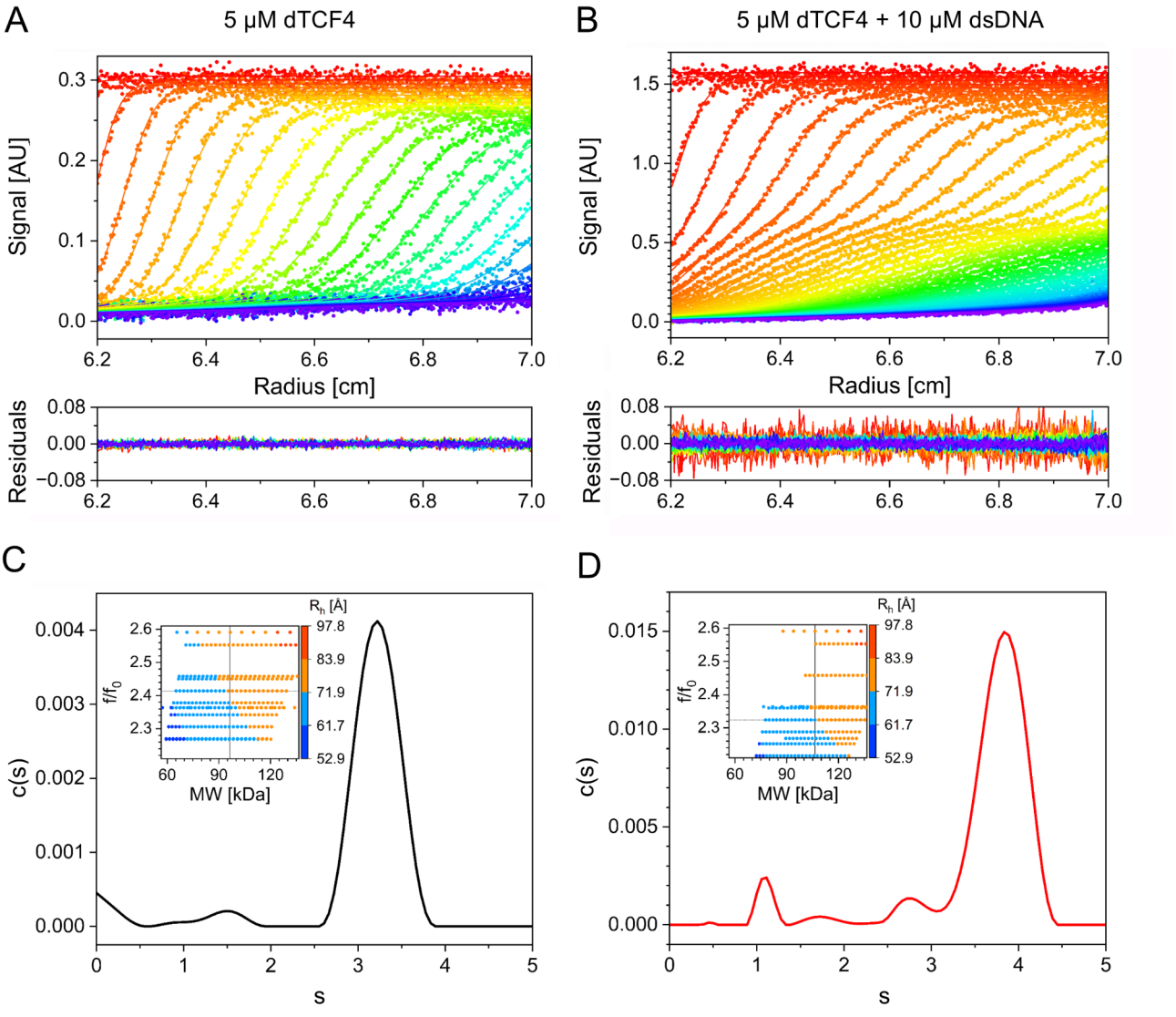

Fig. 6 SV-AUC analysis of free and DNA-bound TCF4. Sedimentation profiles of (**A**) dTCF4 sample without DNA and (**B**) with specific DNA. Sedimentation coefficient (s) distributions of (**C**) free dTCF4 and (**D**) dTCF4 with specific DNA calculated using the c(f/f_0) model for $f/f_0 = 2.41$ and $f/f_0 = 2.32$, respectively. The presented data correspond to dTCF4 analyzed at 5 μM and 5 μM dTCF4 in the presence of 10 μM specific DNA. Insets: determination of the f/f_0 shape factor value, for which the molecular weight distribution generated by SEDFIT centers on the known MW values of (**A**) dTCF4 or the (**B**) dTCF4:dsDNA complex. See Additional file 1: Fig. S8 for details

According to the SEC data, the R_h values for the DNA-free and DNA-bound form of TCF4 are comparable, while SV-AUC experiment indicates that the *apo* form of TCF4 has a larger R_h. In fact, results obtained at similar temperature and solution conditions using different hydrodynamic approaches may not necessarily agree exactly [60], due to intrinsic artifacts of the method, e.g., deviations from ideal SEC behavior due to secondary interactions, including electrostatic [74] and hydrophobic [75], which is reflected by a more than 5-fold greater experimental uncertainty of SEC results compared to AUC. Nevertheless, the results obtained for the *apo* form by both methods are consistent within 2 standard deviations (95% CI) (Fig. 7). Considering that the HDX experiment also indicated a compacting of the basic (b) fragment after DNA binding, we conclude that once the TCF4 dimer binds to DNA, its hydrodynamic radius can slightly decrease. However, even when bound in the complex, dTCF4 exhibits features of very high disorder, since the hydrodynamic dimensions of both forms are still twice as large as for the folded protein of the same chain length (Fig. 7).

DNA binding influences the stability of the TCF4 molecule

It has been reported that some proteins stabilize their structure after binding ligands [78] and it can correlate with the changes in protein flexibility [79]. To investigate the effect of DNA binding to TCF4 on the thermal stability, free and DNA-bound proteins were again analyzed using a thermal shift assay and CD experiments at different temperatures. In the thermal shift experiment, we observed a prominent peak in the fluorescence maximum for DNA-bound TCF4 in the temperature range studied. For free TCF4, the fluorescence was high from the beginning of the heating process, and then, its value increased only slightly (Fig. 8A). The SYPRO® Orange probe exhibits fluorescence after interacting with hydrophobic residues of the protein. The high fluorescence value from the beginning of the process and the lack of the typical transition for free TCF4 may be caused by the exposed hydrophobic residues in the native form of free TCF4. This means that in the *apo* state, TCF4 has exposed regions of the hydrophobic character, which correlates with our HDX results. Interestingly, in the DNA-bound state, the increase in the fluorescence values observed with increasing temperature is sharp and rapid. This suggests that after DNA binding, part of TCF4 folds, leading to the formation of a hydrophobic structure that can bind SYPRO® Orange before a heat-induced conformational transition occurs.

The determination of the first derivative of the fluorescence as a function of temperature allowed the

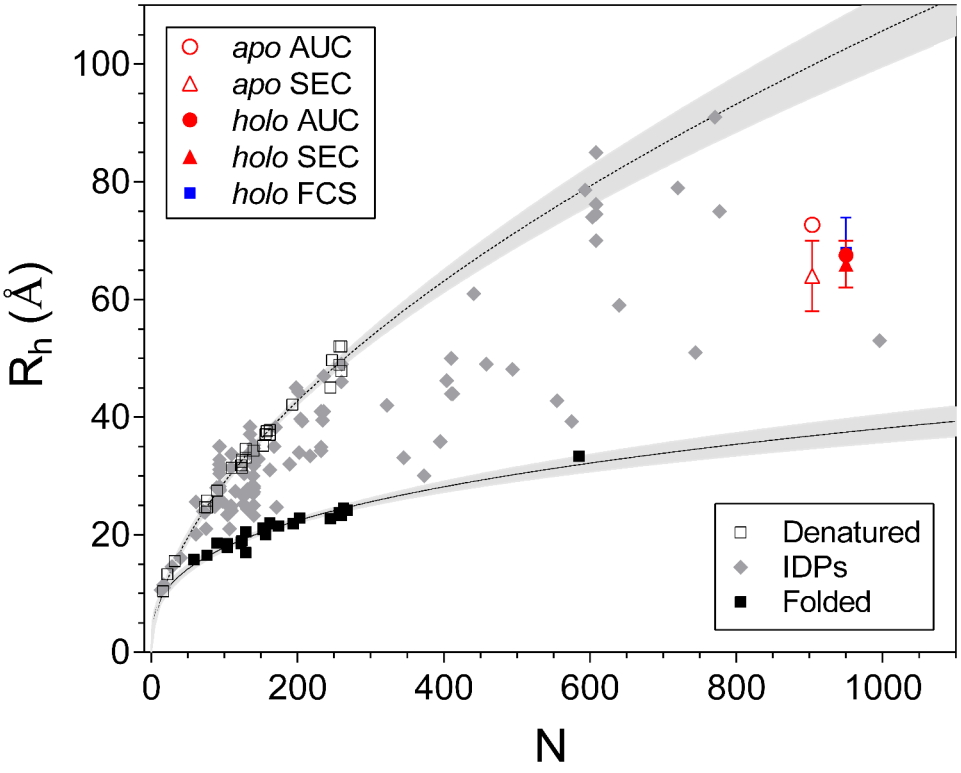

Fig. 7 Hydrodynamic radii, R_h, from AUC, SEC, and FCS for dTCF4 and the dTCF4:dsDNA complex against other proteins. Red empty circle, dTCF4 (AUC); red empty triangle, dTCF4 (SEC); red circle, dTCF4:dsDNA (AUC); red triangle, dTCF4:dsDNA (SEC); blue square, dTCF4:dsDNA (FCS). The points for the complex are shifted to $N = 950$ for clarity. Black empty squares, denatured proteins; gray rhomboids, IDPs; black squares, folded proteins [60, 76, 77]

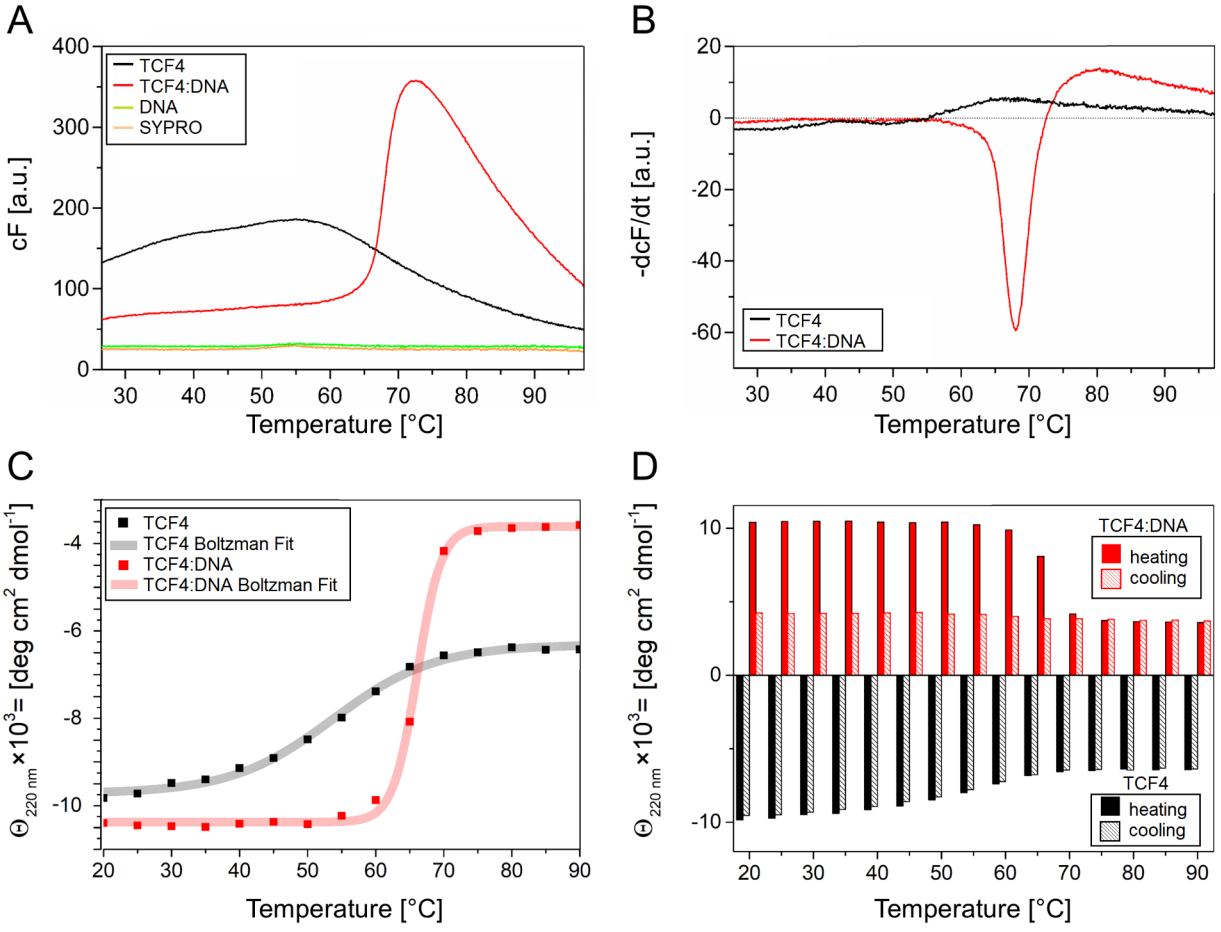

Fig. 8 Investigation of the effect of DNA binding on the thermal stability of TCF4. (**A**) Curves of the dependence of the fluorescence of the SYPRO® Orange probe on the temperature. Black represents the curve for free TCF4, red represents TCF4 in complex with DNA, green represents free DNA, and orange represents free SYPRO® Orange. (**B**) Curves of the first derivatives of fluorescence as a function of temperature. The black color indicates the result for free TCF4, and the red color indicates TCF4 in complex with DNA. (**C**) Curves of the dependence of the molar-residual ellipticity at 220 nm on the temperature (heating process). Black - points and the fitted sigmoidal curve for free TCF4, and red - TCF4 in complex with DNA. (**D**) Bar graph of the dependence of the molar-residual ellipticity at 220 nm on temperature. The black color indicates the values of molar-residual ellipticity for free TCF4 (solid fill — heating process, dashed fill — cooling process), and the red color indicates TCF4 in complex with DNA (solid fill — heating process, dashed fill — cooling process). The ellipticity values for DNA-bound TCF4 are shown as absolute values for simplified readability of the diagram

determination of the melting point of TCF4 in complex with the DNA. The melting temperature was approximately 68 °C. In the case of free TCF4, the melting point was approximately 50 °C (Fig. 8B).

To explore the influence of DNA binding on the stability of TCF4, we decided to use an alternative technique, CD spectroscopy. We recorded CD spectra in the temperature ranges of 20–90 °C and 90–20 °C for free TCF4, and for DNA-bound TCF4. We then plotted the dependence of the ellipticity at 220 nm on temperature during the heating process and fitted sigmoidal curves to the resulting series of points (Fig. 8C). The melting temperatures were approximately 53 °C for free TCF4 and 66 °C for DNA-bound TCF4, which is consistent with the results obtained in the thermal shift assay. We also noted that the decrease in the absolute ellipticity value at 220 nm for free TCF4 was much lower than that for DNA-bound TCF4 (Fig. 8C).

This CD spectroscopy analysis identified an interesting case of free TCF4 behavior, which was revealed by monitoring the ellipticity at 220 nm during cooling. We observed that the absolute value of ellipticity at 220 nm, which decreased with heating, completely returned to its initial value during the cooling process. This effect was not observed for DNA-bound TCF4, which exhibited a typical irreversible transition (Fig. 8D).

Discussion

In this study, we present the results of a detailed biophysical analysis of TCF4 isoform I⁻, which is expressed mainly in the nervous system. This isoform is deficient in amino acids 1-216 in comparison with the canonical isoform B⁻ (UniProt ID P15884-1), and consequently, it

lacks the region comprising AD1, the N-terminal NLS and the CE repressor domain, but it contains the entire bHLH domain [4]. Currently, in the literature, there is no information on the molecular properties of TCF4, although the protein has been proposed previously as an integrator ('hub') of some of bHLH networks regulating important steps of the CNS developmental program [80]. TCF4 has also been associated with the development of neurological disorders, including PTHS. Analyses at genetic level have provided information about the type of mutations that lead to the development of PTHS [7, 29]. Although gene therapy-based approaches have been proposed in recent years for the treatment of PTHS [81, 82], only symptom treatment is being continued.

In the literature, there are examples of some bHLH transcription factors function as transcriptional hubs but their contacts strongly relate to the presence of additional folded motives such as Per-ARNT-Sim (PAS) domains [83]. TCF4 contains no structurally defined domains beside the DNA-binding bHLH domain but still, this protein needs to be able to process various cellular signals and coordinate transcription-related processes. Therefore, it should possess appropriate molecular properties that would allow the protein to form various interactions. The intrinsically disordered character is particularly preferred for proteins with regulatory activity at the crossroads of signaling pathways, as high flexibility allows them to interact with multiple partners in a context-dependent manner [84].

Using the HDX-MS experiment, we demonstrated the fully disordered nature of TCF4 except for the bHLH domain. The disordered conformation is very adaptive, prone to post-translational modifications and often able to fold upon binding [85, 86]. The results of the disorder prediction of the TCF4 presented in our previous study [34] suggested the disordered character of the N-terminal region of TCF4 in contrast to the ordered C-terminal bHLH. To investigate this phenomenon experimentally, we applied HDX coupled with MS since it is known to reveal the conformation of a polypeptide chain and the conformational changes induced by various factors. They are reflected in changes in isotope exchange [87]. The results of HDX-MS presented in this study indicate that binding of the DNA induces secondary structure changes only in the DNA-binding region of TCF4 and short C-terminal fragment of helix 1 of bHLH. Folding of the DNA-binding domain upon binding has been observed for the isolated bHLH domain of TCF4 [21] and other TFs [39–41]. Notably, most of the point mutations associated with PTHS occur in the DNA-binding fragment of the bHLH domain [6, 26, 27, 29, 88–90]. For this reason, further research focused on this part of the protein may provide a deeper understanding of the molecular basis of TCF4-related diseases. Strikingly, compared with analysis of free TCF4, HDX-MS analysis of TCF4 in complex with the E-box did not reveal any secondary structure changes in regions outside the bHLH domain, but it significantly changed the HDX-protection of helix 2, suggesting that binding of the DNA leads to the greater compactness of hydrophobic core of TCF4.

To be able to determine the dissociation constant of TCF4:DNA, we had to obtain a TCF4 free of bacterial nucleic acids. To do this, we had to use a purification method using temporary denaturation, which did not affect the final structure of the purified TCF4 (Additional file 1: Fig. S9), but which effectively removed the nucleic acids (Additional file 1: Fig. S6). This enabled us to demonstrate the ability of TCF4 I^- to bind specific DNA with submicromolar affinity. It has been shown that IDRs of some proteins interact with DNA to ease the interaction of a specific DNA sequence with the binding domain [91–93]. Surprisingly, in the TCF4 case, a comparison of the result of the dissociation constant determined for the isolated bHLH domain (K_D = 0.11 ± 0.04 μM calculated from the FP for the monomer) [21] versus for the full-length protein (K_D = 0.26 ± 0.05 μM calculated from the FP for the monomer) revealed only minor differences.

Notably, dimeric TCF4 I^- is one of the largest IDPs expressed and biophysically analyzed to date (Fig. 7) [60, 76, 77]. Performing shape-related studies of IDPs requires a specific approach [94]. Traditionally, the MW or quaternary structure of globular proteins can be determined on the basis of a direct correlation between the MW and the hydrodynamic properties of the analyzed protein. This principle does not apply to IDPs, which, owing to their dynamic nature, exhibit significantly increased hydrodynamic properties [77]. In this study, we used a less common approach to analyze the conformational changes in disordered proteins induced by ligand binding, i.e., SV-AUC with the f/f_0 model. A similar approach was previously used by Iconaru et al. [95]. Using this method, we were able to demonstrate that the hydrodynamic properties of dimeric TCF4 in its free and DNA-bound forms are very similar, although a slight reduction in R_h after DNA binding is observed, supporting the HDX result suggesting compaction of the hydrophobic core of TCF4 after DNA binding.

Although specific DNA binding does not affect the disordered nature of TCF4 outside the bHLH domain, we observed that it strongly affects the stability of TCF4. In denaturation experiments, we observed that the native-to-denatured transition of the free form of TCF4 is reversible, but once the TCF4:DNA complex is perturbed, the protein irreversibly loses this property. This suggests that in the DNA-free state, the protein is more flexible, whereas in the *holo* state, it presents lower plasticity. The functional consequences of the altered stability of TCF4 in relation to DNA binding are currently

unclear. A flexible structure in the *apo* form clearly has several advantages, i.e., bHLH, which is responsible for both DNA binding and interaction with the dimeric partner, has greater ease of movement. Indeed, the range of motions permitted in an unfolded molecule is less constrained than that permitted in a folded state [96]. This could allow TCF4 molecules in the DNA-free state to dissociate from the wrong dimer partner, providing a second chance to form a complex that can fulfill its function.

We propose that the irreversibility of denaturation of the DNA-bound form of TCF4 could be related to the mechanism by which its activity is inhibited. After transcription initiation, not all the components of the initiation complex dissociate from the promoters [97], and the remaining factors can reinitiate transcription for an extended period of time. Based on our observations indicating that TCF4 is susceptible to irreversible loss of molecular properties only in the DNA-bound state, we propose that certain factors may irreversibly inhibit the activity of some TFs only when they are in the *holo* state. Consequently, TFs can be removed from the pool of active transcription proteins. Notably, the DNA-binding properties of TCF4 and other E proteins are regulated by Ca^{2+}-binding proteins such as calmodulin (CaM), S100a and S100b [98–101]. Ca^{2+}-dependent proteins interact directly with the basic region of the bHLH domain involved in DNA binding [99], and in the presence of Ca^{2+} can selectively inhibit DNA binding by E proteins [98]. Although it has been shown that altering the levels of CaM or Ca^{2+} in cells has a direct effect on E protein-mediated transcription [102, 103], it has not been shown whether CaM binding affects TCF4 stability.

Conclusions

In conclusion, this study is the first in vitro analysis of the molecular properties of the full-length class I bHLH transcription factor. We showed that approximately 90% of TCF4 is composed of disordered regions. Only the HLH domain responsible for dimerization possesses a well-ordered fold. In addition, the basic fragment (b), which is important for E-box binding, actually folds upon binding. These dual structural properties may explain the involvement of TCF4 in the processes that initiate the expression of many genes, allowing TCF4 to act as a "hub" for the regulation of transcription [80]. In this context, we propose that the protein functions in two independent modes. One is related to the function of the bHLH domain and depends on the recognition and binding of the protein to specific DNA sequences, resulting in the anchoring of TCF4 to the appropriate genomic location. In the other mode, TCF4 IDRs act as a flexible platform for multiple interactions with other transcription-related proteins independent of DNA binding. Although further research is needed to gain a comprehensive understanding of how TCF4 uses the flexibility of its IDRs to participate in the regulation of different biological processes, we believe that our study provides a solid basis for further investigations into the mechanism of the biological activity of TCF4 and other bHLH proteins.

Abbreviations

ASCL1	Achaete-scute homolog 1
ATOH1	Atonal bHLH transcription factor 1
bHLH	Basic helix-loop-helix
CaM	Calmodulin
CD	Circular dichroism
cv	Column volumes
DOL	Degree of labeling
dsDNA	Double-stranded DNA
dTCF4	Dimeric TCF4
E-box	Ephrussi box
EMSA	Electrophoretic mobility shift assay
FAM	6-carboxy-fluorescein
FCS	Fluorescence correlation spectroscopy
FP	Fluorescence polarization
HDX-MS	Hydrogen/deuterium exchange mass spectrometry
HES	Hairy and enhancer of split
ID	Intellectual disability
ID-2	DNA-binding protein inhibitor ID-2
IDPs	Intrinsically disordered proteins
IDRs	Intrinsically disordered regions
IMAC	Immobilized metal affinity chromatography
ITF2	Immunoglobulin transcription factor 2
K_D	Dissociation constant
MW	Molecular weight
PAS	Per-ARNT-Sim
PTHS	Pitt-Hopkins syndrome
R_h	Hydrodynamic radius
SEC	Size exclusion chromatography
SEF2	SL3-3 enhancer factor 2
SV-AUC	Sedimentation-velocity analytical ultracentrifugation
TCF4	Transcription factor 4
TF	Transcription factor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12964-025-02154-7.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We are grateful to Karolina Partyk (WUST) for her excellent technical assistance.

Author contributions

N.S. investigation, visualization, writing - original draft preparation, review and editing; B.P.K. methodology development, investigation, visualization, writing - original draft preparation, review and editing; A.N. conceptualization, methodology development, visualization, writing - original draft preparation, review and editing; L.Z. investigation, writing - review and editing; M.D. writing - review and editing; B.G.M. conceptualization, writing - original draft preparation, review and editing; A.O. writing - review and editing; A.T. conceptualization, investigation, visualization, writing - original draft preparation, review and editing;

Funding

This work was supported by a subsidy from The Polish Ministry of Science and High Education for the Faculty of Chemistry of Wroclaw University of Science and Technology. The work of B.P.K. and A.N. was partially supported

by National Science Centre of Poland Sonata-Bis Grant UMO-2016/22/E/NZ1/00656 to A.N.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biochemistry, Molecular Biology and Biotechnology, Faculty of Chemistry, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, Wrocław 50-370, Poland
[2]Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, Warsaw PL-02668, Poland
[3]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a, Warsaw 02-106, Poland

Received: 27 January 2025 / Accepted: 12 March 2025
Published online: 27 March 2025

References

1. Massari ME, Murre C, Helix-Loop-Helix P. Regulators of transcription in eucaryotic. Mol Cell Biol. 2000;20:429–40. https://doi.org/10.1128/mcb.20.2.429-440.2000.
2. Murre C, McCaw PS, Baltimore DA, New DNA. Binding and dimerization motif in Immunoglobulin enhancer binding, daughterless, MyoD, and Myc proteins. Cell. 1989;56:777–83. https://doi.org/10.1016/0092-8674(89)90682-X.
3. Forrest MP, Hill MJ, Quantock AJ, Martin-Rendon E, Blake DJ. The emerging roles of TCF4 in disease and development. Trends Mol Med. 2014;20:322–31. https://doi.org/10.1016/j.molmed.2014.01.010.
4. Sepp M, Kannike K, Eesmaa A, Urb M. Functional diversity of human basic Helix-Loop-Helix transcription factor TCF4 isoforms generated by alternative 5 9 exon usage and splicing. PLoS ONE. 2011;6. https://doi.org/10.1371/journal.pone.0022138.
5. Pscherer A, Dörflinger U, Kirfel J, Gawlas K, Rüschoff J, Buettner R, Schüle R. The Helix-Loop-Helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin receptor II gene. EMBO J. 1996;15:6680–90. https://doi.org/10.1002/j.1460-2075.1996.tb01058.x.
6. De Pontual L, Mathieu Y, Golzio C, Rio M, Malan V, Boddaert N, Soufflet C, Picard C, Durandy A, Dobbie A, et al. Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. Hum Mutat. 2009;30:669–76. https://doi.org/10.1002/humu.20935.
7. Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ. Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. Hum Mutat. 2012;33:1676–86. https://doi.org/10.1002/humu.22160.
8. Petropoulos H, Skerjanc IS. Analysis of the Inhibition of myod activity by ITF-2B and Full-Length E12/E47. J Biol Chem. 2000;275:25095–101. https://doi.org/10.1074/jbc.M004251200.
9. Lu Y, Sheng D-Q, Mo Z-C, Li H-F, Wu N-H, Shen Y-F. A negative regulatory Element-Dependent inhibitory role of ITF2B on IL-2 receptor α gene. Biochem Biophys Res Commun. 2005;336:142–9. https://doi.org/10.1016/j.bbrc.2005.08.050.
10. Furumura M, Potterf SB, Toyofuku K, Matsunaga J, Muller J, Hearing VJ. Involvement of ITF2 in the transcriptional regulation of melanogenic genes. J Biol Chem. 2001;276:28147–54. https://doi.org/10.1074/jbc.M101626200.
11. Flora A, Garcia JJ, Thaller C, Zoghbi HY. The E-Protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc Natl Acad Sci U S A. 2007;104:15382–7. https://doi.org/10.1073/pnas.0707456104.
12. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. Nat Rev Neurosci. 2002;3:517–30. https://doi.org/10.1038/nrn874.
13. Forrest MP, Waite AJ, Martin-Rendon E, Blake DJ. Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. PLoS ONE. 2013;8:e73169. https://doi.org/10.1371/journal.pone.0073169.
14. Akazawa C, Sasai Y, Nakanishi S, Kageyama R. Molecular characterization of a rat negative regulator with a basic Helix-Loop-Helix structure predominantly expressed in the developing nervous system. J Biol Chem. 1992;267:21879–85. https://doi.org/10.1016/S0021-9258(19)36694-3.
15. Sasai Y, Kageyama R, Tagawa Y, Shigemoto R, Nakanishi S. Two mammalian Helix-Loop-Helix factors structurally related to drosophila hairy and enhancer of split. Genes Dev. 1992;6:2620–34. https://doi.org/10.1101/gad.6.12b.2620.
16. Voronova A, Baltimore D, Mutations That Disrupt DNA. Binding and Dimer Formation in the E47 Helix-Loop-Helix Protein Map to Distinct Domains. Proc. Natl. Acad. Sci. U. S. A. 1990;87:4722–4726. https://doi.org/10.1073/pnas.87.12.4722
17. Ellenberger T, Arnaud M, Harrison SC. Crystal structure of transcription factor E47: E-Box recognition by a basic region Helix-Loop-Helix dimer. Genes Dev. 1994;8:970–80. https://doi.org/10.1101/gad.8.8.970.
18. Ephrussi A, Church GM, Tonegawa S, Gilbert W. B Lineage—Specific Interactions of an Immunoglobulin Enhancer with Cellular Factors in Vivo. Science (80-.). 1985;227:134–140. https://doi.org/10.1126/science.3917574
19. Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK. Recognition by max of its cognate DNA through a dimeric B/HLH/Z domain. Nature. 1993;363:38–45. https://doi.org/10.1038/363038a0.
20. Longo A, Guanga GP, Rose RB. Crystal structure of E47-NeuroD1/Beta2 BHLH Domain-DNA complex: heterodimer selectivity and DNA recognition. Biochemistry. 2008;47:218–29. https://doi.org/10.1021/bi701527r.
21. Yang J, Horton JR, Li J, Huang Y, Zhang X, Blumenthal RM, Cheng X. Structural basis for Preferential binding of human TCF4 to DNA containing 5-Carboxylcytosine. Nucleic Acids Res. 2019;47:8375–87. https://doi.org/10.1093/nar/gkz381.
22. Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. Interactions between heterologous Helix-Loop-Helix proteins generate complexes that bind specifically to a common DNA sequence. Cell. 1989;58:537–44. https://doi.org/10.1016/0092-8674(89)90434-0.
23. Henthorn P, Kiledjian M, Kadesch T. Two distinct transcription factors that bind the Immunoglobulin enhancer ME5/KE2 motif. Sci (80-). 1990;247:467–70. https://doi.org/10.1126/science.2105528.
24. Corneliussen B, Thornell A, Hallberg B, Grundström T. Helix-Loop-Helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. J Virol. 1991;65:6084–93. https://doi.org/10.1128/jvi.65.11.6084-6093.1991.
25. Yoon SO, Chikaraishi DM. Isolation of two E-Box binding factors that interact with the rat tyrosine hydroxylase enhancer. J Biol Chem. 1994;269:18453–62. https://doi.org/10.1016/S0021-9258(17)32330-X.
26. Amiel J, Rio M, Pontual L. Mutations in TCF4, encoding a class I basic Helix-Loop-Helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet. 2007;80:988–93. https://doi.org/10.1086/515582. RedonR.Malan, V.; Boddaert, N.; Plouin, P.; Carter, N.P.; Lyonnet, S.; Munnich, A.; et al.
27. Zweier C, Peippo MM, Hoyer J, Sousa S, Bottani A, Clayton-Smith J, Reardon W, Saraiva J, Cabral A, Göhring I, et al. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins Syndrome). Am J Hum Genet. 2007;80:994–1001. https://doi.org/10.1086/515583.
28. Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, Frenck W, Radlwimmer B, Lichter P, Engels H, et al. Severe mental retardation with breathing abnormalities (Pitt–Hopkins Syndrome) is caused by haploinsufficiency of the neuronal BHLH transcription factor TCF4. Hum Mol Genet. 2007;16:1488–94. https://doi.org/10.1093/hmg/ddm099.
29. Sepp M, Pruunsild P, Timmusk T. Pitt–Hopkins Syndrome-Associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to Dominant-Negative effects. Hum Mol Genet. 2012;21:2873–88. https://doi.org/10.1093/hmg/dds112.
30. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OPH, Mors O, Mortensen PB, et al. Common Variants

Conferring Risk Schizophrenia Nat. 2009;460:744–7. https://doi.org/10.1038/nature08186.
31. Hu X, Zhang B, Liu W, Paciga S, He W, Lanz TA, Kleiman R, Dougherty B, Hall SK, McIntosh AM, et al. A survey of rare coding variants in candidate genes in schizophrenia by deep sequencing. Mol Psychiatry. 2014;19:858–9. https://doi.org/10.1038/mp.2013.131.
32. Basmanav FB, Forstner AJ, Fier H, Herms S, Meier S, Degenhardt F, Hoffmann P, Barth S, Fricker N, Strohmaier J, et al. Investigation of the role of TCF4 rare sequence variants in schizophrenia. Am J Med Genet Part B Neuropsychiatr Genet. 2015;168:354–62. https://doi.org/10.1002/ajmg.b.32318.
33. Kim S, Morgunova E, Naqvi S, Goovaerts S, Bader M, Koska M, Popov A, Luong C, Pogson A, Swigut T, et al. DNA-Guided transcription factor cooperativity shapes face and limb mesenchyme. Cell. 2024;187:692–e71126. https://doi.org/10.1016/j.cell.2023.12.032.
34. Greb-Markiewicz B, Kazana W, Zarębski M, Ożyhar A. The subcellular localization of BHLH transcription factor TCF4 is mediated by multiple nuclear localization and nuclear export signals. Sci Rep. 2019;9:15629. https://doi.org/10.1038/s41598-019-52239-w.
35. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry. 2006;45:6873–88. https://doi.org/10.1021/bi0602718.
36. Minezaki Y, Homma K, Kinjo AR, Nishikawa K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. J Mol Biol. 2006;359:1137–49. https://doi.org/10.1016/j.jmb.2006.04.016.
37. Quong MW, Massari ME, Zwart R, Murre CA. New Transcriptional-Activation motif restricted to a class of Helix-Loop-Helix proteins is functionally conserved in both yeast and mammalian cells. Mol Cell Biol. 1993;13:792–800. https://doi.org/10.1128/mcb.13.2.792.
38. Massari ME, Jennings PA, Murre C. The AD1 transactivation domain of E2A contains a highly conserved helix which is required for its activity in both Saccharomyces cerevisiae and mammalian cells. Mol Cell Biol. 1996;16:121–9. https://doi.org/10.1128/mcb.16.1.121.
39. Weiss MA, Ellenberger T, Wobbe CR, Lee JP, Harrison SC, Struhl K. Folding transition in the DMA-Binding domain of GCN4 on specific binding to DNA. Nature. 1990;347:575–8. https://doi.org/10.1038/347575a0.
40. Podust LM, Krezel AM, Kim Y. Crystal structure of the CCAAT Box/Enhancer-Binding protein B activating transcription Factor-4 basic leucine zipper heterodimer in the absence of DNA. J Biol Chem. 2001;276:505–13. https://doi.org/10.1074/jbc.M005594200.
41. Neil O K.T., Shuman J.D., Ampe C, DeGrado W.F. DNA-Induced increase in the. Alpha.-Helical content of C/EBP and GCN4. Biochemistry. 1991;30:9030–4. https://doi.org/10.1021/bi00101a017.
42. Romero O, Dunker K. Sequence data analysis for long disordered regions prediction in the calcineurin family. Genome Inf Ser Workshop Genome Inf. 1997;8:110–24.
43. Mészáros B, Erdös G, Dosztányi Z. IUPred2A: Context-Dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 2018;46:W329–37. https://doi.org/10.1093/nar/gky384.
44. Erdős G, Dosztányi Z. Analyzing protein disorder with IUPred2A. Curr Protoc Bioinforma. 2020;70. https://doi.org/10.1002/cpbi.99.
45. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021;49:W297–303. https://doi.org/10.1093/nar/gkab408.
46. Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Sønderby CK, Sommer MOA, Winther O, Nielsen M, Petersen B, et al. NetSurfP-2.0: improved prediction of protein structural features by integrated deep learning. Proteins Struct Funct Bioinforma. 2019;87:520–7. https://doi.org/10.1002/prot.25674.
47. Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. From protein sequence to dynamics and disorder with dynamine. Nat Commun. 2013;4:2741. https://doi.org/10.1038/ncomms3741.
48. Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. The dynamine webserver: predicting protein dynamics from sequence. Nucleic Acids Res. 2014;42:W264–70. https://doi.org/10.1093/nar/gku270.
49. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic Acids Res. 2019;47:W402–7. https://doi.org/10.1093/nar/gkz297.
50. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82. https://doi.org/10.1038/s41592-022-01488-1.
51. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with alphafold. Nature. 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.
52. Weeks SD, Drinker M, Loll PJ. Ligation independent cloning vectors for expression of SUMO fusions. Protein Expr Purif. 2007;53:40–50. https://doi.org/10.1016/j.pep.2006.12.006.
53. Unger T, Jacobovitch Y, Dantes A, Bernheim R, Peleg Y. Applications of the restriction free (RF) cloning procedure for molecular manipulations and protein expression. J Struct Biol. 2010;172:34–44. https://doi.org/10.1016/j.jsb.2010.06.016.
54. Bond SR, Naus CC. RF-Cloning.Org: an online tool for the design of Restriction-Free cloning projects. Nucleic Acids Res. 2012;40:W209–13. https://doi.org/10.1093/nar/gks396.
55. Schneider CA, Rasband WS, Eliceiri KW. NIH image to imageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5. https://doi.org/10.1038/nmeth.2089.
56. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970. https://doi.org/10.1038/227680a0.
57. Fairbanks G, Steck TL, Wallach DFH. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry. 1971;10:2606–17. https://doi.org/10.1021/bi00789a030.
58. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. Trends Biochem Sci. 2008;33:220–9. https://doi.org/10.1016/j.tibs.2008.02.002.
59. Petrášek Z, Schwille P. Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. Biophys J. 2008;94:1437–48. https://doi.org/10.1529/biophysj.107.108811.
60. Białobrzewski MK, Klepka BP, Michaś A, Cieplak-Rotowska MK, Staszałek Z, Niedźwiecka A. Diversity of hydrodynamic radii of intrinsically disordered proteins. Eur Biophys J. 2023;52:607–18. https://doi.org/10.1007/s00249-023-01683-8.
61. Taylor JR. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. 2nd Edition; University Science Books, Sausalito, 1997; ISBN 9780935702750.
62. Ryder SP, Recht MI, Williamson JR. Quantitative analysis of Protein-RNA interactions by gel mobility shift. Methods Mol Biol. 2008;488:99–115. https://doi.org/10.1007/978-1-60327-475-3_7.
63. Niedzwiecka A, Marcotrigiano J, Stepinski J, Jankowska-Anyszka M, Wyslouch-Cieszynska A, Dadlez M, Gingras A-C, Mak P, Darzynkiewicz E, Sonenberg N, et al. Biophysical studies of EIF4E cap-Binding protein: recognition of MRNA 5′ cap structure and synthetic fragments of EIF4G and 4E-BP1 proteins. J Mol Biol. 2002;319:615–35. https://doi.org/10.1016/S0022-2836(02)00328-5.
64. Puchała W, Burdukiewicz M, Kistowski M, Dąbrowska KA, Badaczewska-Dawid AE, Cysewski D, Dadlez M. HaDeX: an R package and Web-Server for analysis of data from Hydrogen–Deuterium exchange mass spectrometry experiments. Bioinformatics. 2020;36:4516–8. https://doi.org/10.1093/bioinformatics/btaa587.
65. Schimmel CR, Cantor PR, Biophysical Chemistry II. 10– 2. Techniques for the study of biological structure and function, chap. In; W. H. Freeman and Co.; 1980.
66. Laue T, Shah B, Ridgeway T, Pelletier S. Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins. In Analytical ultracentrifugation in biochemistry and polymer science; S.E. Harding, J.C.H. and A.J.R., Ed.; Royal Society of Chemistry: Cambridge, U.K., 1992; pp. 90–125 ISBN 0851863450.
67. Philo JS, SEDNTERP. A calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. Eur Biophys J. 2023;52:233–66. https://doi.org/10.1007/s00249-023-01629-0.
68. Vaitiekunas P, Crane-Robinson C, Privalov PL. The energetic basis of the DNA double helix: A combined microcalorimetric approach. Nucleic Acids Res. 2015;43:8577–89. https://doi.org/10.1093/nar/gkv812.
69. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta - Proteins Proteom. 2005;1751:119–39. https://doi.org/10.1016/j.bbapap.2005.06.005.
70. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. Protein Sci. 2002;11:739–56. https://doi.org/10.1110/ps.4210102.
71. Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins. J Mol Biol. 2021;433:167208. https://doi.org/10.1016/j.jmb.2021.167208.
72. Chalmers MJ, Busby SA, Pascal BD, Southern MR, Griffin PR. A Two-Stage differential hydrogen deuterium exchange method for the rapid characterization of protein/ligand interactions. J Biomol Tech. 2007;18:194–204.
73. Cieplak-Rotowska MK, Tarnowski K, Rubin M, Fabian MR, Sonenberg N, Dadlez M, Niedzwiecka A. Structural dynamics of the GW182 Silencing domain

including its RNA recognition motif (RRM) revealed by Hydrogen-Deuterium exchange mass spectrometry. J Am Soc Mass Spectrom. 2018;29:158–73. https://doi.org/10.1007/s13361-017-1830-9.
74. Dubin PL, editor.; 1st ed.; Elsevier: Amsterdam, 1988; ISBN 9780080858449.
75. Bywater RP, Marsden NV. In: Heftmann E, editor. Chromatography: fundamentals and applications of chromatographic and electrophoretic methods. Amsterdam: Elsevier; 1988.
76. Khaymina SS, Kenney JM, Schroeter MM, Chalovich JM. Fesselin is a natively unfolded protein. J Proteome Res. 2007;6:3648–54. https://doi.org/10.1021/pr070237v.
77. Waszkiewicz R, Michaś A, Białobrzewski MK, Klepka BP, Cieplak-Rotowska MK, Staszałek Z, Cichocki B, Lisicki M, Szymczak P, Niedzwiecka A. Hydrodynamic radii of intrinsically disordered proteins: fast prediction by minimum dissipation approximation and experimental validation. J Phys Chem Lett. 2024;15:5024–33. https://doi.org/10.1021/acs.jpclett.4c00312.
78. Waldron TT, Murphy KP. Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. Biochemistry. 2003;42:5058–64. https://doi.org/10.1021/bi034212v.
79. Celej MS, Montich GG, Fidelio GD. Protein stability induced by ligand binding correlates with changes in protein flexibility. Protein Sci. 2003;12:1496–506. https://doi.org/10.1110/ps.0240003.
80. Quednow BB, Brzózka MM, Rossner MJ. Transcription factor 4 (TCF4) and schizophrenia: integrating the animal and the human perspective. Cell Mol Life Sci. 2014;71:2815–35. https://doi.org/10.1007/s00018-013-1553-4.
81. Kim H, Gao EB, Draper A, Berens NC, Vihma H, Zhang X, Higashi-Howard A, Ritola KD, Simon JM, Kennedy AJ, et al. Rescue of behavioral and electrophysiological phenotypes in a Pitt-Hopkins syndrome mouse model by genetic restoration of Tcf4 expression. Elife. 2022;11:1–24. https://doi.org/10.7554/eLife.72290.
82. Papes F, Camargo AP, de Souza JS, Carvalho VMA, Szeto RA, LaMontagne E, Teixeira JR, Avansini SH, Sánchez-Sánchez SM, Nakahara TS, et al. Transcription factor 4 Loss-of-Function is associated with deficits in progenitor proliferation and cortical neuron content. Nat Commun. 2022;13. https://doi.org/10.1038/s41467-022-29942-w.
83. Möglich A, Ayers RA, Moffat K. Structure and signaling mechanism of Per-ARNT-Sim domains. Structure. 2009;17:1282–94. https://doi.org/10.1016/j.str.2009.08.011.
84. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16:18–29. https://doi.org/10.1038/nrm3920.
85. Tompa P, Schad E, Tantos A, Kalmar L. Intrinsically disordered proteins: emerging interaction specialists. Curr Opin Struct Biol. 2015;35:49–59. https://doi.org/10.1016/j.sbi.2015.08.009.
86. Uversky VN. Analyzing IDPs in interactomes. Methods Mol Biol. 2020;2141:895–945. https://doi.org/10.1007/978-1-0716-0524-0_46.
87. Vinciauskaite V, Masson GR. Fundamentals of HDX-MS. Essays Biochem. 2023;67:301–14. https://doi.org/10.1042/EBC20220111.
88. Zweier C, Sticht H, Bijlsma EK, Clayton-Smith J, Boonen SE, Fryer A, Greally MT, Hoffmann L, den Hollander NS, Jongmans M, et al. Further delineation of Pitt-Hopkins syndrome: phenotypic and genotypic description of 16 novel patients. J Med Genet. 2008;45:738–44. https://doi.org/10.1136/jmg.2008.060129.
89. Giurgea I, Missirian C, Cacciagli P, Whalen S, Fredriksen T, Gaillon T, Rankin J, Mathieu-Dramard M, Morin G, Martin-Coignard D, et al. TCF4 deletions in Pitt-Hopkins syndrome. Hum Mutat. 2008;29:242–51. https://doi.org/10.1002/humu.20859.
90. Whalen S, Héron D, Gaillon T, Moldovan O, Rossi M, Devillard F, Giuliano F, Soares G, Mathieu-Dramard M, Afenjar A, et al. Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. Hum Mutat. 2012;33:64–72. https://doi.org/10.1002/humu.21639.
91. de Jonge WJ, Patel HP, Meeussen JVW, Lenstra TL. Following the tracks: how transcription factor binding dynamics control transcription. Biophys J. 2022;121:1583–92. https://doi.org/10.1016/j.bpj.2022.03.026.
92. Chappleboim M, Naveh-Tassa S, Carmi M, Levy Y, Barkai N. Ordered and disordered regions of the origin recognition complex direct differential in vivo binding at distinct motif sequences. Nucleic Acids Res. 2024;52:5720–31. https://doi.org/10.1093/nar/gkae249.
93. Hurieva B, Kumar DK, Morag R, Lupo O, Carmi M, Barkai N, Jonas F. Disordered sequences of transcription factors regulate genomic binding by integrating diverse sequence grammars and interaction types. Nucleic Acids Res. 2024;52:8763–77. https://doi.org/10.1093/nar/gkae521.
94. Maiti S, Singh A, Maji T, Saibo NV, De S. Experimental methods to study the structure and dynamics of intrinsically disordered regions in proteins. Curr Res Struct Biol. 2024;7:100138. https://doi.org/10.1016/j.crstbi.2024.100138.
95. Iconaru LI, Das S, Nourse A, Shelat AA, Zuo J, Kriwacki RW. Small molecule sequestration of the intrinsically disordered protein, P27Kip1, within soluble oligomers. J Mol Biol. 2021;433:167120. https://doi.org/10.1016/j.jmb.2021.167120.
96. Frauenfelder H, Petsko GA, Tsernoglou D, Temperature-Dependent. X-Ray diffraction as a probe of protein structural dynamics. Nature. 1979;280:558–63. https://doi.org/10.1038/280558a0.
97. Pomp W, Meeussen JVW, Lenstra TL. Transcription factor exchange enables prolonged transcriptional bursts. Mol Cell. 2024;84:1036–e10489. https://doi.org/10.1016/j.molcel.2024.01.020.
98. Corneliussen B, Holm M, Waltersson Y, Onions J, Hallberg B, Thornell A, Grundström T. Calcium/Calmodulin Inhibition of Basic-Helix-Loop-Helix transcription factor domains. Nature. 1994;368:760–4. https://doi.org/10.1038/368760a0.
99. Onions J, Hermann S, Grundström T. Basic Helix-Loop-Helix protein sequences determining differential Inhibition by calmodulin and S-100 proteins. J Biol Chem. 1997;272:23930–7. https://doi.org/10.1074/jbc.272.38.23930.
100. Larsson G, Schleucher J, Onions J, Hermann S, Grundström T, Wijmenga SS. A novel target recognition revealed by calmodulin in complex with the basic Helix–Loop–Helix transcription factor SEF2-1/E2-2. Protein Sci. 2001;10:169–86. https://doi.org/10.1110/ps.28401.
101. Larsson G, Schleucher J, Onions J, Hermann S, Grundström T, Wijmenga SS. Backbone dynamics of a symmetric calmodulin dimer in complex with the calmodulin-Binding domain of the Basic-Helix-Loop-Helix transcription factor SEF2-1/E2-2: A highly dynamic complex. Biophys J. 2005;89:1214–26. https://doi.org/10.1529/biophysj.104.055780.
102. Hauser J, Sveshnikova N, Wallenius A, Baradaran S, Saarikettu J, Grundström T. B-Cell receptor activation inhibits AID expression through calmodulin Inhibition of E-Proteins. Proc Natl Acad Sci. 2008;105:1267–72. https://doi.org/10.1073/pnas.0708220105.
103. Hauser J, Saarikettu J, Grundström T. Calcium regulation of myogenesis by differential calmodulin Inhibition of basic Helix-Loop-Helix transcription factors. Mol Biol Cell. 2008;19:2509–19. https://doi.org/10.1091/mbc.e07-09-0886.

Publisher's note