

**Physicochemical studies on interaction between intrinsically
disordered regions in transcription factors Sp1 and TAF4**

転写因子 Sp1 と TAF4 の天然変性領域を介した
相互作用に関する物理化学的研究

2016

Emi Hibino

Contents

Abbreviations	4
Preface	5
 Chapter 1	
NMR analysis of transcriptional activation domains in Sp1 and TAF4	
1-1. Introduction	9
1-2. Materials and Methods	10
1-3. Results	
1-3-1. ^1H - ^{15}N HSQC spectra of Q-domains in Sp1 and TAF4	14
1-3-2. CD spectra of Q-domains in Sp1 and TAF4	17
1-3-3. SAXS analysis of Q-domains in Sp1 and TAF4	19
1-4. Discussion	
1-4-1. Structural features of Q-domains in Sp1 and TAF4	21
1-4-2. TAF4N/C and Sp1-QB may be divided into several fragment proteins	21
1-4-3. TAF4N/C was partially structured, but largely disordered	22
 Chapter 2	
Identification of heteromolecular binding sites in transcription factors Sp1 and TAF4 using NMR spectroscopy	
2-1. Introduction	24
2-2. Materials and Methods	25
2-3. Results	
2-3-1. Interaction between Q-domains of Sp1 and TAF4 elucidated by NMR spectroscopy	29
2-3-2. Surface plasmon resonance (SPR)	31
2-3-3. Optimization of experimental conditions and NMR signal assignment	32
2-3-4. Identification of the Sp1-interaction site in the TAF4 molecule	35
2-3-5. The interaction between fragment proteins of Sp1-QB and TAF4N/C	39
2-3-6. No significant conformational change upon binding	46
2-4. Discussion	
2-4-1. The interaction between Q-domains of Sp1 and TAF4	49

2-4-2. C-terminal half of the Sp1-QB domain is responsible for the interaction with TAF4	49
2-4-3. TAF4-Q1 is responsible for the interaction with Sp1-QB	50
2-4-4. Quantitative analysis of the interaction between Q-domains of Sp1 and TAF4	53
2-4-5. Self-association of Sp1-QB versus the heterogeneous interaction with TAF4	53
2-4-6. A rare interaction mode of the disordered region	55
 Chapter 3	
Elucidation of mechanism for molecular recognition on the formation of Sp1-TAF4 complex	
3-1. Introduction	58
3-2. Materials and Methods	59
3-3. Results	
3-3-1. Analyses of the interactions of Sp1-QBc mutants with TAF4-Q12	60
3-3-2. Comparing NMR spectra of wild type and mutant proteins of Sp1-QBc	61
3-3-3. The analysis of Sp1-QBc WA and the complex formation of wild type Sp1-QBc with TAF4-Q12	64
3-3-4. Investigation of conformation changes in Sp1	65
3-4. Discussion	
3-4-1. Re-examination of the interaction of Sp1-QBc mutants to TAF4 by NMR	66
3-4-2. A possible formation of transient hydrophobic core	67
3-4-3. The conformational changes introduced to the region distant from the binding site	68
 Conclusions	70
Acknowledgements	71
References	72

Abbreviations

CD	circular dichroism
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
DTT	dithiothreitol
GST	glutathione S-transferase
HSQC	heteronuclear single-quantum coherence
IDP	intrinsically disordered protein
IPTG	isopropyl- β -D-thiogalactopyranoside
K_a	association constant
k_{ex}	exchange rate constant
L ₃ A ₃	Sp1 L467A/L469A/L472A mutant
Ni-NTA	nickel-nitrilotriacetic acid
NOE	nuclear Overhauser effect
PF	Photon Factory
Q-domain	glutamine-rich domain
QNA ₃	Sp1 Q468A/Q470A/N471A mutant
R_g	radius of gyration
SAXS	small-angle X-ray scattering
Sp1	specificity protein 1
SPR	surface plasmon resonance
TAF4	TBP-associated factors 4
TAF4N/C	central fragment protein corresponding to human TAF4 (408 and 838)
TBP	TATA box-binding protein
TFIID	transcription factor IID
Ub	human ubiquitin
WA	Sp1 W464A mutant
YUH-1	yeast ubiquitin hydrolase-1

Preface

The expression of genes is controlled in a temporally and specifically coordinated manner, and this precise regulation is accomplished predominantly at the level of transcription. A gene in a eukaryotic cell typically consists of two types of DNA elements: a common core-promoter element that is proximal to the transcription start site, and gene-specific enhancer sequences that are located at more distal positions¹. The core-promoter element is recognized by the general transcriptional machinery to form a pre-initiation complex, which contains RNA polymerase II and the general transcription initiation factors TFIIA, -B, -D, -E, -F, and -H^{2,3} (**Fig. P-1**).

In contrast, enhancer elements are recognized by gene-specific transcriptional regulators, which typically consist of a DNA-binding domain and one or more activation domains^{4,5}. In the absence of these activators, the expression of most genes *in vivo* is silenced by the chromatin structure as well as the effects of other repressors. Furthermore, RNA polymerase II and cognate general initiation factors, without the aid of activators, only exhibit a low level of transcriptional activity *in vitro*. Therefore, the mechanisms by which enhancer selective transcriptional factors stimulate the transcriptional activity of the general transcriptional machinery need to be elucidated in order to obtain a detailed understanding of the regulation of gene expression.

The general transcription factor TFIID is considered to play a central role during the formation of a pre-initiation complex because it is the first component to recognize and bind the core promoter, and also provides a scaffold upon which the remaining transcriptional factors assemble⁶⁻⁸ (**Fig. P-1**). TFIID is a multisubunit complex comprising the TATA box-binding protein (TBP) and multiple TBP-associated factors (TAFs)⁹⁻¹¹. TAFs have been identified biochemically as stably associating components with TBP and were originally named according to their molecular weight in electrophoresis¹². A number of genetic and biochemical approaches have revealed that TAF4 (also known as TAF_{II}130 or TAF_{II}135 in humans, and TAF_{II}110 in *Drosophila*) interacts with cellular transcriptional activators such as Sp1 and the cAMP-response element-binding protein (CREB)¹³⁻¹⁵. These interactions are considered to participate in the recruitment of the pre-initiation complex to the promoter, thereby increasing transcriptional levels.

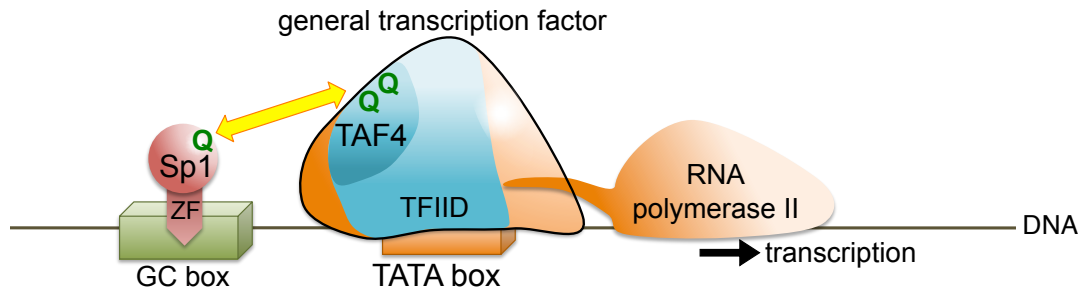


Figure P-1. Schematic diagram of transcription initiation. Sp1 and GC-box are colored in red and green, respectively. TAF4 and TFIID are colored in light blue, and the general transcription factor complex, TATA-box and RNA polymerase II are colored in orange. Gray line represents DNA and “Q” colored in green shows Q-domains. Zinc fingers (ZF) in Sp1 binds to GC-box on DNA.

In order to initiate transcription, the complex of general transcription factors recruits RNA polymerase II to the transcription initiation site. TFIID bound to TATA-box on DNA is a component of the general transcription factor complex. TFIID is also a multisubunit complex composed of TBP and several TAFs, including TAF4. Sp1 and TAF4 are reported to interact each other through their Q-domains.

Although the homologues of TAF4 in humans (hTAF_{II}130/135) and that in *Drosophila* (dTAF_{II}110) show limited sequence similarities^{16,17}, they share several functional and structural properties¹³. Previous studies revealed that *Drosophila* dTAF_{II}110 directly interacts with the human transcriptional activator Sp1 and enhances the transcriptional levels of the genes located downstream^{13,18}. A sequence analysis showed that both proteins contain four glutamine-rich domains (Q-domains) (Q1, Q2, Q3 and Q4), as well as two highly consensus regions, CI and CII with similarities of 68% and 72%, respectively¹⁵ (**Fig. P-2A**). An *in vitro* binding analysis and yeast two-hybrid assays indicated the involvement of one or more of the four Q-domains, Q1-Q4, in interactions with transcriptional activators. The central region of human TAF4 (written in “TAF4N/C” in this study described as **Fig. P-2A**) was found to be sufficient for interactions with the activation domain of human Sp1¹⁷.

The promoter-specific transcription factor Sp1 is expressed ubiquitously and plays a primary role in regulating the expression of many genes^{19–21}. Genetic and biochemical

studies on Sp1 have identified two potent activating regions, termed QA and QB, both of which are characterized by an abundance of glutamine residues^{22,23} (**Fig. P-2B**). Sp1 also possesses three C₂H₂-type zinc fingers at the C-terminal (**Fig. P-2B**), which interact with GC-boxes on the target gene located distally from the transcriptional initiation site²⁴. The interaction between Sp1, which recognizes and binds the upstream GC-box, and TAF4 in the general transcriptional factor via each Q-domain is considered to be important for the recruitment of RNA polymerase II to the transcription initiation site as well as the activation of transcription (**Fig. P-1**).

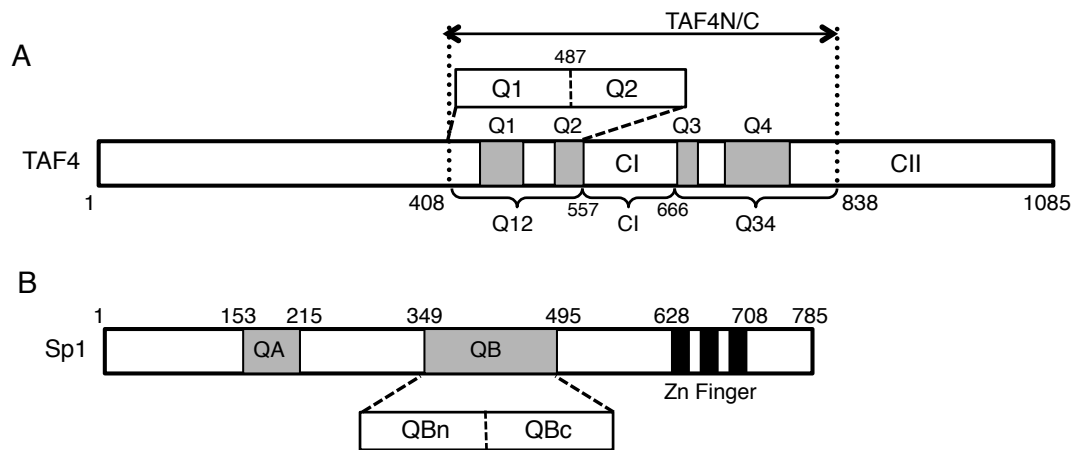


Figure P-2. (A) Schematic drawing of the primary structure of human TAF4 (hTAF_{II}130). Four Q-domains present in the molecule, Q1, Q2, Q3 and Q4, are indicated in gray. The positions of the highly conserved regions CI between Q2 and Q3, and CII at the C-terminus are also indicated. The region of “TAF4N/C”, which used in this study as containing all Q-domains, is indicated by a two-direction arrow. The region of fragment proteins of TAF4N/C, Q12, CI and Q34 also indicated (see chapter 1 and 2). TAF4-Q12 is also divided into two fragments, TAF4-Q1 and TAF4-Q2 (see chapter 2). (B) Schematic representation of the primary sequence of the transcription factor Sp1. Two Q-domains, Sp1-QA and Sp1-QB, and three zinc finger domains are indicated in gray and black, respectively. Sp1-QB is divided into two fragments, Sp1-QBn and Sp1-QBc (see chapter 1 and 2).

In chapter 1, the author revealed that the each Q-domain in Sp1 and TAF4 is intrinsically disordered protein (IDP) by the experiments using high-resolution NMR, small angle X-ray scattering (SAXS) and circular dichroism (CD) spectroscopy. IDPs have neither regular secondary nor tertiary structures under physiological conditions²⁵. An increasing number of IDPs has recently been found in many organisms, particularly higher eukaryotic organisms.

While an IDP is largely unstructured by itself, marked conformational changes often occur upon binding to an interacting partner in general, which is known as the “coupled-folding and binding mechanism”. High-resolution multidimensional NMR provides the chemical shift values of ¹H, ¹⁵N and ¹³C. ¹H and ¹⁵N chemical shifts of amide group in each amino acid constructing that protein reflect sensitively the environment around that amide group. On the other hand, ¹³C chemical shift values of the main chain and side chain are mainly affected by the secondary structure of that amino acid residue. In chapter 2, the author elucidated the binding site, interaction mode and association constant in the interaction between Q-domains in Sp1 and TAF4 by using multidimensional NMR. The author also measured association constant of the interaction by SPR experiments. Both methods showed that the association constant was on the order of 10⁴ M⁻¹, which suggested that the interaction was relatively weak.

In order to elucidate mechanisms of molecular recognition of Sp1 by TAF4 through their Q-domains, the author analyzed the interaction between several mutant proteins of Sp1 and TAF4-Q12 in chapter 3. By analyzing the ¹H-¹⁵N HSQC spectra, {¹H}-¹⁵N NOE values and ¹³C chemical shift changes in detail, the author confirmed that a clustering of hydrophobic residues was important for the interaction. On the basis of these observations, the author proposed a model, in which wild type Sp1-QBc is in fast equilibrium between two states. One can bind to TAF4-Q12 and has low mobility, and the other cannot bind and has high mobility, which can describe the results obtained. Moreover, the author also assumed that the wild type Sp1-QBc had allosteric effect in the presence of TAF4-Q12.

Chapter 1

NMR analysis of transcriptional activation domains in Sp1 and TAF4

1-1. Introduction

The expression of eukaryotic genes is precisely controlled by interactions between general transcriptional factors and promoter-specific transcriptional activators. Interactions between the fourth element of TATA-box binding protein-associated factor (TAF4), an essential subunit of the general transcription factor TFIID, and site-specific transcriptional activators, such as Sp1, are considered to be important for regulating the expression levels of genes of interest¹³. Early biochemical and genetic studies indicated that the glutamine-rich domains (Q-domains) in Sp1 are responsible for the interaction with Q-domains in TAF4¹⁷ (**Fig. P-1**).

Q-domains are characterized by a high degree (25% or more) of glutamine residues in the primary sequence. Moreover, they have been detected in many transcription-activating proteins, and are considered to represent one of the common motifs involved in molecular interactions^{26,27}. The two Q-domains in Sp1 (Sp1-QA and -QB) (**Fig. P-2B**) and four in TAF4 (TAF4-Q1, -Q2, -Q3 and -Q4) (**Fig. P-2A**) are also considered to be important for the interaction between these proteins²⁸. Sp1 also contains three zinc fingers of C₂H₂-type at N-terminal²⁹ and TAF4 also contains two highly conserved regions (CI and CII)^{30,31}, as shown in **Fig. P-2A**.

Although previous genetic and biochemical studies showed that Sp1 and TAF4 interacted with each other via Q-domains *in vitro* and *in vivo*, the molecular structures of these proteins remain unclear, and quantitative analyses on this interaction are limited. In this chapter, the author investigated the molecular architectures of the Q-domains in Sp1 and TAF4 using high-resolution NMR spectroscopy, CD and small angle X-ray scattering (SAXS) experiments. The author found that the isolated Q-domains of Sp1 and TAF4 were classified as intrinsically disordered proteins (IDPs), which have neither regular secondary nor tertiary structures under physiological conditions²⁵.

1-2. Materials and Methods

Materials

¹⁵N-Ammonium chloride and deuterium oxide (D₂O) were purchased from SI Science (Saitama, Japan). Other reagents were purchased from Nacalai Tesque (Kyoto, Japan).

Protein expression and purification

There are several GenBank entries for the sequence of human Sp1, and the author followed the numbering of amino acid residues for “isoform a” composed of 785 residues (NCBI Reference Sequence: NP_612482.2).

The gene of human TAF4 is also found as several entries in the database. Previous biochemical studies have referred to the sequence reported by Tanese *et al.*¹⁵, which corresponds to a protein of 947 amino acid residues (GenBank entry of U75308). By using this sequence, the central region (residues from 270 to 700) of human TAF4 (hTAF_{II}130) was found to be sufficient for interactions with the activation domain of human Sp1¹⁷. The author herein prepared the fragment protein corresponding to the same region (from 270 to 700) of human TAF4, and referred to it as TAF4N/C. However, the author used a different entry for the sequence encoding a protein with 1085 residues (NCBI Reference Sequence: NP_003176.2), and thus, the numbering of amino acid residues differs from those in previous studies (from 408 to 838 in this study). The chemically synthesized DNA encoding the central region of TAF4 [408-838], in which codon usage was optimized for *E. coli* B-strain, was purchased from Life Technologies (CA, USA).

The expression plasmids for Sp1-QA (Sp1 [153-215]) was constructed as fusion proteins with glutathione S-transferase (GST) at the N-terminal by inserting the gene of interest into the pGEX-1N (Amrad, Hawthorn, Australia) or modified pGEX-3X vector. The protein contained factor Xa cleavage site (Xa-site) followed by the three-residue N-terminal extension and C-terminal extension of a FLAG-octapeptide-tag followed by a hexahistidine-tag (His6-tag). The pGEX-Sp1-QA vector obtained was introduced into the *E. coli* strain BL21(DE3)pLysS.

The expression plasmids for Sp1-QB (Sp1 [349-495]), TAF4N/C, and TAF4-Q12 (TAF4 [408-557]) were constructed as fusion protein with human ubiquitin (Ub)

contained N-terminal extension of a His6-tag. The encoding region for TAF4N/C, TAF4-Q12 and TAF4-Q34 were amplified by PCR to introduce recognition sites for Bsp TI at the 5'-end and Eco RI at the 3'-end. The each amplified DNA fragment was inserted into the pET-His6-Ub expression vector between the Bsp TI and Eco RI sites. The pET-His6-Ub-Sp1-QB, pET-His6-Ub-TAF4N/C and pET-His6-Ub-TAF4-Q12 vectors obtained were introduced into the *E. coli* strain Rosetta2(DE3)pLysS purchased from Novagen (Darmstadt, Germany).

The expression plasmids for TAF4-CI (TAF4 [554-669]) and TAF4-Q34 (TAF4 [666-838]) were constructed with a His6-tag followed by Xa-site at the N-terminus. The encoding regions for TAF4-CI and TAF4-Q34 were amplified by PCR to introduce the recognition sites for Bsp TI at the 5'-end and Eco RI at the 3'-end. The amplified DNA fragment for each was inserted into the pET28-His6-Xa vector between the Bsp TI and Eco RI sites. The pET28-His6-Xa-TAF4-CI and pET28-His6-Xa-TAF4-Q34 vectors obtained were introduced into *E. coli* strain BL21-Gold(DE3).

Transformed bacteria were grown in Lucia broth medium. Uniformly ^{15}N -labeled proteins were prepared from cells grown in M9 minimal medium with 0.5 g/L [^{15}N]- NH_4Cl . Bacterial cells were grown at 37 °C, and isopropyl- β -D-thiogalactopyranoside (IPTG) was added to the medium when the absorbance at 600 nm reached 0.6. After the IPTG was added, cells were further incubated at 37 °C for 3 hr. Cells were harvested by centrifugation at 5,000 g for 15 min at 4 °C.

The collected bacterial cells were suspended in buffer A (300 mM NaCl, 50 mM sodium dihydrogen phosphate, 50 mM di-sodium hydrogen phosphate, pH 8.0) containing 1 mM phenylmethanesulfonyl fluoride, and disrupted by sonication with intermittent pulses for 1 min (pulse of 0.5 s, interval of 0.5 s, output level of 7) three times using an ultrasonic disruptor equipped with a TP-012 standard tip (UD-201, Tomy, Tokyo, Japan). The cell debris was removed by centrifugation at 10,000 g for 60 min at 4 °C, and the cell extract was loaded onto Ni-NTA resin. The column was washed with buffer A containing 20 mM imidazole, and bound protein was eluted with buffer A containing 500 mM imidazole.

In the case of Sp1-QA, the eluted protein was dialyzed against buffer B (140 mM NaCl, 20 mM sodium dihydrogen phosphate, 20 mM di-sodium hydrogen phosphate, pH 7.3), and digested by a protease factor Xa (Novagen) to cleave the peptide bond between the GST-tag and Sp1-QA. The protein was incubated for 16 hr at 20 °C at a

ratio of 1 unit of enzyme per 400 µg of substrate protein. The digested protein mixture was loaded onto Glutathione Sepharose 4B (GE Healthcare UK, Ltd, Buckinghamshire, England), and flow-through fractions were collected. After the GST-tag moiety was removed, the Sp1-QA was further purified by size-exclusion chromatography using a column of Superdex 75 10/300 GL (GE Healthcare) equilibrated with buffer B. The purity of the protein was confirmed to be >95% by size-exclusion chromatography and SDS-PAGE. The purified proteins were collected and concentrated with Amicon Ultra (Millipore, MA, USA), and stored at 4 °C.

In the case of Sp1-QB, TAF4N/C and TAF4-Q12, the eluted protein was dialyzed against buffer A, and digested the yeast ubiquitin carboxyl-terminal hydrolase 1 (YUH-1)^{32,33}. The enzyme was added to the dialyzed substrate protein at a ratio of 1 mol of enzyme per 0.5 mol of TAF4N/C and per 2 mol of Sp1-QB and TAF4-Q12, respectively. The protein mixture containing the enzyme and 1 mM DTT was incubated at 37 °C for 4 hr. The digested protein was added with solid ammonium sulfate up to 18% saturation, and the precipitate was collected by centrifugation at 25,000 g for 60 min. Precipitated protein was dissolved in buffer A and dialyzed against buffer A. The protein was purified by Ni-NTA and then further purified by size-exclusion chromatography in the same way as Sp1-QA.

In the case of TAF4-CI and TAF4-Q34, the eluted protein was dialyzed against buffer B, and digested by a protease factor Xa to cleave the peptide bond between the His6-tag and TAF4-CI or TAF4-Q34. The protein was incubated for 16 hr at 20 °C at a ratio of 1 unit of enzyme per 200 µg of substrate protein. The digested protein mixture was added with solid ammonium sulfate up to 30% saturation for TAF4-CI and 18% for TAF4-Q34, and the precipitate was collected by centrifugation at 25,000 g for 60 min. Precipitated protein was dissolved in buffer A and dialyzed against buffer A. The protein was purified by Ni-NTA and then further purified by size-exclusion chromatography in the same way as Sp1-QA.

NMR experiments

NMR spectra were measured on a Bruker DMX600 spectrometer at 4 °C equipped with a triple-axis gradient and triple-resonance TXI probe. The sample concentration was 50 µM ¹⁵N-labeled protein in buffer B containing 10% D₂O. Spectra were processed with nmrPipe and analyzed with nmrDraw³⁴.

CD spectroscopy

CD spectra were measured on a Jasco J-820 spectropolarimeter. A standard quartz cell with an assembling cell composed of a pair of quartz plates with a 0.1-mm path length, which had been calibrated by measuring the absorption of a protein solution with known concentration, was used to record spectra at each protein concentration of 50 μM in buffer B. The results obtained were expressed as apparent ellipticity with the unit of millidegrees (mdeg).

Small-Angle X-ray Scattering (SAXS)

SAXS measurements were performed with the spectrometer installed at BL-10C of Photon Factory (PF), a synchrotron radiation facility of Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK). The scattering intensity in a q -range between 0.02 and 0.2 $\text{\AA}^{-1}$ was recorded with PILATUS 2M detector. The q -range is estimated by the diffraction from the standard sample, silver behenate. Here, ($q = 4\pi\sin\theta/\lambda$, where 2θ and λ ($= 0.1488$ nm) is the scattering angle and the X-ray wavelength, respectively. Measured two-dimensional data were converted into one-dimensional one by circulation average and then the scattering profile, $I(q)$, of protein was obtained by the standard procedure of data correction for cell and background scatterings, transmission, and beam intensity.

During the X-ray exposure, the protein solutions were continuously flowed in a 1 mm thickness cell with the rate of 0.3 $\mu\text{L/s}$ to avoid the radiation damage by the X-ray beam. It is confirmed that all samples were free from the radiation damage in the exposure time of 120 sec.

1-3. Results

1-3-1. ^1H - ^{15}N HSQC spectra of Q-domains in Sp1 and TAF4

In order to reveal features of Q-domains in Sp1 and TAF4 in detail by using high-resolution NMR spectroscopy, the author first measured ^1H - ^{15}N HSQC spectra of Sp1-QA, Sp1-QB and TAF4N/C in buffer B (140 mM NaCl and 20 mM sodiumphosphate, pH 7.3) under physiological conditions, that is, at pH 7.3 and 37 °C. However, the signal-to-noise ratio was considerably low in all spectra (the spectra of Sp1-QA and Sp1-QB are not shown, and the spectrum of TAF4N/C is shown in **Fig. 1-1A**). In order to find the optimal conditions for the analysis, the author measured ^1H - ^{15}N HSQC spectra of Sp1-QA, Sp1-QB (data not shown) and TAF4N/C (**Fig. 1-1A-C**) at various temperatures with fixed pH, and found that signal intensity markedly increased at lower temperatures at pH 7.3.

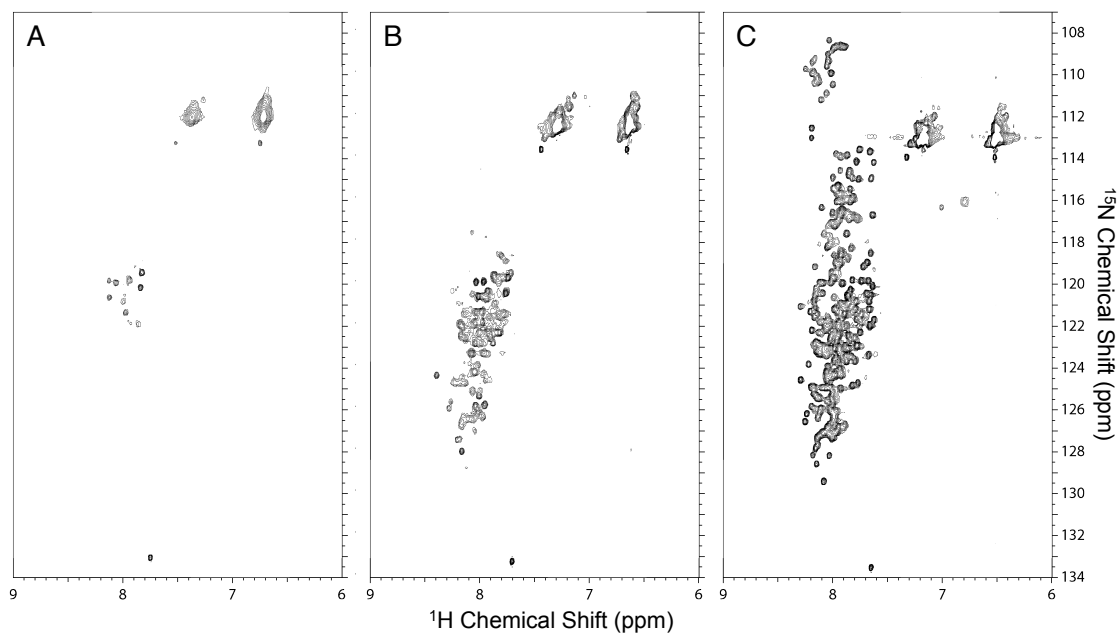


Figure 1-1. ^1H - ^{15}N HSQC spectra of TAF4N/C measured at pH 7.3 and 37 °C (A), 25 °C (B) and 4 °C (C), respectively. The concentration of ^{15}N -TAF4N/C was 50 μM . The number of peaks increased as the temperature decreased.

The author partly attributed these behaviors to the rapid exchange of amide hydrogen with solvent water at higher temperatures because the decrease in intensity was more significant for amide signals than aliphatic protons (**Fig. 1-2**). This was reasonable if rigid hydrogen bonds were lacking in these proteins. Therefore, the author performed NMR experiments under nearly physiological condition at pH 7.3 and 4 °C in chapter 1. The author confirmed that the difference in temperature (37 °C and 4 °C) did not affect the structure of these proteins using CD spectroscopy as described in detail in the next section **1-3-2**.

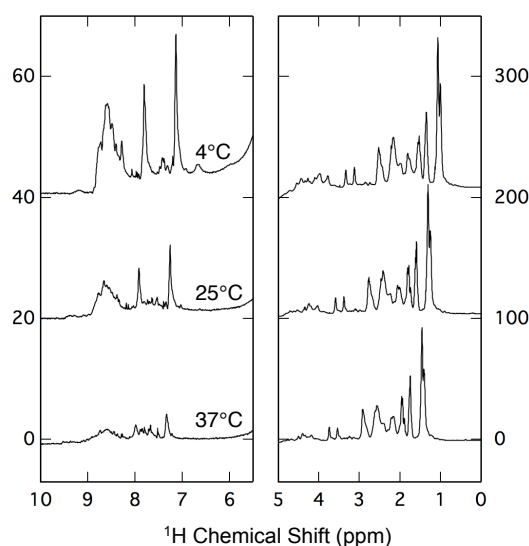


Figure 1-2. 1D ^1H NMR spectra of TAF4N/C recorded at 4 °C, 25 °C and 37 °C. The concentration of ^{15}N -TAF4N/C was 50 μM . Left and right panels correspond to the region of amide and aliphatic protons, respectively.

In order to elucidate the structural features of the Q-domains of Sp1 (QA and QB) and TAF4N/C expressed in *E. coli*, the author measured the ^1H - ^{15}N HSQC spectra under nearly physiological conditions at pH 7.3 and 4 °C. All resonance peaks were poorly dispersed along the ^1H chemical shift axis and appeared within the range of 8.5 and 7.5 ppm, indicating that the protein had neither secondary nor tertiary structures that were stabilized by regular hydrogen bonds and the rigid packing of side chains (**Fig. 1-3**). This result suggests that all Sp1-QA, Sp1-QB and TAF4N/C are intrinsically disordered under the conditions examined.

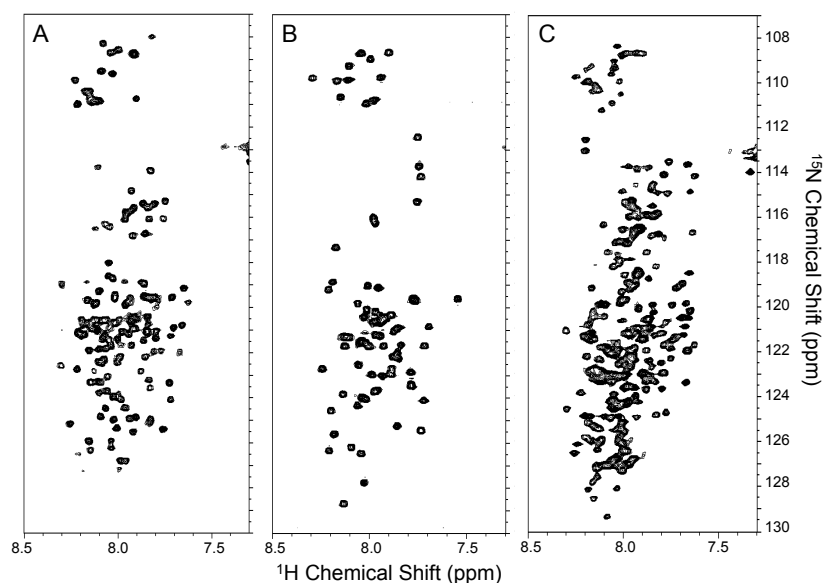


Figure 1-3. ^1H - ^{15}N HSQC of Sp1-QA (A), Sp1-QB (B) and TAF4N/C (C) measured at 4 °C and pH 7.3. Panel C is the same as **Fig.1-1C** except for the display range. The concentrations of the proteins were 50 μM .

In order to confirm that these proteins are indeed intrinsically disordered, the author prepared three fragment proteins by dividing the full-length TAF4N/C into three parts (TAF4-Q12, TAF4-CI and TAF4-Q34 shown in **Fig. P-2A**) and measured ^1H - ^{15}N HSQC spectra of these fragments under the same conditions. The superposition of ^1H - ^{15}N HSQC spectra obtained by three fragment proteins was almost identical to the spectrum of whole TAF4N/C (**Fig. 1-4**). This result indicates the absence of any significant interaction between the fragment proteins, the “dividability” of the whole protein into segments, and the “reproducibility” of the characteristics of the whole protein, which was consistent with TAF4N/C being largely disordered under the conditions examined.

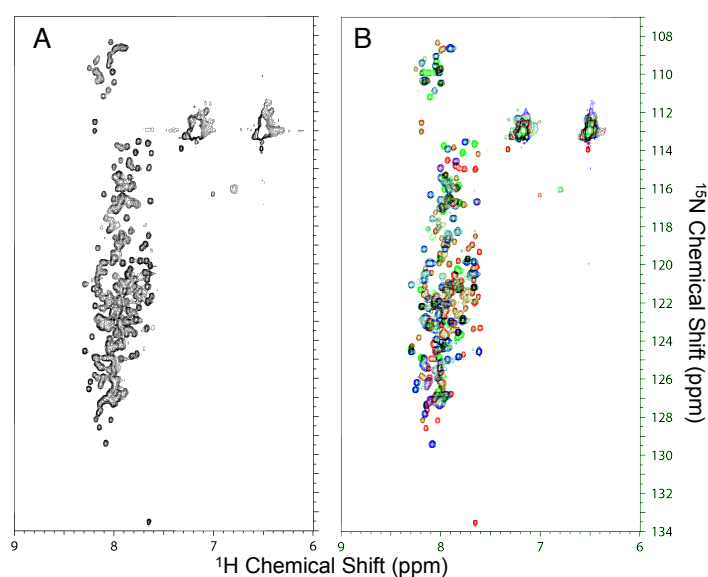


Figure 1-4. (A) ^1H - ^{15}N HSQC spectrum of ^{15}N -TAF4N/C recorded at pH 7.3 and 4 °C. (B) Overlay of the ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q12 (red), ^{15}N -TAF4-CI (green), and ^{15}N -TAF4-Q34 (blue) recorded separately at pH 7.3 and 4 °C. The concentrations of all the proteins were 50 μM . Panel A is the same as **Fig. 1-1C**, which is shown for comparison to panel B.

1-3-2. CD spectra of Q-domains in Sp1 and TAF4

In order to confirm the result of the NMR spectra, in which all of Q-domains in Sp1 and TAF4 were IDPs, the author examined the secondary structures of Sp1-QA, Sp1-QB and TAF4N/C by CD spectra (**Fig. 1-5A-C**). Also, to investigate that the temperature did not affect the structure, CD spectra were measured under physiological condition, at pH 7.3 and 37 °C. Whereas the signal intensities of NMR spectra are markedly affected by the experimental conditions such as temperature and pH value, those of CD spectra are not affected by these conditions. Therefore it is possible to obtain the spectra at any temperature and pH conditions. The CD spectra of both Sp1-QA and Sp1-QB showed a deep minimum at slightly less than 200 nm, suggesting that these proteins were largely disordered. On the other hand, the spectrum of TAF4N/C showed a large minimum around 208 nm as well as a small but significant shoulder around 222nm, suggesting that the protein is not fully but partially disordered.

The author also measured CD spectra of TAF-Q12, TAF4-CI and TAF-Q34 separately to elucidate structural features of TAF4N/C further as well as NMR spectra.

The both spectra of TAF4-Q12 and TAF4-Q34 represented to be disordered (red and blue lines in **Fig. 1-5D**, respectively). On the other hand, the CD spectrum of TAF4-CI showed large double minima at 222 and 208 nm (black line in **Fig. 1-5D**), suggesting that it adopts a predominantly α -helical structure. This was consistent with the X-ray crystallographic study, which indicated that the protein consisted of a five-helix “wedge” structure with a large hydrophobic groove³⁰. The CD measurements of Q-domains in Sp1 and TAF4 under physiological condition showed that these Q-domains were IDPs. This result was the same as the results of NMR experiments measured under nearly physiological condition, at 4 °C.

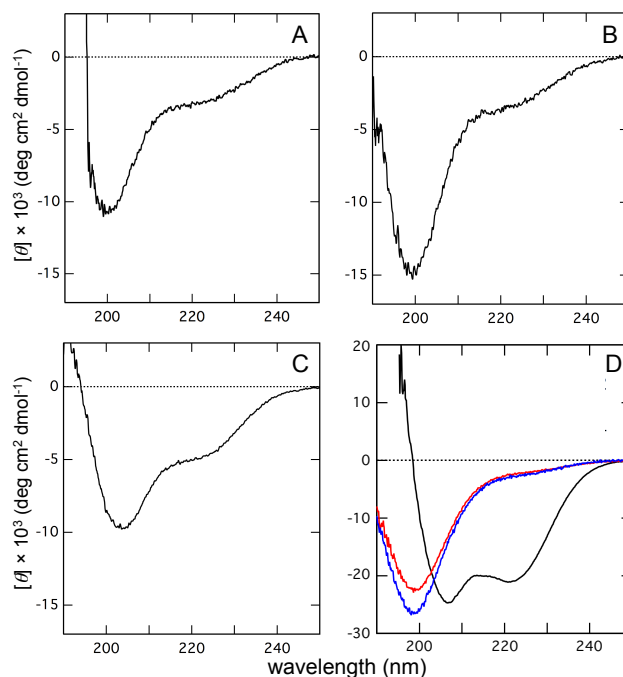


Figure 1-5. (A-C) Far-UV-CD spectra of Sp1-QA (A), Sp1-QB (B) and TAF4N/C (C). (D) Far UV-CD spectra of TAF4-Q12 (red), TAF4-CI (black), and TAF4-Q34 (blue). The concentrations of the proteins were 50 μ M.

1-3-3. SAXS analysis of Q-domains in Sp1 and TAF4

In order to further elucidate the structures of Sp1-QA, Sp1-QB and TAF4N/C at pH 7.3, the author also performed SAXS experiments that provide overall conformational properties of protein molecules. The author first plotted the logarithm of scattering intensity ($I(q)$) against the square of magnitude of scattering vector (q^2) (**Fig. 6A**). It is known that in the very narrow q -region ($< 1.3/R_g$), the linear dependency of $I(q)$ against q^2 (Guinier plot) provides the radius of gyration (R_g) of the solute. From the Guinier plot shown in **Fig. 6A**, the R_g values for Sp1-QA, Sp1-QB and TAF4N/C were estimated to be 25.6 ± 1.5 , 36.2 ± 0.3 and 49.0 ± 7.4 Å, respectively. In the case of Sp1-QA and Sp1-QB, these experimental values were in good agreement with those expected for the urea or guanidine hydrochloride denatured states of ~ 30 and ~ 40 Å, which were estimated for proteins with 92 and 172 amino acid residues³⁵. On the other hand, the R_g value of TAF4N/C was significantly smaller than that expected for the urea or guanidine hydrochloride denatured state of protein with 431 amino acid residues (~ 70 Å)³⁵. These results suggest that both Sp1-QA and Sp1-QB were largely disordered under the physiological conditions, that is, these proteins were typical IDPs, which was also consistent with that by the NMR experiments (**1-3-1**). On the other hand, TAF4N/C was shown to have bipartite aspects, that is, two largely disordered regions (TAF4-Q12 and TAF4-Q34) are connected by predominantly α -helical region (TAF4-CI), which was consistent with the CD spectra of the fragment proteins of TAF4N/C (**Fig. 1-5D**).

The author then expressed the scattering profile of these proteins by the Kratky plot, which provides the information on the overall shape of the protein (**Fig. 6B**). The Kratky plot for the typical native conformation with a globular shape exhibits a distinct peak, the position of which is dependent on R_g , while the plot for a molecule with random-coil like structure gives a plateau and then raises monotonically³⁶. None of the plot for Sp1-QA, Sp1-QB and TAF4N/C showed any clear peaks, suggesting that these proteins are largely disordered, which was in accord with the conclusion from NMR and CD spectra.

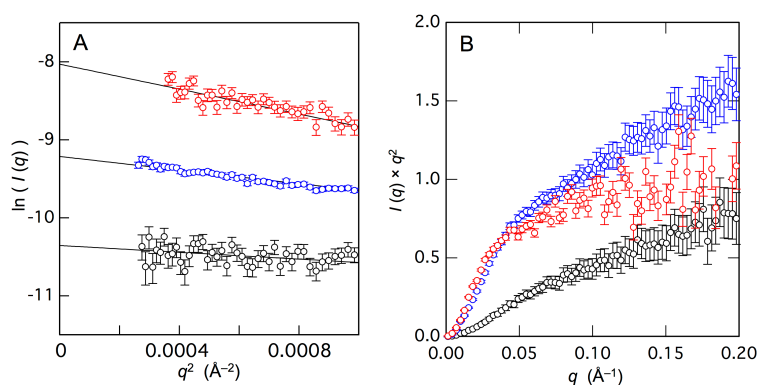


Figure 1-6. (A) The Guinier plots of SAXS of Sp1-QA (black), Sp1-QB (blue), and TAF4N/C (red) shown in open circles. The solid lines indicate the results of linear curve fitting for the data within the Guinier-region ($q \times R_g < 1.3$). (B) The Kratky plots of Sp1-QA (black), Sp1-QB (blue), and TAF4N/C (red). In panel A and B, the concentrations of all the proteins were 2 mg/mL.

1-4. Discussion

1-4-1. Structural features of Q-domains in Sp1 and TAF4

Q-domains, which are often found in transcription factors, are characterized by a high degree (25% or more) of glutamine residues in the primary sequence. The NMR experiments showed that Q-domains in Sp1, QA and QB, and Q-domains in TAF4 were intrinsically disordered under nearly physiological conditions at 4 °C and pH 7.3. The reason why the author performed NMR characterizations of these domains at 4 °C is because the peak intensities in the ¹H-¹⁵N HSQC spectra were markedly decreased at higher temperatures (**Fig. 1-1**). In order to confirm no effect of temperature on the structure of Q-domains, CD spectra were measured instead of NMR spectra. CD spectra measured under the physiological conditions at 37 °C and pH 7.3 showed that Q-domains in Sp1 and TAF4 were in random-coil state as well as NMR spectra at 4 °C and pH 7.3 (**Fig.1-5A-C**). Since the result from CD spectra under the physiological condition was identical with that from NMR spectra under the nearly physiological condition, while CD spectrum shows just overall structure of a protein, the author considers that the “nearly physiological condition” at pH 7.3 and 4 °C is virtually the same as “physiological condition” at pH 7.3 and 37 °C, and that these Q-domains are IDPs.

The native structure of a globular protein is characterized by well-ordered secondary structures stabilized by hydrogen bonds, which are typically buried inside the protein molecule, otherwise they are destabilized by non-specific hydrogen bonding with the surrounding water molecules. The interior of a water-soluble protein molecule mostly consists of hydrophobic amino acids, which is referred to as a “hydrophobic core”, and its formation is involved in the specific and rigid packing of side chains of amino acids. Therefore, a certain number of hydrophobic residues are necessary for the formation of a typical “native” structure³⁷. The Q-domains in Sp1 and TAF4 may not have had a sufficient number of hydrophobic residues because of the large number of glutamine residues.

1-4-2. TAF4N/C and Sp1-QB may be divided into several fragment proteins

The author confirmed the “dividability” of the whole protein into segments and the “reproducibility” of the characteristics of the whole protein by these fragmented

proteins using ^1H - ^{15}N HSQC spectra. The superposition of the ^1H - ^{15}N HSQC spectra of the three fragment proteins, TAF4-Q12, TAF4-CI, and TAF4-Q34 (see **Fig. P-2A**), was indistinguishable from the spectrum of TAF4N/C (**Fig. 1-4**). Furthermore, the author and colleagues has elucidated previously that Sp1-QB was also successfully divided into two parts, Sp1-QBn and Sp1-QBc (see **Fig. P-2B**)³⁸. These results were not surprising because many proteins are built of structural “domains” that are formed by a group of residues located spatially close to each other. These structural domains are known to form a compact three-dimensional structure, which often may be independently stable and folded from other parts of the molecule. These disordered protein fragments may be divided without the loss of structure and function. Similarly to the results of NMR and CD spectra, these results also indicated that the Q-domains in TAF4 and Sp1 were classified as typical IDPs without any rigid secondary or tertiary structures.

1-4-3. TAF4N/C was partially structured, but largely disordered

Based on the ^1H - ^{15}N HSQC spectrum of TAF4N/C and the fragmentation experiments, the author concluded that TAF4N/C was classified as an IDP. On the other hand, other experimental observations suggest the presence of partially structured regions in TAF4N/C. One of them was the CD spectrum of TAF4N/C, which showed a small, but significant shoulder around 222 nm, indicating the presence of some amount of α -helical structure (**Fig. 1-5C**). Another argument was based on the R_g value, which was estimated by a Guiner plot of SAXS measurements (**Fig. 1-6A**). The R_g value calculated by experimental data ($49.0 \pm 7.4 \text{ \AA}$) was significantly smaller than that expected for the fully denatured state of the protein with 431 amino acid residues estimated by a collection of urea or guanidine hydrochloride denatured proteins ($\sim 70 \text{ \AA}$)³⁵.

These apparent discrepancies may be rationalized by measuring the CD spectra of three fragment proteins separately (**Fig. 1-5D**). Two fragment proteins that are dominated by the Q-rich sequences, TAF4-Q12 and TAF4-Q34, showed deep minima at less than 200 nm, suggesting that these two regions are largely disordered. On the other hand, the fragment protein corresponding to the highly conserved region, TAF4-CI, showed deep minima at 222 and 208 nm, which are characteristics of proteins

with predominantly α -helical structure. This result was in good agreement with the X-ray crystallographic study, which indicated that the protein consists of a five-helix “wedge” structure³⁰. Therefore, the author considered the structure of TAF4N/C to be “bipartite”, consisting of two highly disordered segments of the Q-rich sequence connected by a highly conserved α -helical structure located in the center of the molecule.

Chapter 2

Identification of heteromolecular binding sites in transcription factors Sp1 and TAF4 using NMR spectroscopy

2-1. Introduction

Previous genetic and biochemical studies showed that Sp1 and TAF4 interacted with each other via their Q-domains. The interaction sites between Sp1 and TAF4 have been estimated by yeast two-hybrid assays using several deletion mutants³⁹. Note that this method evaluates the interaction between proteins by the relative expression level of the reporter gene such as β -galactosidase, and a quantitative analysis is difficult, since this method sometimes detects a pseudo-positive interaction.

A ^1H - ^{15}N HSQC spectrum is 2D-NMR spectrum, whose horizontal and vertical axes show ^1H and ^{15}N chemical shifts, respectively. The ^1H and ^{15}N chemical shifts are derived from amide groups of amino acids, which construct the ^{15}N -labeled protein. In this spectrum, one peak corresponds to one amino acid residue. ^1H - ^{15}N HSQC spectra are useful for detection of protein interactions, because unlabeled proteins are not visible in the spectrum. That is, by evaluating the change in the intensities and/or position of peaks, it is possible to reveal whether each ^{15}N -labeled amino acid constructing the ^{15}N -labeled protein is affected by the addition of unlabeled proteins or not. Moreover, ^1H and ^{15}N chemical shift values sensitively reflect changes in the environment surrounding the nuclear spin. Therefore, changes in the ^1H and ^{15}N chemical shift values upon the addition of the unlabeled protein directly reveal the interaction between the proteins, as well as a possible conformational change.

Most IDPs are known to undergo a large conformational change to form secondary structures when they interact with their partner biomolecules. This mechanism is known as “coupled folding and binding”^{40,41}. ^{13}C chemical shift analysis is useful for identifying the conformation changes of main chain, since ^{13}C chemical shifts reflect dihedral angles of main chain of proteins. The author elucidated that all Q-domains in Sp1 and TAF4 were IDPs in chapter 1. That is, the interaction between Q-domains in Sp1 and TAF4 might be classified into the interaction between IDPs.

In chapter 2, the author investigated the interaction region and mode between Q-domains in Sp1 and TAF4 in detail using high resolution NMR spectroscopy. As a result, the interaction regions in Sp1 and TAF4 were elucidated to present at C-terminal

half of QB domain in Sp1 and Q1 domain in TAF4, respectively. And surprisingly, this binding mode is not categorized into the well-accepted concept of the “coupled-binding and folding” mechanism. Moreover, from NMR titration and SPR experiments, the association constant was estimated to be on the order of 10^4 M^{-1} . The author and colleagues have revealed previously that Sp1-QB molecules interact with themselves, and that the self-association constant was on the order of 10^3 M^{-1} ³⁸. The interaction sites in Sp1-QB for the homo-oligomerization of Sp1 and the hetero-oligomerization between Sp1 and TAF4 were largely overlapped. These results suggested that the self-association of Sp1-QB domains and the heterogeneous interaction between Sp1-QB and TAF4N/C competed with each other.

2-2. Materials and Methods

Materials

¹⁵N-Ammonium chloride, ¹³C₆-D-glucose, deuterium oxide (D₂O), and sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) were purchased from SI Science. Other reagents were purchased from Nacalai Tesque.

Protein preparation

The expression and purification of Sp1-QA, Sp1-QB, TAF4N/C, TAF4-Q12, TAF4-CI and TAF4-Q34 were described in chapter 1.

The expression plasmid for Sp1-QBn (Sp1 [342-424]) was constructed as fusion proteins with GST at the N-terminal, which was similar to Sp1-QA. The expression and purification of Sp1-QBn was also performed similar to Sp1-QA.

The expression plasmids for Sp1-QBc (Sp1 [421-495]) and TAF4-Q1 (TAF4 [408-487]) were constructed as fusion protein with Ub contained N-terminal extension of a His6-tag, which were similar to Sp1-QB, TAF4N/C or TAF4-Q12. The expression and purification of Sp1-QBc and TAF4-Q1 was also performed similar to Sp1-QB, TAF4N/C or TAF4-Q12

The expression plasmid for TAF4-Q2 (TAF4 [486-557]) was constructed with a His6-tag followed by Xa-site at the N-terminus, which was similar to TAF4-CI or TAF4-Q34. The expression and purification of TAF4-Q2 was also performed similar to

TAF4-CI or TAF4-Q34.

For the preparation of uniformly ^{15}N , ^{13}C -labeled proteins, bacterial cells were grown in M9 minimal medium supplied with 0.5 g/L $[^{15}\text{N}]\text{-NH}_4\text{Cl}$ and 2 g/L.

All proteins were dialyzed against buffer B (140 mM NaCl, 20 mM Na_2HPO_4 , 20 mM NaH_2PO_4 , pH 7.3) or buffer C (20 mM sodium acetate, pH 5.0) depended on experiments.

Protein concentration measurements

Protein concentrations were measured by absorption at 277-280 nm on the assumption that the total extinction coefficients of proteins were the same as the linear combination of the extinction coefficients of tryptophan and tyrosine residues present in the proteins⁴². Since several fragment proteins (TAF4-Q1, TAF4-Q12 and TAF4-Q34) did not contain any tryptophan or tyrosine residues, their concentrations were measured as follows. First, tyrosine-containing mutant proteins were prepared. In the case of TAF4-Q1 and TAF4-Q12, a tyrosine residue was added at the N-terminus (called TAF4-YQ1 and TAF4-YQ12, respectively). In the case of TAF4-Q34, F828 located between Q3 and Q4 was replaced by tyrosine (named TAF4-Q34-F828Y). Next, the concentrations of the purified tyrosine-containing mutants TAF4-YQ1, TAF4-YQ12 and TAF4-Q34-F828Y were determined by the absorption at 277 nm. Finally, the concentrations of wild-type tyrosine-less proteins were estimated by comparing the relative peak intensity of 1D-NMR spectra to those recorded for tyrosine-containing mutants with known concentrations.

Surface plasmon resonance (SPR)

SPR measurements were performed using a ProteOn XPR36 system (Bio-Rad) at 15 °C (the lowest temperature that is able to be set by the instrument) with an HTG sensor chip equilibrated with buffer B as a running buffer at a flow rate of 100 $\mu\text{L}/\text{min}$. The HTG sensor chip was first activated with 10 mM NiSO_4 for 4 min. One hundred and twenty five microliters of the ligand protein (20 μM Sp1-QA and Sp1-QB with His6-tag at the C-terminus) was injected and immobilized on the HTG chip. The analyte protein (TAF4N/C) was then injected for 150 sec at a constant flow rate of 100 $\mu\text{L}/\text{min}$, and sensorgrams were recorded as the association phase. The concentration of analyte loaded was 2.0, 1.2 and 0.4 μM for Sp1-QA, and 2.0, 1.6, 0.8 and 0.4 μM for Sp1-QB.

The dissociation of the bound analyte followed subsequently for 300 sec at the same flow rate. Kinetic parameters were calculated using a global fitting analysis with the assumption of the 1:1 Langmuir binding model using equations [1] and [2] for the association and dissociation phases, respectively.

$$R = \frac{k_a CR_{\max}}{k_a C + k_d} [1 - \exp\{-(k_a C + k_d)t\}] \quad [1]$$

$$R = R_0 [1 - \exp\{-k_d t\}] \quad [2]$$

where t is time, R is the time-dependent response of the SPR signal, R_0 , and $R_{\max}$ are the initial (just after the dissociation phase) and the maximum (reaches at the infinite time in the association phase) values of the SPR signal, respectively, C is the concentration of the analyte, and k_a and k_d are the association and dissociation rate constants between the analyte and ligand, respectively.

Small-Angle X-ray Scattering (SAXS)

SAXS measurements were performed with the spectrometer installed at BL-10C of Photon Factory (PF) as described in chapter 1.

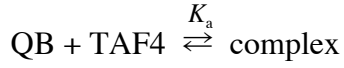
Circular dichroism (CD) spectroscopy

CD spectra were measured on a Jasco J-820 spectropolarimeter performed as described in chapter 1.

NMR experiments

NMR experiments were performed on a Bruker DMX600 spectrometer as described in chapter 1. Sequence-specific resonance assignments were obtained by analyzing the 3D-HNCACB, 3D-CBCA(CO)NH, 3D-HNCO, 3D-HN(CA)CO, and 3D-hNcocaNH spectra recorded at pH 5.0 and 37 °C⁴³⁻⁴⁷. The concentrations used were 100 μM ¹⁵N, ¹³C-labeled protein in buffer C containing 10% D₂O. The chemical shift value was referenced to DSS. Spectra were processed with nmrPipe⁴⁸, and analyzed with XIPP, a new version of PIPP⁴⁹, which was kindly gifted by Dr. Garrett.

In the titration experiment, the author assumed 1:1 binding between Sp1-QBc and TAF4-Q12 as



where QB, TAF, and complex represent a monomeric ^{15}N -Sp1-QBc, unlabeled TAF4-Q12, and the complex between them, respectively. This defines the dimeric association constant K_a as

$$K_a = \frac{[\text{complex}]}{[\text{QB}][\text{TAF}]}$$

By using K_a , and the total concentrations of ^{15}N -Sp1-QBc and unlabeled TAF4-Q12, $[\text{QB}]_0$ and $[\text{TAF}]_0$, and the saturated chemical shift perturbation, $\delta\Delta_{\text{max}}$, the chemical shift perturbation, $\delta\Delta$, of ^{15}N -Sp1-QBc was represented as,

$$\Delta\delta = \frac{\delta\Delta_{\text{max}} \{([\text{QB}]_0 + [\text{TAF}]_0 + 1/K_a) - \sqrt{([\text{QB}]_0 + [\text{TAF}]_0 + 1/K_a)^2 - 4[\text{QB}]_0[\text{TAF}]_0}\}}{2[\text{QB}]_0} \quad [3]$$

The observed differences in the chemical shift values were curve-fitted to equation [3] as the function of the total concentration of proteins ($[\text{QB}]_0$ and $[\text{TAF}]_0$) to estimate K_a value.

2-3. Results

2-3-1. Interaction between Q-domains of Sp1 and TAF4 elucidated by NMR spectroscopy

The intensity and position (chemical shift value) of each NMR peak sensitively reflects changes in the environment surrounding the nuclear spin. Therefore, the author expected a detailed analysis of NMR spectra to reveal the intermolecular interaction between the Q-domains in Sp1 and TAF4. The author used the same experimental conditions as in chapter 1, that is, at pH 7.3 and 4 °C (called as nearly physiological condition). In order to detect the interactions, the author initially measured the ^1H - ^{15}N HSQC spectra of Sp1-QA in the absence and presence of a three molar excess amount of unlabeled TAF4N/C. As a result, the spectra of Sp1-QA were not changed between with and without unlabeled TAF4N/C. This result suggested that Sp1-QA and TAF4N/C did not interact with each other, or that the affinity of the two proteins is too weak to be detected under the conditions examined (**Fig. 2-1A**).

The author performed a similar experiment in order to examine the interaction between ^{15}N -Sp1-QB and unlabeled TAF4N/C. In contrast to Sp1-QA, the intensity of several peaks of ^{15}N -Sp1-QB markedly decreased when an equimolar amount of unlabeled TAF4N/C was added (**Fig. 2-1B**). Severe line broadening may be resulted from a large molecular weight of the complex with unlabeled TAF4N/C (see below). However, the result clearly demonstrated that Sp1-QB and TAF4N/C interacted with each other. In order to identify the region of the interaction, peak intensities in the presence of unlabeled TAF4N/C relative to those recorded in its absence were plotted against the residue number of Sp1-QB (**Fig. 2-1C**). The results revealed that the residues, whose intensities were significantly decreased, were located in the region from the center to the C-terminal part of the molecule. This suggests that the observed decrease in the intensity was site-specific, and that these residues in Sp1-QB may represent the binding site for TAF4N/C.

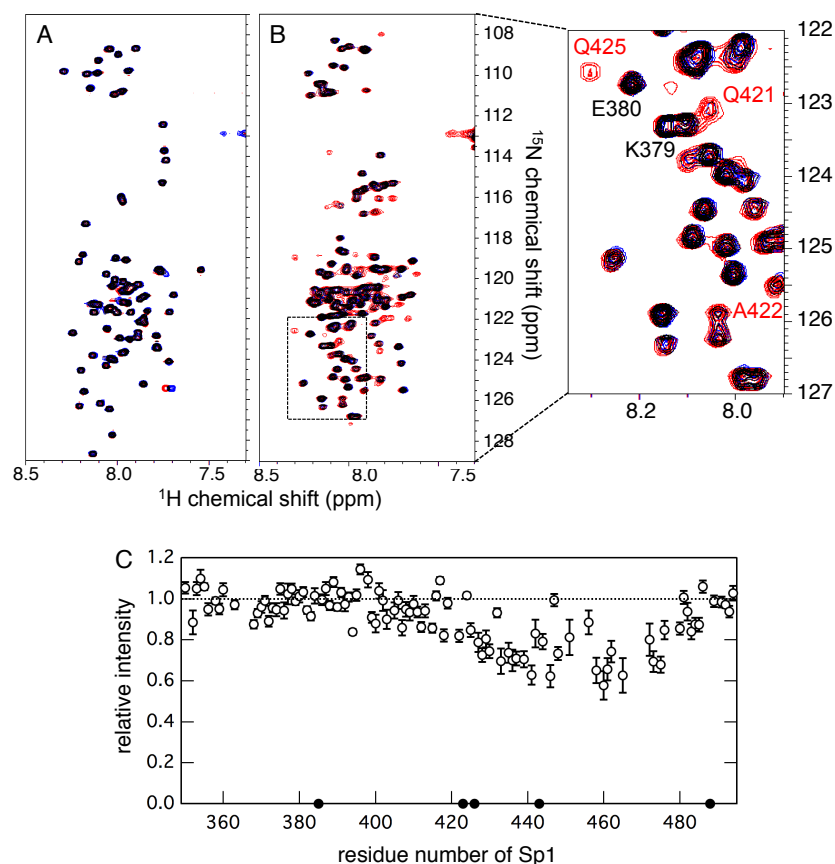


Figure 2-1. (A) Overlay of the ^1H - ^{15}N HSQC spectra of ^{15}N -labeled Sp1-QA in the absence (red) and presence (blue) of a three molar excess amount of unlabeled TAF4N/C measured at pH 7.3 and 4 °C. The concentrations of ^{15}N -Sp1-QA in both spectra were 50 μM . The peaks colored black indicate that both chemical shifts and the intensities of the peaks were not changed. (B) Overlay of the ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QB domains in the absence (red) and presence (blue) of an equimolar amount of unlabeled TAF4N/C measured at pH 7.3 and 4 °C. The concentrations of ^{15}N -Sp1-QB in both spectra were 50 μM . The region indicated by the dashed box in the spectrum is expanded in the right panel. The intensities of several peaks of ^{15}N -Sp1-QB markedly decreased with the addition of unlabeled TAF4N/C. (C) The relative peak intensity of ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QB plotted against the residue number of Sp1. Intensity in the presence of the same concentration of unlabeled TAF4N/C relative to that in its absence is shown. The closed circle shows residues that could not be analyzed due to signal overlap.

2-3-2. Surface Plasmon Resonance (SPR)

The author also measured SPR in order to obtain more direct evidence for the interaction and association constant for binding between Sp1-QB and TAF4N/C. An SPR analysis requires the immobilization of the “ligand” molecule on the sensor chip. In order to minimize the effects of immobilization on the structure and function of the ligand protein, the author utilized a His6-tag attached at the C-terminus of the Sp1-QB protein via a flexible linker sequence. Injections of various concentrations of TAF4N/C as the “analyte” resulted in marked increases in resonance, demonstrating the association between immobilized Sp1-QB and TAF4N/C (**Fig. 2-2A**). After the injection of TAF4N/C for 150 s, the same buffer B without the analyte was loaded for 300 s. Gradual decreases were observed in resonance, suggesting the slow dissociation of the bound analytes from the immobilized ligand protein.

All sensorgrams were globally fitted to equations [1] and [2] for the association and dissociation phases, respectively. The fit converged well, and single association and dissociation rate constants were estimated to be $4.40 \pm 0.08 \text{ M}^{-1} \text{ s}^{-1}$ and $(3.03 \pm 0.06) \times 10^{-4} \text{ s}^{-1}$, respectively. The equilibrium association constant between Sp1-QB and TAF4N/C was estimated to be $(1.45 \pm 0.04) \times 10^4 \text{ M}^{-1}$ from these association and dissociation rate constants. These results suggest a significant interaction between Sp1-QB and TAF4N/C. On the other hand, there were no detectable interactions between Sp1-QA and TAF4N/C by SPR experiments performed under the same conditions (**Fig. 2-2B**), which was consistent with the result of NMR experiments of ^{15}N -Sp1-QA and unlabeled TAF4N/C.

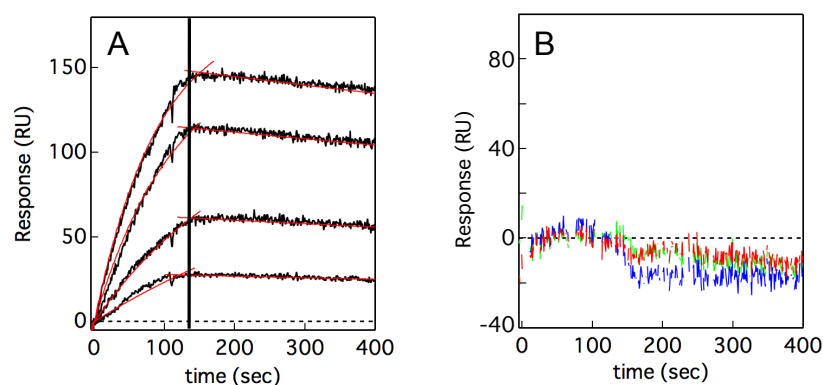


Figure 2-2. (A) Kinetics of the association between Sp1-QB and TAF4N/C, and their dissociation were measured by SPR. Sp1-QB was immobilized on an HTG chip via the His6-tag at the C-terminus. Black lines show sensorgrams recorded after the loading of TAF4N/C at various concentrations for 150 s, followed by the flow of its absence. Red lines indicate the results of global fitting of four sensorgrams by equations [1] and [2]. (B) The interaction between Sp1-QA and TAF4N/C was measured by the SPR. Sp1-QA was immobilized on an HTG chip via the His6-tag at the C-terminus. The lines represent the sensorgrams recorded after loading TAF4N/C for 150 s at concentrations of 2.0 (red), 1.2 (green) and 0.4 μM (blue), respectively, followed by the flow of its absence for 300 s.

2-3-3. Optimization of experimental conditions and NMR signal assignment

In order to elucidate the molecular mechanisms underlying the interaction between Sp1 and TAF4 in more detail, the author attempted to perform NMR experiments similar to section 2-3-1, but the opposite combination of labeling to identify the residues responsible for the interaction. That is, ^{15}N -TAF4N/C was titrated by unlabeled Sp1-QB. In the first step of the analysis, the author needed to assign the NMR signals of TAF4N/C, but its molecular weight was approximately 45 kDa, and the signal assignment of TAF4N/C is challenging. The author had found in chapter 1 that TAF4N/C was an IDP, and that the entire protein could be divided into three fragment proteins (Q12, CI, and Q34, described in **Fig. P-2A**) without significant perturbations to the structure. Therefore, the author assigned the NMR signal of TAF4N/C by taking advantage of analyzing several fragment proteins separately.

The author attempted to analyze a short fragment protein corresponding to the Q12

region of TAF4 under nearly physiological condition, at pH 7.3 and 4 °C (**Fig. 2-3A**). However, the analysis under these conditions was unsuccessful, because most NMR peaks disappeared in the ^1H - ^{15}N HSQC spectra and signal intensities were very weak in triple-resonance experiments such as 3D-HNCACB and 3D-hNcocaNH (data not shown). The author attributed the disappearance of peaks to the exchange of labile NH protons with water molecules at pH 7.3 and/or chemical exchange broadening by the transient formation of dimers or higher oligomers. After a comprehensive screening of experimental conditions, the author found that an increase in temperature and decrease in pH value markedly improved the quality of ^1H - ^{15}N HSQC spectra (**Fig. 2-3**).

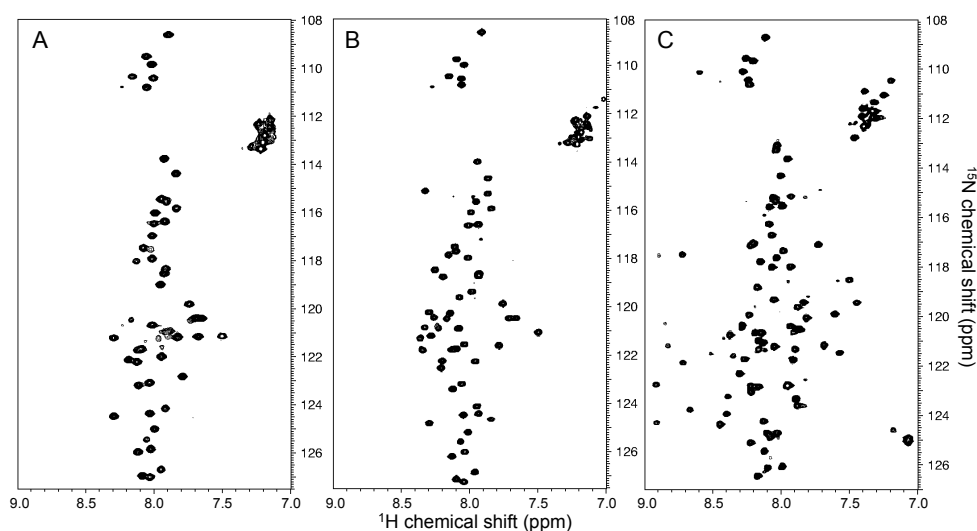


Figure 2-3. ^1H - ^{15}N HSQC spectra of ^{15}N -labeled TAF4-Q12 measured at pH 7.3 and 4 °C (A), pH 5.0 and 4 °C (B), and pH 5.0 and 37 °C (C). The concentration of TAF4-Q12 was 50 μM .

During the process to optimize the experimental conditions, the author found that pH 5.0 and 37 °C gave the most promising spectra, while an increase in pH by one unit (pH 6) significantly decreased the signal-to-noise ratio (**Fig. 2-4**). Therefore, the author selected these experimental conditions (pH 5.0 and 37 °C) although it is slightly more acidic than physiological conditions.

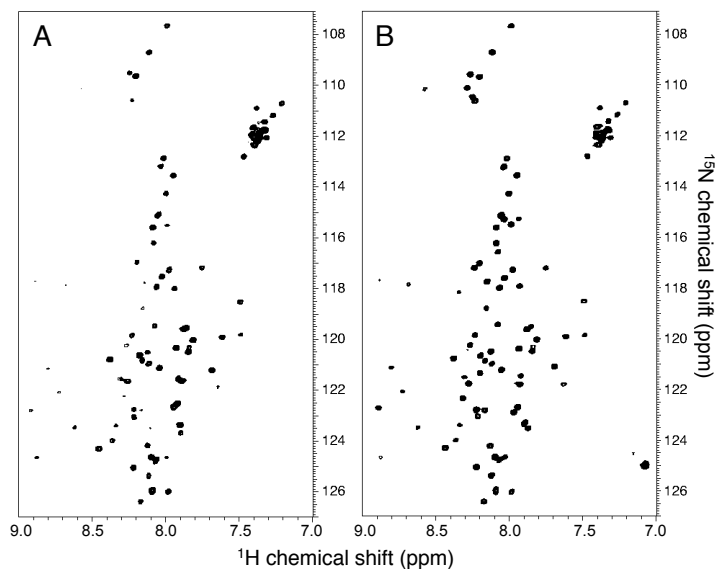


Figure 2-4. ^1H - ^{15}N HSQC spectra of ^{15}N -labeled TAF4-Q12 measured at pH 6.0 (A) and pH 5.0 (B). The concentration of TAF4-Q12 was 50 μM

In order to confirm that pH changes did not affect the structure of Q-domains, the CD spectra were recorded at pH 7.3 and 5.0 and these spectra were indistinguishable from each other (**Fig. 2-5A**). The author also performed SAXS experiments and analyzed the scattering profile to estimate the values of the radius of gyration (R_g) (**Fig. 2-5B**). The R_g of Sp1-QB at pH 7.3 and pH 5.0 were $36.2 \pm 0.3 \text{ \AA}$ and $37.2 \pm 0.5 \text{ \AA}$, respectively. The R_g of TAF4N/C at pH 7.3 and pH 5.0 were $49.0 \pm 7.4 \text{ \AA}$ and $49.7 \pm 0.6 \text{ \AA}$, respectively. The results of both CD and SAXS experiments elucidated that the overall structures of Sp1-QB and TAF4N/C were not significantly affected under the different pH conditions.

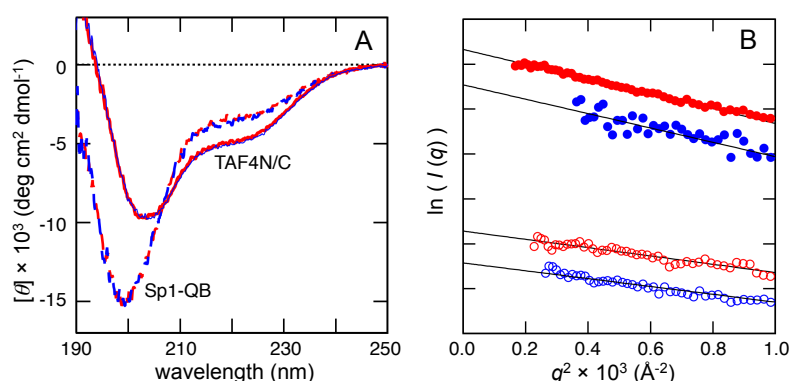


Figure 2-5. (A) Far UV-CD spectra of TAF4N/C (solid line) and Sp1-QB (dotted line) measured at pH 7.3 (blue) and pH 5.0 (red). The concentrations of both proteins were 50 μ M. Note that each spectrum recorded at different pH values is overlaid indistinguishably. (B) Guinier plots of the SAXS profiles of TAF4N/C (closed circle) and Sp1-QB (open circle). Each sample was measured at pH 5.0 (red) and pH 7.3 (blue). The concentrations of both proteins were 2 mg/mL. The lines indicate the results of linear curve fitting for data within the Guinier region ($q \leq R_g/1.3$). The R_g values of TAF4N/C and Sp1-QB are calculated from the slopes of these lines.

2-3-4. Identification of the Sp1-interaction site in the TAF4 molecule

The author measured the ^1H - ^{15}N HSQC spectra for TAF4-CI and TAF4-Q34 in the absence and presence of more than a two-fold amount of unlabeled Sp1-QB, however, no changes were observed in chemical shift values or peak intensities (**Fig. 2-6B,C**). Therefore, the author concluded that these regions did not contribute significantly to the interaction between TAF4 and Sp1. Note that only a limited number of peaks appeared in the NMR spectrum of TAF4-CI (approx. 60, while the expected number was 136; **Fig. 2-6B**). The author attributed this apparent discrepancy to chemical exchange broadening of TAF4-CI, possibly originating from the formation of a transient dimer or higher oligomeric structure under the experimental conditions used. The presence of a large hydrophobic groove in TAF4-CI may contribute to these possible intermolecular interactions³⁰.

The author then performed a similar experiment to examine the interaction between TAF4-Q12 and Sp1-QB. In contrast to the results obtained for TAF4-CI and TAF4-Q34, the ^1H - ^{15}N HSQC spectra of TAF4-Q12 measured in the absence and presence of an equimolar amount of Sp1-QB showed marked differences in the

chemical shift values of several residues, suggesting a specific interaction between TAF4-Q12 and Sp1-QB (**Fig. 2-6A**). Note that a careful examination of the spectrum of TAF4-Q12 revealed that several peaks appeared in a lower field region (with higher values for the chemical shift) in the ^1H chemical shift axis, suggesting that weak hydrogen bonds might be formed in these residues. Based on the apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{1H}})^2 + (\Delta\delta_{\text{15N}}/5)^2}$, plotted against residue number of TAF4, the chemical shift values of a stretch of residues (455–480) corresponding to the C-terminal half of the Q1 domain were markedly perturbed, suggesting a possible binding site for Sp1-QB (**Fig. 2-6D**).

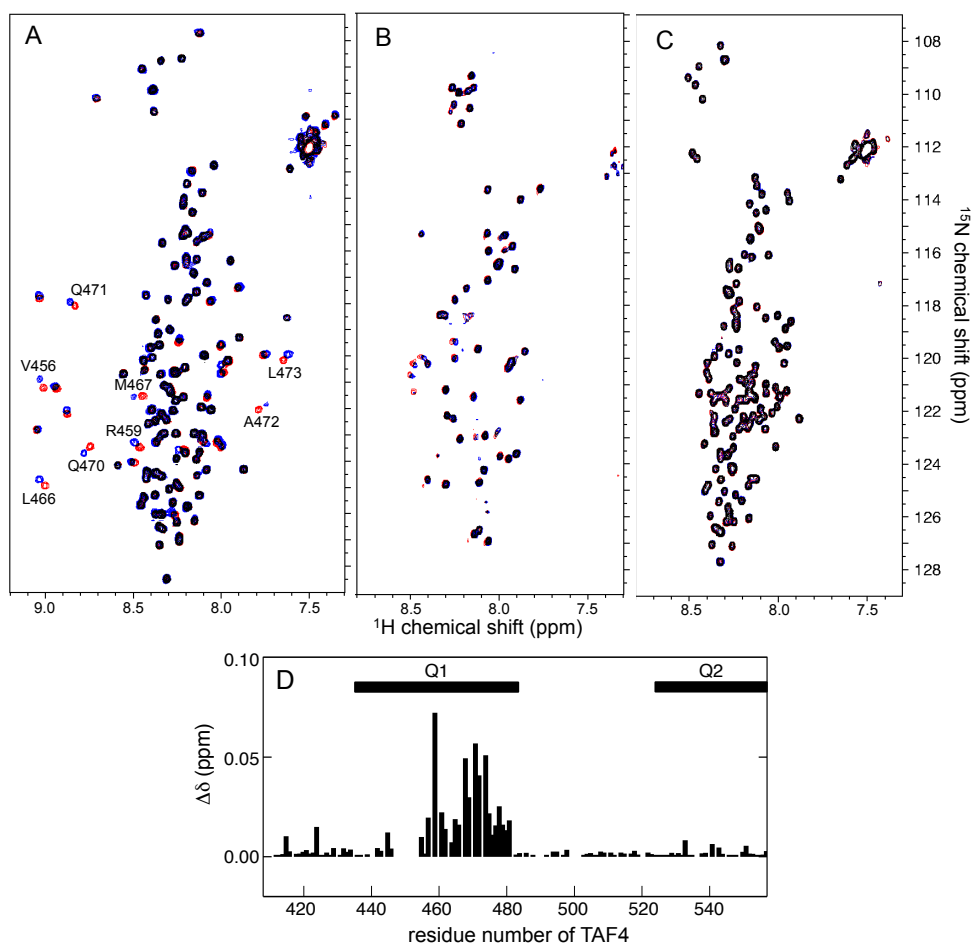


Figure 2-6. (A-C) The ¹H-¹⁵N HSQC spectra of ¹⁵N-TAF4-Q12 (A), ¹⁵N-TAF4-CI (B), and ¹⁵N-TAF4-Q34 (C) measured at pH 5.0 and 37 °C. In each panel, two spectra are superimposed, which are recorded in the absence (red) and presence (blue) of unlabeled Sp1-QB. The concentrations of all the ¹⁵N-labeled proteins were 50 μM. The peaks observed in black indicate that neither the chemical shifts nor intensities of the peaks changed. Peaks that showed a marked difference in chemical shift values are labeled by the residue name and number. (D) The apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{1H})^2 + (\Delta\delta_{15N}/5)^2}$, calculated from the ¹H-¹⁵N HSQC spectra of ¹⁵N-TAF4-Q12 measured in the absence and presence of unlabeled Sp1-QB (panel A).

The distribution of changes in chemical shift values indicated that only the Q1 domain in TAF4-Q12 contributed significantly to the interaction with Sp1-QB. In order to clarify whether the residues responsible for the interaction with Sp1-QB are exclusively located in the regions corresponding to TAF4-Q1, as observed in **Fig. 2-6D**, the author prepared two fragment proteins, TAF4-Q1 and TAF4-Q2 (see **Fig. P-2A**), and recorded their ^1H - ^{15}N HSQC spectra separately in the absence and presence of unlabeled Sp1-QB (**Fig. 2-7A** and **2-7B**). Apparent chemical shift differences in the absence and presence of Sp1-QB were plotted against residue numbers in **Fig. 2-7C** and **2-7D**. Although chemical shift values for the residues located in the C-terminal half of TAF4-Q1 changed markedly in the presence of an equimolar amount of Sp1-QB, no significant changes were observed in the chemical shift values or intensities of the signals for TAF4-Q2 even in the presence of a 5-fold excess amount of unlabeled Sp1-QB. Furthermore, the distribution of the change in the averaged chemical shift change ($\Delta\delta$) along the residue number showed almost the same pattern as that obtained from the interaction between TAF4-Q12 and Sp1-QB (**Fig. 2-6D**), which confirmed that the residues responsible for binding are solely located in the region of TAF4-Q1 and also that TAF4-Q2 does not significantly contribute to this interaction.

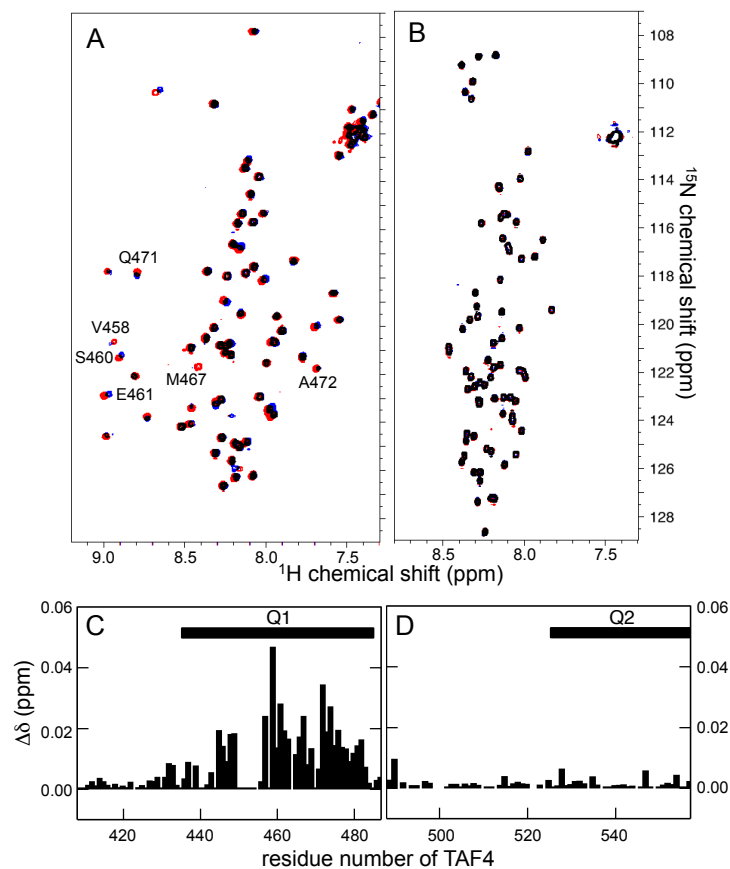


Figure 2-7. (A) ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q1 in the absence (red) and presence (blue) of an equimolar amount of unlabeled Sp1-QB. The concentration of ^{15}N -TAF4-Q1 was 100 μM . (B) ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q2 in the absence (red) and presence (blue) of a five-fold excess amount of Sp1-QB. The concentration of ^{15}N -TAF4-Q2 was 50 μM . The peaks observed in black indicate that neither the chemical shifts nor intensities of the peaks were changed. Peaks that showed marked differences in chemical shift values are labeled by the residue name and number. (C, D) The apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{1H}})^2 + (\Delta\delta_{\text{15N}}/5)^2}$, of ^{15}N -TAF4-Q1 and ^{15}N -TAF4-Q2 measured in the absence and presence of unlabeled Sp1-QB (panel A and B, respectively).

2-3-5. The interaction between fragment proteins of Sp1-QB and TAF4N/C

The author examined the interaction between Sp1 and TAF4 using smaller fragment proteins at pH 5.0 and 37 $^{\circ}\text{C}$, and successfully observed the NMR signals of the complex. Therefore, the author performed a similar experiment to that described in

section **2-3-1**, that is, ^1H - ^{15}N HSQC spectra were recorded for ^{15}N -Sp1-QB in the absence and presence of an excess amount of unlabeled TAF4-Q12, TAF4-CI, and TAF4-Q34. The author did not detect any significant change regardless of the absence or presence of unlabeled TAF4-CI or TAF4-Q34 (**Fig. 2-8B** and **2-8C**). In contrast, the author found that the chemical shift values of several peaks markedly changed following the addition of unlabeled TAF4-Q12, suggesting a specific interaction between TAF4-Q12 and Sp1-QB (**Fig. 2-8**). In addition, the apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{1H}})^2 + (\Delta\delta_{\text{15N}}/5)^2}$, plotted against residue number of Sp1, the chemical shift values of residues corresponding to the C-terminal half of the QB were markedly perturbed, suggesting a possible binding site for TAF4-Q12 (**Fig. 2-8D**). In order to confirm the idea that the binding region is exclusively located in the C-terminal half of the Sp1-molecule, the author prepared two fragment proteins, Sp1-QBn and Sp1-QBc (see **Fig. P-2B**).

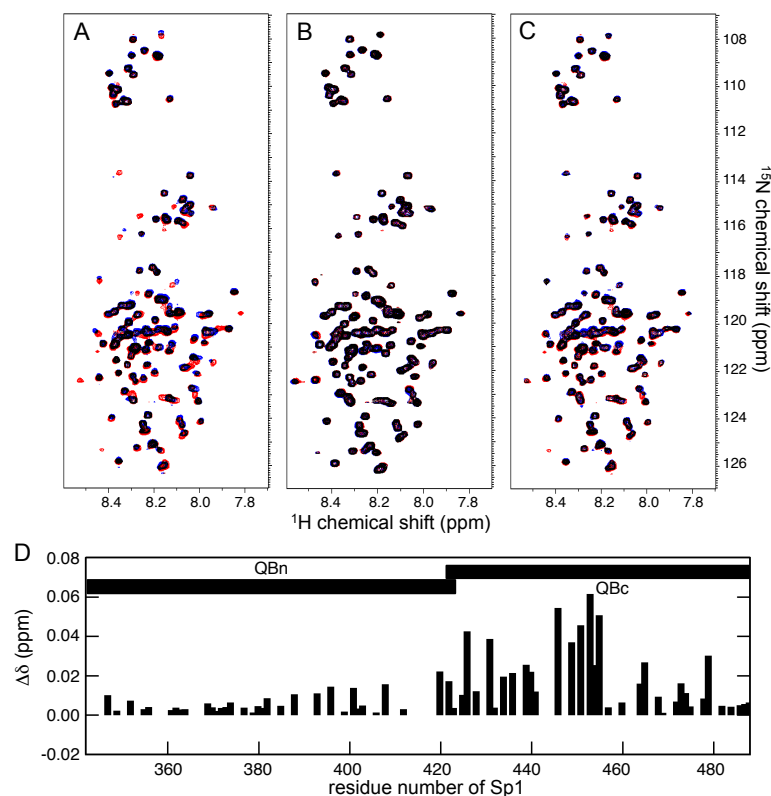


Figure 2-8. (A-C) Overlay of ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-QB in the absence (red) and presence (blue) of unlabeled TAF4-Q12 (A), TAF4-CI (B), and TAF4-Q34 (C). The concentrations of ^{15}N -Sp1-QB in all spectra were 50 μM . The peaks observed in black indicate that neither the chemical shifts nor intensities of the peaks were changed. (D) The apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{1H}})^2 + (\Delta\delta_{\text{15N}}/5)^2}$, of ^{15}N -Sp1-QB obtained from spectra measured in the absence and presence of unlabeled TAF4-Q12 (panel A).

In order to identify the binding region, ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBn and ^{15}N -Sp1-QBc were measured in the absence and presence of unlabeled Q-domains in TAF4. No significant differences were observed in the chemical shift values or peak intensities in the spectra of Sp1-QBn, even in the presence of a 5-fold molar excess amount of unlabeled TAF4N/C, suggesting that the N-terminal half of Sp1-QB (Sp1-QBn) did not contribute significantly to the interaction (**Fig. 2-9A**). On the other hand, several peaks showed a large displacement in the chemical shift value for the spectra of the C-terminal half of Sp1-QB (Sp1-QBc) with increases in the concentration of unlabeled TAF4-Q12 added (**Fig. 2-9B**). Furthermore, the plot of the averaged

chemical shift change against residue number showed that the most affected residues were located exclusively from the center to the C-terminus of Sp1-QBc (**Fig. 2-9C** and **2-9D**). These results were consistent with the results obtained at pH 7.3 and 4 °C, which also suggested the importance of the residues located from the center to the C-terminus of Sp1-QB (**Fig. 2-1C**).

Next, the author measured several ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBc in the presence of various amounts of unlabeled TAF4-Q12, and found that the chemical shift values of ^{15}N -Sp1-QBc were dependent on the concentration of unlabeled TAF4-Q12 present in the solution. In order to analyze these results in more detail, the author made two assumptions: (1) the formation of a 1:1 complex between Sp1-QBc and TAF4-Q12, and (2) “a fast exchange regime”, namely, the apparent chemical shift change was proportional to the concentration of the resultant equimolar complex. Based on these assumptions, the author analyzed the apparent chemical shift change as the concentration of TAF4-Q12 added (**Fig. 2-9E**) using equation [3]. The results of curve fitting were converged well, and binding constants for several representative residues were estimated to be $1.4 - 2.8 \times 10^4 \text{ M}^{-1}$.

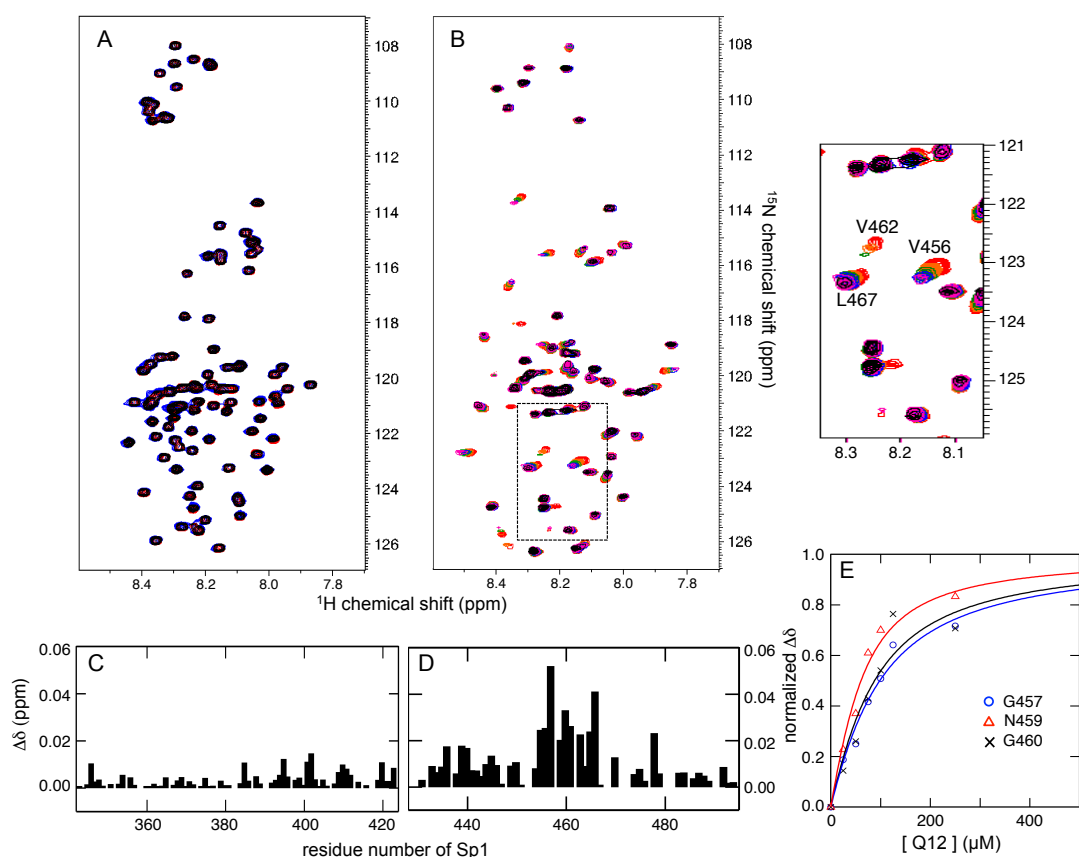


Figure 2-9. (A) Overlay of ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBn in the absence (red) and presence (blue) of unlabeled TAF4N/C. The concentration of ^{15}N -Sp1-QBn was 50 μM . The peaks observed in black indicate that neither the chemical shifts nor intensities of the peaks changed. (B) ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBc measured in the presence of various concentrations of unlabeled TAF4-Q12. The concentration of ^{15}N -Sp1-QBc was 50 μM . Spectra in the absence and the presence of 0.5, 1.0, 1.5, 2.0, and 2.5 molar ratio of unlabeled TAF4-Q12 to ^{15}N -Sp1-QBc are colored in red, orange, green, blue, purple and black, respectively. The region indicated by the dashed box in the spectrum is expanded in the right panel. The chemical shifts of several peaks change linearly. (C, D) The apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}}/5)^2}$, of ^{15}N -Sp1-QBn (C) and ^{15}N -Sp1-QBc (D) obtained from spectra measured in the absence and presence of unlabeled TAF4-Q12 (panel A and B, respectively). (E) The relative chemical shift change for several representative residues of ^{15}N -Sp1-QBc as a function of the added unlabeled TAF4-Q12. Lines are the theoretical curves assuming 1:1 binding and a fast exchange regime (proportionality of the apparent chemical shift change to the amount of the complex). Estimated binding constants are 1.36×10^4 , 2.80×10^4 , and $1.59 \times 10^4 \text{ M}^{-1}$ for G457, N459, and G460, respectively.

The author also performed a similar experiment using the opposite combination of isotopic labeling, that is, unlabeled Sp1-QBn or Sp1-QBc was added separately to ^{15}N -TAF4-Q12 and examined whether any changes in the spectra would be observed (**Fig. 2-10**). The results obtained were the same as those expected, namely, a marked change in the chemical shift values of ^{15}N -TAF4-Q12 was only observed when unlabeled Sp1-QBc was added (**Fig. 2-10B**). The author also measured several ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q12 in the presence of various amounts of unlabeled Sp1-QBc, and found that the peaks, which shifted by the addition of unlabeled TAF4-Q12, were located on a straight line (**Fig. 2-10B**), as well as ^{15}N -Sp1-QBc in **Fig. 2-9B**. Collectively, these results suggest that the interaction between TAF4 and Sp1 is very specific and highly localized. In Sp1, only the C-terminal half (Sp1-QBc) was needed for the interaction with TAF4. On the other hand, in TAF4, only the C-terminal half of the first Q-domain (TAF4-Q1) was exclusively important for the interaction with Sp1. The author summarized interactions between several fragment proteins in Table 1.

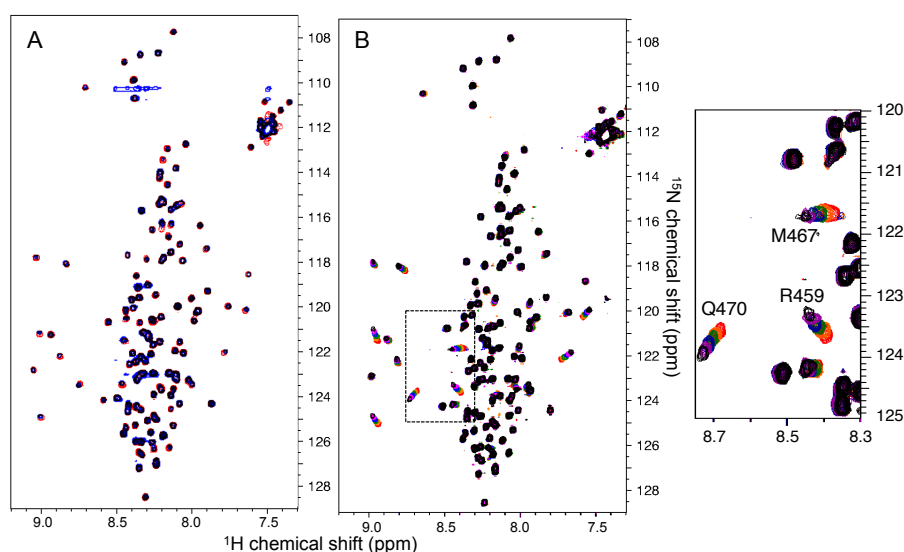


Figure 2-10. (A) The ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q12 measured in the absence (red) and presence (blue) of unlabeled Sp1-QBn. The concentration of ^{15}N -TAF4-Q12 was 50 μM . (B) The ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q12 measured at various concentrations of the unlabeled Sp1-QBn. The concentration of ^{15}N -TAF4-Q12 was 50 μM . Spectra in the absence and the presence of 0.5, 1.0, 1.5, 2.0, 2.5 molar ratio of unlabeled Sp1-QBc to ^{15}N -TAF4-Q12 are colored in blue, green, yellow, orange, red, and black, respectively. The region indicated by the dashed box in the spectrum is expanded in the right panel.

observed protein	added protein	molar ratio ^a	change in spectra	corresponding figure
¹⁵ N-Sp1-QA	TAF4N/C	1 : 3	n.d. ^b	Fig. 2-1A
¹⁵ N-Sp1-QB	TAF4N/C	1 : 1	disappeared ^c	Fig. 2-1B, 2-1C
¹⁵ N-Sp1-QB	TAF4-Q12	1 : 1	moved ^d	Fig. 2-8A, 2-8D
¹⁵ N-Sp1-QB	TAF4-CI	1 : 2.5	n.d.	Fig. 2-8B
¹⁵ N-Sp1-QB	TAF4-Q34	1 : 2.5	n.d.	Fig. 2-8C
¹⁵ N-Sp1-QBn	TAF4N/C	1 : 5	n.d.	Fig. 2-9A, 2-9D
¹⁵ N-Sp1-QBc	TAF4-Q12	1 : 0.5–2.5	moved	Fig. 2-9B, 2-9D
¹⁵ N, ¹³ C-Sp1-QBc	TAF4-Q12	1 : 1	n.d.	Fig. 2-11A
¹⁵ N-TAF4-Q12	Sp1-QB	1 : 1.5	moved	Fig. 2-6A, 2-6D
¹⁵ N-TAF4-Q12	Sp1-QBn	1 : 5	n.d.	Fig. 2-10A
¹⁵ N-TAF4-Q12	Sp1-QBc	1 : 0.5–2.5	moved	Fig. 2-10B
¹⁵ N-TAF4-CI	Sp1-QB	1 : 3	n.d.	Fig. 2-6B
¹⁵ N-TAF4-Q34	Sp1-QB	1 : 2	n.d.	Fig. 2-6C
¹⁵ N-TAF4-Q1	Sp1-QB	1 : 0.8	moved	Fig. 2-7A, 2-7C
¹⁵ N-TAF4-Q2	Sp1-QB	1 : 5	n.d.	Fig. 2-7B, 2-7D
¹⁵ N, ¹³ C-TAF4-Q1	Sp1-QB	1 : 0.8	n.d.	Fig. 2-11B

Table 1. Summary for intermolecular interactions investigated in chapter 2.

^a the ratio of [observed protein] : [added protein]

^b not detected

^c disappearance of several peaks in a concentration-dependent manner

^d remarkable change in chemical shift values for several residues

2-3-6. No significant conformational change upon binding

The chemical shift values of NMR signals are very sensitive indicators that reflect any changes in the electron density of the nucleus of interest, and very subtle changes in the local environment may result in a large difference in the chemical shift value. Therefore, slight changes in chemical shift value in ^1H - ^{15}N HSQC spectra are not necessarily linked to conformational changes in a protein. On the other hand, the chemical shift values of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ are known to be influenced sensitively by the ϕ and ψ dihedral angles of the residue of interest, and have been used to analyze secondary structures at residual resolution⁵⁰. Therefore, the author measured 3D-HNCACB spectra in order to obtain the chemical shift values of $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ for ^{15}N , ^{13}C -Sp1-QBc, and ^{15}N , ^{13}C -TAF4-Q1, respectively, in the absence and presence of unlabeled TAF4-Q12 or Sp1-QB. Differences in the ^{13}C chemical shift values of Sp1-QBc and TAF4-Q1 were shown in **Fig. 2-11**. No significant differences in ^{13}C chemical shift values were found throughout the molecules in both proteins, even for residues that showed a marked change in the chemical shift values in ^1H - ^{15}N HSQC spectra. These results suggest that the interaction between Sp1 and TAF4 is not accompanied by any significant conformational changes in either protein, at least at the level of the secondary structure.

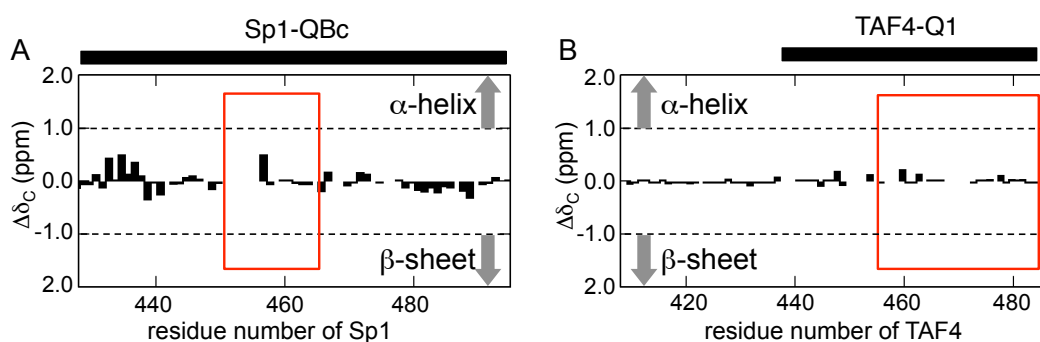


Figure 2-11. Changes in ^{13}C chemical shift values $\Delta\delta_{\text{C}}$ calculated by $\Delta\delta_{\text{C}} = \Delta\delta_{\text{C}\alpha} - \Delta\delta_{\text{C}\beta}$. $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shift values were obtained from 3D-HNCACB spectra. The concentrations of the ^{15}N , ^{13}C -labeled proteins were 100 μM . Red boxes represent the binding site determined by ^1H - ^{15}N HSQC experiments. In this plot, the values more than +1 or less than -1 suggest a significant propensity for the formation of α -helix or β -sheet, respectively. (A) Changes in the chemical shift values of ^{13}C of ^{15}N , ^{13}C -Sp1-QBc in the absence and presence of an equimolar amount of unlabeled TAF4-Q12. (B) Changes in the chemical shift values of ^{13}C of ^{15}N , ^{13}C -TAF4-Q1 in the absence and presence of an equimolar amount of unlabeled Sp1-QBc.

In order to confirm this, the author also measured the CD spectra of Sp1-QBc (93 a.a.) and TAF4-Q1 (80 a.a.). Since all residues in a protein contribute to the signal of the CD spectrum, the author used these minimal fragment proteins of interest in order to facilitate the detection of any conformational changes. Otherwise, the presence of many residues that do not contribute to the interaction may obscure potentially small changes in the spectra. The author measured the CD spectra of Sp1-QBc, TAF4-Q1, and their equimolar mixture, and compared the spectrum of the mixture with the sum of the two separately measured spectra (**Fig. 2-12**). The intensity of CD was expressed as an apparent ellipticity value with the unit of mdeg because two proteins consisted of different numbers of amino acid residues. The two spectra, obtained from the mixed sample and calculated sum of two separately measured spectra, were indistinguishable from each other. These results are in good agreement with the analysis of the ^{13}C -chemical shift values, and suggest that there were no significant conformational changes in Sp1 or TAF4 at the level of the secondary structure even when they interacted with each other.

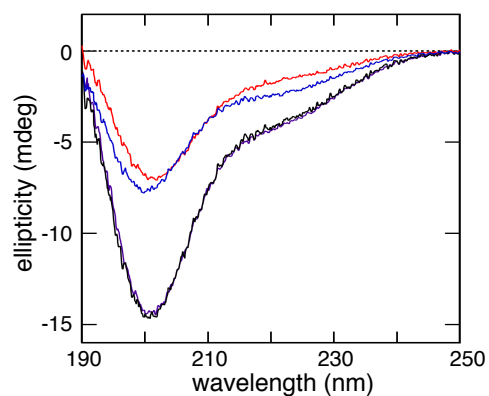


Figure 2-12. Far UV-CD spectra of TAF4-Q1 (red), Sp1-QBc (blue), and their equimolar mixture (purple). The concentrations of all the proteins were 50 μ M. The black line represents the simple sum spectrum of TAF4-Q1 and Sp1-QBc that was recorded separately. Note that the purple and black lines are indistinguishable from each other.

2-4. Discussion

2-4-1. The interaction between Q-domains of Sp1 and TAF4

Titration experiments monitored by the ^1H - ^{15}N HSQC spectra of Sp1-QB revealed that the intensities of several residues were markedly decreased in the presence of unlabeled TAF4N/C in a concentration-dependent manner (**Fig. 2-1B**). Moreover, the plot of relative peak intensities against residue numbers shown in **Fig. 2-1C** suggests that the interaction site is located between the center and C-terminus of QB in Sp1. These results are consistent with previous biochemical studies in which a fragment protein of Sp1 corresponding to the C-terminal half of the QB domain (residues 428-549) was suggested to interact with dTAF_{II}110, a *Drosophila* homologue of human hTAF_{II}130 (TAF4)^{14,15}.

In contrast to Sp1-QB, neither the chemical shift nor peak intensity of ^{15}N -Sp1-QA changed in the presence of unlabeled TAF4N/C up to a three molar excess amount (**Fig. 2-1A**). On the other hand, Hoey *et al.*¹³ reported that Sp1-QA and Sp1-QB interacted with dTAF_{II}110. In addition, the central region of human hTAF_{II}130 (TAF4) has also been suggested to interact with the QA and QB domains¹⁷. Both studies analyzed the interaction between the Q domains of Sp1 and TAF using a yeast two-hybrid system, in which many other cellular components may have contributed to the interaction. Therefore, the author concluded that Sp1-QA did not interact by itself with TAF4N/C, at least under the conditions examined.

2-4-2. C-terminal half of the Sp1-QB domain is responsible for the interaction with TAF4

The author considered that the decrease in peak intensity of ^{15}N -Sp1-QB with addition of unlabeled TAF4N/C resulted from the increase of the molecular weight of the complex. Therefore, the author also examined the interaction between Sp1-QB and TAF4 using smaller fragment proteins under slightly acidic conditions and a higher temperature (pH 5.0 and 37 °C) to obtain spectra of Sp1-QB in complex, as summarized in **Table 1**. The chemical shift values of only the residues within the range from the central to C-terminal part of Sp1-QB significantly changed by the addition of unlabeled TAF4-Q12 (**Fig. 2-8A**). The importance of the C-terminal region in Sp1-QB was further confirmed using smaller fragment proteins in which Sp1-QB was divided into two parts,

Sp1-QBn and Sp1-QBc. The addition of unlabeled Sp1-QBn did not significantly affect the chemical shift values or intensities of the peaks of the ^1H - ^{15}N HSQC spectra of ^{15}N -TAF4-Q12, suggesting that these proteins did not strongly interact with each other (**Fig. 2-10A**). On the other hand, remarkable changes in chemical shift values for many residues were observed when unlabeled Sp1-QBc was added to ^{15}N -TAF4-Q12, indicating that only residues that consist of Sp1-QBc are sufficient for the interaction between Sp1-QB and TAF4N/C (**Fig. 2-10B**).

The importance of the Sp1-QBc region was further evidenced by reverse experiments, in which the ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBn or ^{15}N -Sp1-QBc were compared in the absence and presence of unlabeled TAF4-Q12. No significant change was observed in the spectra of ^{15}N -Sp1-QBn, even in the presence of a five-fold molar excess amount of unlabeled TAF4N/C (**Fig. 2-9A, 2-9C**). On the other hand, marked differences were noted for many peaks in the spectra of ^{15}N -Sp1-QBc in a concentration-dependent manner of unlabeled TAF4-Q12 (**Fig. 2-9B, 2-9D**). These results suggest that residues located in the center to C-terminal part of Sp1-QB are important for the interaction with TAF4.

2-4-3. TAF4-Q1 is responsible for the interaction with Sp1-QB

The author also elucidated the interaction region of TAF4 with Sp1-QB, as summarized in **Table 1**. There are four Q-domains in the central region of TAF4, and one or more of these Q-domains are considered to be involved in the interaction with various transcriptional activators such as Sp1 and CREB^{51,52}. The author herein prepared several types of ^{15}N -labeled fragment proteins from TAF4N/C and examined whether they interact with Sp1 using ^1H - ^{15}N HSQC experiments, which suggest that only the fragment proteins that contain the region corresponding to the Q1-domain in TAF4 showed marked changes in chemical shift values (**Fig. 2-6A, 2-6D, 2-7A, 2-7C**). On the other hand, the other regions did not show any significant change in spectra between the absence and presence of unlabeled counterparts of the hetero-oligomer (**Fig. 2-6B, 2-6C**). A detailed analysis of the distribution of the residues showing marked changes in chemical shift values following the addition of Sp1-QB revealed that the effects of Sp1-QB were specific and localized within a narrow range of 455–480 of TAF4.

As described in chapter 1, the comparison of CD spectra suggested that while the

The author also analyzed the primary sequence of TAF4, and found that not only the highly conserved region (CI), but also the Q1 region was conserved very well between vertebrates including fish and amphibia (**Fig. 2-14**), suggesting the importance of this region in the function of TAF4.

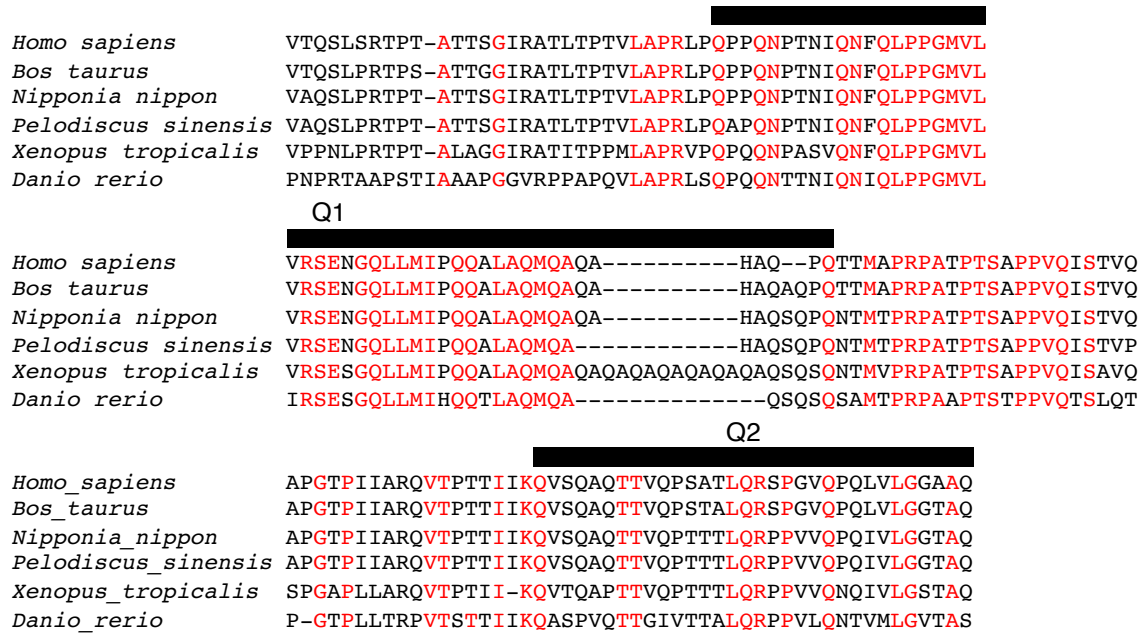


Figure 2-14. A sequence alignment analysis of the TAF4-Q12 region among various species by ClustalW 2.1⁵⁶. The GenBank Accession Number or NCBI Reference Sequence ID for the sequences are EAW75406.1, XP_015321532.1, XP_009459856.1, XP_006126225.1, AAI21309.1, and XP_005163125.1, respectively. Identical residues among different species are indicated by red.

In contrast to Q1 domain, the other three Q-domains (Q2, Q3, and Q4) as well as the central highly conserved region, CI, did not show any evidence for an interaction with Sp1-QB. TAF4 is known to interact with a number of transcriptional activators, which are considered to recognize different parts of the TAF4 molecule³⁹. For example, the highly conserved region CI has been reported to bind to the short hydrophobic motif DΨΨζζΨΦ present in transcriptional regulators³⁰. Therefore, regions other than the C-terminal part of the Q1 domain may be important for interactions with a number of transcriptional regulators in the versatile regulatory protein TAF4.

2-4-4. Quantitative analysis of the interaction between Q-domains of Sp1 and TAF4

The author examined the interaction between Sp1-QB and TAF4N/C using SPR, which provides the association and dissociation rate constants, k_a and k_d . Sensorgrams were analyzed by the standard Langmuir model, which assumes 1:1 binding between the ligand and analyte. The binding constant for Sp1-QB and TAF4N/C was calculated to be $1.5 \times 10^4 \text{ M}^{-1}$, indicating a significant interaction between these proteins.

The author found the chemical shift perturbations for several residues of ^{15}N -Sp1-QBc were markedly changed in the presence of unlabeled TAF4-Q12 in a concentration-dependent manner. The author analyzed changes in the chemical shift perturbation of the ^1H - ^{15}N HSQC spectra of Sp1-QBc recorded in the presence of various concentrations of unlabeled TAF4-Q12 on the basis of the assumption that binding stoichiometry between Sp1-QBc and TAF4-Q12 is 1:1. By using this assumption, chemical shift perturbation as a function of the concentration of TAF4-Q12 added was analyzed with equation [3]. From this analysis, the author estimated that the binding constant between Sp1-QBc and TAF4-Q12 was $1.4\text{--}2.8 \times 10^4 \text{ M}^{-1}$. These values were consistent with the estimation by SPR experiments between full-length Sp1-QB and TAF4N/C at pH 7.3 and 15 °C.

2-4-5. Self-association of Sp1-QB versus the heterogeneous interaction with TAF4

Q-domains are found in many nuclear proteins such as transcriptional factors, and are considered to be important for molecular recognition. The promoter-specific transcriptional factor Sp1 also possesses two Q-domains, QA and QB, which have been suggested to be responsible for the self-association by Sp1 as well as the interaction with TAF4. Biochemical studies indicated that the self-association of Sp1 enhances its transcriptional activity synergistically, and this effect is known as “superactivation”^{23,51}. On the other hand, the interaction between Sp1 and TAF4 (hTF_{II}130), which is one of the components of the general transcription factor TFIID, is considered to recruit RNA polymerase II to the transcription initiation site and, thus, activate transcription.

The author and colleagues previously found that Sp1-QB self-associated under nearly physiological condition, at pH 7.3 and 4 °C³⁸. This process was analyzed using analytical ultracentrifugation, and the binding constant was estimated to be 4.5×10^3

M^{-1} on the basis of a simple homo-dimerization model. The author and colleagues also examined the signal intensities of the 1H - ^{15}N HSQC spectra of Sp1-QB at various concentrations, and found that the binding site was localized from the center to the C-terminal of the molecule³⁸. The binding site for the self-association of Sp1-QB was similar to that for the interaction with TAF4N/C or TAF4-Q12 (**Fig. 2-1C, 2-8D**). This result suggests that the self-association of Sp1-QB domains and the heterogeneous interaction between Sp1-QB and TAF4N/C compete with each other. It is important to note that the binding constant for the interaction of Sp1-QB with TAF4N/C was stronger than that for the self-association of Sp1-QB, suggesting that a heterodimer composed of Sp1-QB and TAF4N/C is predominantly formed even if the two processes compete. Therefore, the author propose the following interaction model as possible events during the process of activation of certain DNA sequences by Sp1. [1] The Sp1 molecule first recognizes and binds via Zn-fingers to the GC-box located upstream of the transcription initiation site⁵⁷. [2] Sp1 molecules that bind to the same DNA will self-associate with each other via their QB regions, and this process will synergistically enhance affinity for the other transcriptional factors. [3] TAF4 preferentially interacts with multiple Sp1 molecules, resulting in the recruitment of RNA polymerase as well as the dissociation of the preformed Sp1 homo-oligomer. [4] The dissociation of the Sp1 homo-oligomer results in the decrease of the number of Sp1 molecules near TAF4, which may contribute to preventing prolonged transcriptional activation. Such a negative-feedback mechanism is considered to be important for the proper regulation of the timing of gene activation.

The association constant of the interaction between Q-domains in Sp1 and TAF4 is on the order of $10^4 M^{-1}$. Moreover, the binding constant of self-association of Sp1-QB was estimated to be on the order of $10^3 M^{-1}$ by the author and colleagues previously³⁸. Since both binding constants are relatively low, the author assumes that there are some concentration effects by DNA *in vivo*. In the case of Sp1, it has three zinc fingers at its C-terminus (see **Fig. P-2B**) which can bind to a GC-box on DNA (see **Fig. P-1**). GC-boxes located upstream (approximately -70)⁵⁸ and 1.7 kb downstream of the transcriptional start site²³. It has been considered that the Sp1 molecules bind to the “distal” GC-box synergistically interacts with another Sp1 molecules at the “proximal” binding site. Moreover, Sp1 has synergetic effect known as “superactivation”, which is considered a result of the interaction between Sp1 molecules via the Q-domains^{23,51}.

That is, Sp1 form homo-oligomer on DNA. On the other hand, in the case of TAF4, TAF4 is component of the general transcription factor TF_{II}D which also contains the TATA-box binding protein (TBP). In general, TATA-box also located upstream of the transcriptional start site (approximately -30)⁵⁸. Therefore, both Sp1 and TAF4 exist in close spatial proximity near the transcription initiation site (**Fig. P-1**). If the interaction is detected using Sp1 containing Q-domains and zinc fingers and TF_{II}D containing TAF4 and TBP, the association constant relatively becomes large by support of concentration effects by DNA.

2-4-6. A rare interaction mode of the disordered region

The author identified the important regions for the interaction unambiguously as Sp1-QBc and TAF4-Q1 (**Table 1**). As described above, the chemical shift values of NMR signals are very sensitive indicators that reflect changes in the electron density of the nucleus of interest, and therefore, slight changes in chemical shift values in ¹H-¹⁵N HSQC spectra are not necessarily linked to conformational changes in a protein. The author analyzed the change in the chemical shift values of ¹³C_α, ¹³C_β, and ¹³C', which are sensitive indicators of backbone ϕ and ψ dihedral angles of the residue of interest. In contrast to the results obtained for chemical shift differences in ¹H-¹⁵N HSQC spectra, no significant changes were observed in these ¹³C chemical shift values, suggesting that the structures of Sp1-QB and TAF4N/C do not significantly change when they interact with each other (**Fig. 2-11**). The result of CD spectra also suggests that the interaction between Sp1-QBc and TAF4-Q1 is not accompanied by the significant conformational changes, in which a 1:1 mixture of Sp1-QBc and TAF4-Q1 was almost identical to the sum of two separately measured spectra of two proteins (**Fig. 2-12**).

Among these proteins, the interacting residues in TAF4-Q1 were suggested to partially fold into strand-turn-strand followed by an α-helix based on analyses of backbone chemical shift values using CSI 2.0/3.0 and TALOS+. On the other hand, the residues in Sp1-QBc did not show any preference for the formation of the secondary structure in the interaction region with TAF4-Q1, although it was predicted to have some β-strands in the different region from the binding site (**Fig. 2-15**).

	GILQAISQETLQNQLQLQAVPNSGPII	IIRTPTVGPNGQVSWQT	LQLQNLQV
CSI 2.0	CCCCCCCCCCCCCCCCCCBBBB	CCCCCCCCCCCBBB	CCCCCCCCC
CSI 3.0	CCCCCCCCCCCCCCCCCCBBBBB	CCCCCCCCCCBBBH	CCCCCCCCC
TALOS+	XCCCCCCCCCCCCCCCCCCBBB	CCCCCCCCCXCHHH	CXXX
	QNPQAQTITLAPMQGVSLGQT		
CSI 2.0	CCCCCCCCCCCCCCCCCC		
CSI 3.0	CCCCCCCCCCCCCCCCCC		
TALOS+	XXCCCCCCCCCCCCCCCCCC		

Figure 2-15. The prediction of secondary structures for Sp1-QBc by backbone chemical shift values of ^1H , ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $^{13}\text{C}'$ using CSI 2.0, CSI 3.0 and TALOS+⁵³⁻⁵⁵. C : random coil, H : α -helix, B : β -sheet, X : undefined. The interaction region was represented by red box.

Recent advances in bioinformatics and genome sequencing suggest that an increasing number of eukaryotic proteins are intrinsically disordered. Many of these IDPs are found to be involved in crucial and important cellular functions such as transcriptional regulation, translation, and signal transduction. One of the important structural features for these IDPs are “flexibility”, which has been suggested to enable them to access a broad conformational space to interact with a wide array of biomolecular targets. Many IDPs undergo a disorder-to-order transition to form well-defined structures upon binding to their cellular targets. This process is called “coupled-folding and binding”, and is suggested to be a common mechanism for IDPs to interact with their target molecules. The author herein revealed the binding regions in TAF4N/C and Sp1-QB, and found that at least the binding region of Sp1-QBc did not change its conformation, even after binding to its target molecule, TAF4-Q1. The results of the present study suggest that this binding mode is not categorized into the well-accepted concept of the “coupled-binding and folding” mechanism, which might be because the interaction through IDPs but not between an IDP and a folded protein. The interactions though an IDP under the “coupled-binding and folding” mechanism are reported in the case of interactions between IDPs and folded proteins^{25,59}.

Note that the association constant between Q-domains in Sp1 and TAF4 is relatively weak, that is, on the order of 10^4 M^{-1} (see **2-4-4.**). In the experiments of the

detection of conformational changes using ^{13}C chemical shift values, the concentrations of ^{15}N , ^{13}C -labeled proteins and unlabeled proteins were around 100 μM because of experimental limitations in triple-resonance NMR experiments (see **Fig. 2-11**). That is, almost half amount of molecules does not form the complex formation. However, in the case of Sp1-QBc, the ^1H chemical shift values of the peaks that shifted by the addition of unlabeled TAF-Q12 were around 8.2 ppm (see **Fig. 2-9B**), indicating that the corresponding residues are random-coil like state. On the other hand, in the case of TAF4-Q1, the maximum of changes in ^{13}C chemical shift values, $\Delta\delta_{\text{C}}$, was only around 0.03 ppm (see **Fig. 2-11B**). When $\Delta\delta_{\text{C}}$ value is greater than 1 ppm, formation of secondary structure is suggested. From the results of NMR experiments of Sp1-QBc and TAF4-Q1, the author assumes that both Sp1-QBc and TAF4-Q1 do not form any secondary structure even if all Sp1-QBc and TAF4-Q1 molecules are involved in the complex formation.

Chapter 3

Elucidation of mechanism for molecular recognition on the formation of Sp1-TAF4 complex

3-1. Introduction

Early molecular biological study using the site-directed mutagenesis by Gill *et al.*¹⁸ identified the residues involved in the interaction of Sp1-QBc with *Drosophila* TAF4 (dTAF_{II}110); the substitutions of amino acids were introduced into Sp1-QBc and then the activities of interactions were evaluated by yeast two-hybrid assays. They found that two mutations of Sp1-QBc, namely W464A and L467A/L469A/L472A, resulted in the complete loss of binding ability to *Drosophila* TAF4. On the other hand, Sp1-QBc-Q468A/Q470A/N471A mutation was not significantly affected the interaction. That is, hydrophobic residues in this region, W464, L467, L469 and L472, were reported to be important for the interaction. Note that the precision of the detection is relatively low and the detection of the interaction is somehow indirect because this method relies on the expression levels of the downstream gene.

On the other hand, the author identified that the interaction region of Sp1 was around between 450 and 465 residue number by analyzing ¹H-¹⁵N HSQC spectra in chapter 2. The author focused on the difference between the region determined by NMR in chapter 2 and that reported by Gill *et al.*¹⁸. Notably, the binding region estimated by NMR does not contain L467, L469 and L472 residues. In chapter 3, the author attempted to elucidate the effect of these hydrophobic residues on the interaction of Sp1-QBc with TAF4 in expectation of understanding mechanisms of molecular recognitions.

3-2. Materials and Methods

Protein preparation

The expression plasmids for Sp1-QBc-W464A, Sp1-QBc-L467A/L469A/L472A and Sp1-QBc-Q468A/Q470A/N471A mutant proteins were constructed as fusion protein with Ub contained N-terminal extension of a His6-tag using the QuickChange method with pET-His6-Ub-QBc as a template. The pET-His6-Ub-Sp1-QBc-W464A, pET-His6-Ub-Sp1-QBc-L467A/L469A/L472A and pET-His6-Ub-Sp1-QBc-Q468A/Q470A/N471A vectors obtained were introduced into *E. coli* strain Rosetta2(DE3)/pLysS (Novagen). The expression and purification of three kinds of Sp1-QBc mutants were performed as wild type Sp1-QBc described in chapter 1.

Sp1-QBc (Sp1 [424-495]) and TAF4-Q12 (TAF4 [408-557]) were constructed as fusion protein with Ub contained N-terminal extension of a His6-tag. The expression and purification of these proteins were described in chapter 1.

All proteins were dialyzed against buffer C (20 mM sodium acetate, pH 5.0).

NMR experiments

^1H - ^{15}N HSQC spectra were measured and processed as described in chapter 1 and 2. Steady-state heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE spectra were measured with or without ^1H -saturation pulses at 4.7 ppm. The irradiation was performed at the strength of 25 kHz for 5 seconds. The NOE values were calculated from the ratios of cross peak volumes ($I_{\text{NOE}}/I_{\text{noNOE}}$) in the spectra recorded with (I_{NOE}) and without (I_{noNOE}) ^1H -saturation. All spectra were measured at 37 °C.

3-3. Results

3-3-1. Analyses of the interactions of Sp1-QBc mutants with TAF4-Q12

Referring to the previous report by Gill *et al.*¹⁸ in which the interactions were investigated by yeast two-hybrid assays using Sp1-QBc mutants, the author attempted to re-examine the interactions between Sp1-QBc mutants and TAF4-Q12 using purified proteins by NMR spectroscopy. In the first step of the experiments, three kinds of Sp1-QBc mutant proteins (Sp1-QBc-W464A (Sp1-QBc-WA), Sp1-QBc-L467A/L469A/L472A (Sp1-QBc-L₃A₃) and Sp1-QBc-Q468A/Q470A/N471A (QNA₃)) were prepared by *E. coli* overexpression systems. The three kinds of proteins were labeled uniformly with ¹⁵N and then the interactions with unlabeled TAF4-Q12 were analyzed by ¹H-¹⁵N HSQC spectra performed as described in chapter 2. The results showed that both spectra of ¹⁵N-Sp1-QBc-WA and ¹⁵N-Sp1-QBc-L₃A₃ mutant proteins did not change in peak intensities and positions even by the addition of excess amount of unlabeled TAF4-Q12 (**Fig. 3-1A** and **3-1B**). On the other hand, in the case of ¹⁵N-Sp1-QBc-QNA₃ mutant proteins, the position and peak intensities of the several peaks changed by the addition of equimolar amount of TAF4-Q12, which was similar to ¹H-¹⁵N HSQC spectra of the complex formed by wild type Sp1-QBc with TAF4-Q12 described in chapter 2 (**Fig. 3-1C**). These results suggest that both Sp1-QBc-WA and Sp1-QBc-L₃A₃ mutant proteins lost the binding affinity for TAF4-Q12. On the other hand, Sp1-QBc-QNA₃ mutant was interacted with TAF4-Q12 in a similar manner to the wild type Sp1-QBc protein. These results were consistent with the previous report using yeast two-hybrid systems¹⁸.

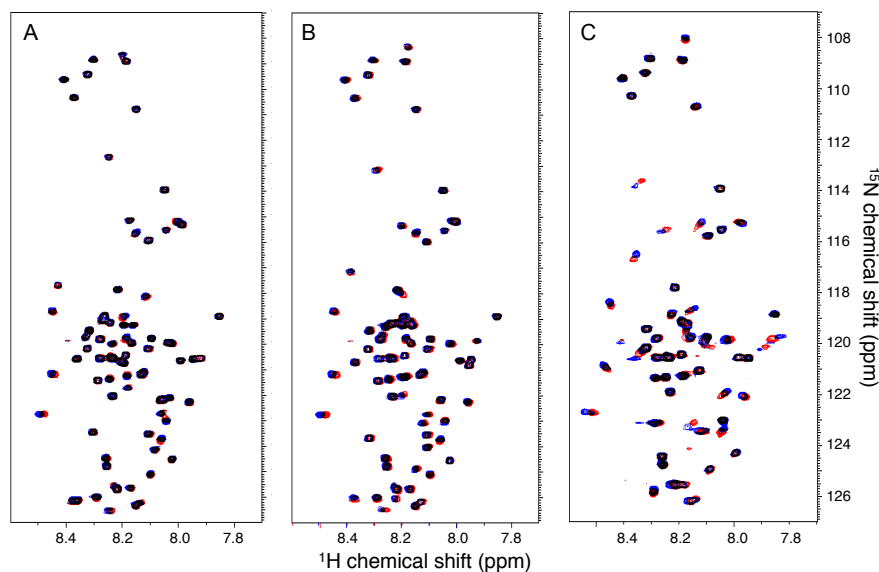


Figure 3-1. (A-B) The ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBc-WA (A) and ^{15}N -Sp1-QBc-L₃A₃ (B) mutant proteins measured in the absence (red) and presence of 5-fold amount of unlabeled TAF4-Q12 (blue). The concentrations of all the ^{15}N -labeled proteins were 50 μM . (C) The ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBc-QNA₃ mutant proteins measured in the absence (red) and presence of equimolar amount of unlabeled TAF4-Q12 (blue). The concentration of ^{15}N -Sp1-QBc-QNA₃ mutant protein was 50 μM .

3-3-2. Comparing NMR spectra of wild type and mutant proteins of Sp1-QBc

In the previous section (3-3-1), the author confirmed that the results of previous mutational study by Gill *et al.*¹⁸ were consistent with the results of the NMR titration experiments. That is, Sp1-QBc-WA and Sp1-QBc-L₃A₃ mutants lost the affinity for TAF4-Q12, while Sp1-QBc-QNA₃ mutant bound to TAF4-Q12. However, these results were somehow confusing because these mutated residues analyzed here were located outside of the binding site elucidated by the NMR titration experiments performed on the wild type protein as shown in chapter 2. Therefore, the author examined the conformations of the mutant proteins by comparing the ^1H - ^{15}N HSQC spectra with that of wild-type Sp1-QBc. The superposition of the spectra of wild type and mutant proteins revealed that most of the peaks of the intensities and the chemical shift values did not change. Therefore, the author considered that the structure of the protein was not significantly perturbed by these mutations, although several residues did change their

chemical shift values (**Fig. 3-2**). Furthermore, the author found that the residues, which were affected by their mutations, were overlapped each other. However, these changes in the ^1H - ^{15}N HSQC spectra did not explain by itself the loss of function in Sp1-QBc-W464A and Sp1-QBc-L₃A₃ mutants, because almost the same changes in the spectra were observed in Sp1-QBc-QNA₃ mutants, which bound to TAF4-Q12. A careful comparison of the spectra between the mutant proteins and wild type protein revealed that the residues affected by the mutations were largely overlapped. The author noticed that those affected residues in ^{15}N -labeled mutant proteins corresponded to those in wild type ^{15}N -Sp1-QBc affected significantly by the addition of unlabeled TAF4-Q12 (see **Fig. 3-4A** and **3-4B** in section **3-3-3**).

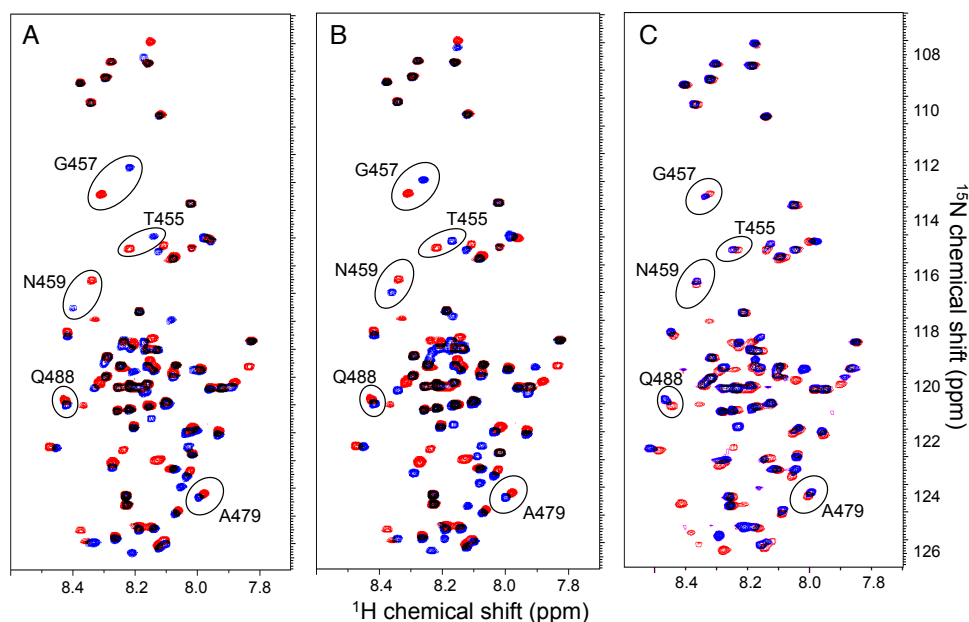


Figure 3-2. The overlapped ^1H - ^{15}N HSQC spectra of ^{15}N -Sp1-QBc-WA (A), ^{15}N -Sp1-QBc-L₃A₃ (B), and ^{15}N -Sp1-QBc-QNA₃ (C) mutant proteins with wild type ^{15}N -Sp1-QBc. The concentrations of the proteins were 50 μM . The spectra of ^{15}N -Sp1-QBc wild type and ^{15}N -labeled Sp1-QBc mutant proteins are colored in red and blue, respectively. Several peaks of those that showed marked differences in chemical shift values are labeled by the residue name and number.

In order to further examine the relationship between the complex formed by wild type Sp1-QBc with TAF4-Q12 and function-loss mutant proteins (Sp1-QBc-WA and Sp1-QBc-L₃A₃), the ¹H-¹⁵N HSQC spectra of those were superimposed (**Fig. 3-3**). Surprisingly, the overlapped ¹H-¹⁵N HSQC spectra showed that the affected peaks of Sp1-QBc mutant proteins were shifted in opposite direction compared with the complex formed by wild type Sp1-QBc with TAF4-Q12. Moreover, the four peaks were located on a straight line (right panel in **Fig. 3-3**). That is, the affected peaks of wild type Sp1-QBc in the absence of TAF-Q12 located between those of wild type Sp1-QBc in the presence of TAF4-Q12 and Sp1-QBc mutant proteins.

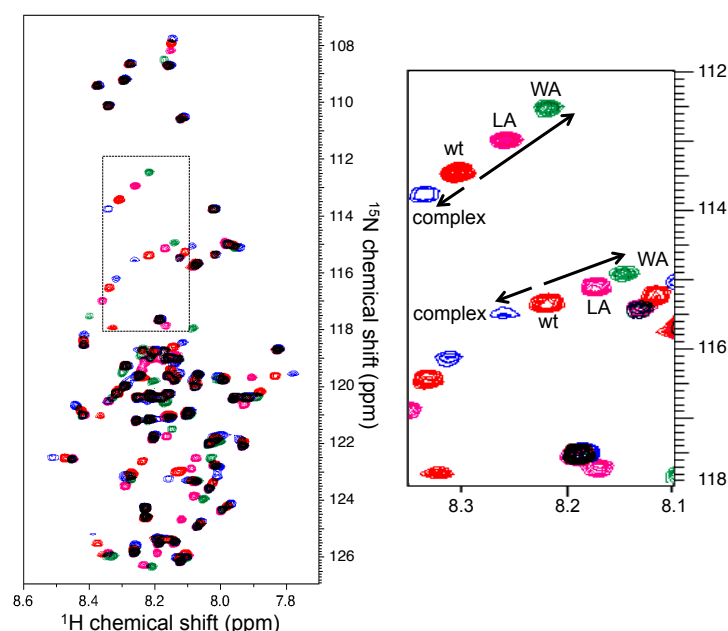


Figure 3-3. ¹H-¹⁵N HSQC spectra of ¹⁵N-Sp1-QBc-WA (green) and ¹⁵N-Sp1-QBc-L₃A₃ (magenta), ¹⁵N-Sp1-QBc in the presence and absence of unlabeled TAF4-Q12 (red and blue, respectively). The concentrations of the ¹⁵N-labeled proteins were 50 μM. The region indicated by the dotted box in the spectrum is expanded in the right panel. The “wt”, “complex”, “LA” and “WA” represent the peaks of wild type Sp1-QBc in the absence of TAF4-Q12, wild type Sp1-QBc in the presence of TAF4-Q12, Sp1-QBc-L₃A₃ and Sp1-QBc-WA, respectively. Arrows represent the direction from the peaks of wild type Sp1-QBc in the absence of TAF4-Q12 to those of the complex formed by wild type Sp1-QBc with TAF4-Q12 or the mutant proteins.

3-3-3. The analysis of Sp1-QBc-WA and the complex formation of wild type Sp1-QBc with TAF4-Q12

In order to elucidate why the Sp1-QBc-WA and Sp1-QBc-L₃A₃ mutant proteins lost the affinity for TAF4-Q12, the author attempted to examine their structural properties further in detail at the residue level. In order to clarify the results, the author focused on the Sp1-QBc-WA mutant, because this mutant protein showed larger chemical shift changes than those of the Sp1-QBc-L₃A₃ mutant protein (**Fig. 3-3**). The apparent chemical shift perturbations of the Sp1-QBc-WA mutant protein in the absence of TAF4-Q12 compared with wild type Sp1-QBc in the absence of TAF4-Q12 were plotted against residue number of Sp1 (**Fig. 3-4B**). For comparison, the apparent chemical shift changes of wild type Sp1-QBc upon the addition of unlabeled TAF4-Q12 were also shown in **Fig. 3-4A**. Those affected residues were fairly overlapped with the binding site to TAF4-Q12. It should be noted that, the directions of the chemical shift changes were opposite between the mutation and the binding with TAF4-Q12 (see **Fig. 3-3**).

Next, the author examined whether the Sp1-QBc-WA mutation and the complex formation of wild type Sp1-QBc with TAF4-Q12 affected the dynamics of the protein molecule. The relaxation rate constants of ¹⁵N-spins are mostly affected by the directly bound ¹H-spins. Therefore, the cross-relaxation rate constant of ¹⁵N-spin is the sensitive indicator for the mobility of ¹⁵N-¹H vector^{60,61}. The author measured heteronuclear {¹H}-¹⁵N NOE values of the Sp1-QBc-WA mutant and wild type Sp1-QBc in the absence and presence of unlabeled TAF4-Q12 (**Fig. 3-4C** and **3-4D**). In these plot, the negatively larger values suggest the higher mobility of the corresponding ¹⁵N-¹H vectors. The {¹H}-¹⁵N NOE values of ¹⁵N-Sp1-QBc-WA alone at the binding region were more negative and those of ¹⁵N-Sp1-QBc in the presence of TAF4-Q12 were a little more positive than those of ¹⁵N-Sp1-QBc in the absence of TAF4-Q12 (represented by red boxes in **Fig. 3-4C** and **3-4D**). Similarly to the changes in the chemical shift values, the results were again the opposite direction between the Sp1-QBc-WA mutation and the complex formation with TAF4-Q12. That is, the mobility of Sp1-QBc-WA mutant protein at the binding region increased compared with wild type Sp1-QBc, which suggests to be related to the low affinity of Sp1-QBc-WA mutant. On the other hand, it is natural that the mobility of the complex formed by wild type Sp1-QBc with TAF4-Q12 at the binding region decreased because the TAF4-Q12 should restrain the

movement of the binding site in Sp1-QBc.

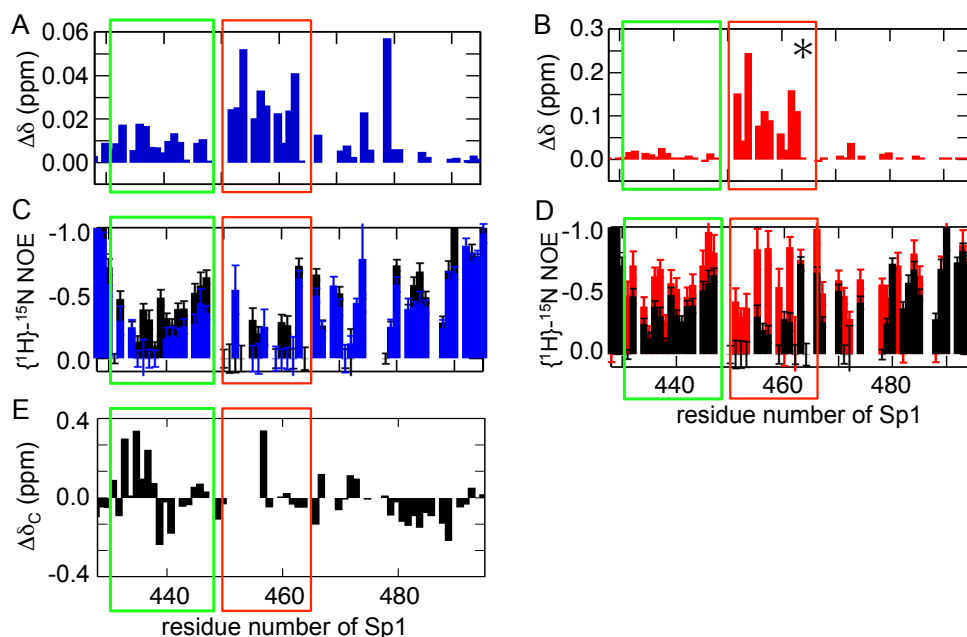


Figure 3-4. Comparisons between wild type Sp1-QBc in the presence of TAF4-Q12 and Sp1-QBc-WA mutant protein in the absence of TAF4-Q12. In all figures, red and light green boxes represent the interaction regions determined by ^1H - ^{15}N HSQC experiments in chapter 2 and the “A-region” (see **3-3-4**), respectively. (A) The apparent chemical shift change, $\Delta\delta = \sqrt{(\Delta\delta_{\text{1H}})^2 + (\Delta\delta_{\text{15N}}/5)^2}$, calculated of wild type ^{15}N -Sp1-QBc measured in the absence and presence of unlabeled TAF4-Q12, which is same as **Fig. 2-9D**. (B) The apparent chemical shift change of ^{15}N -Sp1-QBc-WA mutant protein alone compared with ^{15}N -Sp1-QBc in the absence of TAF4-Q12 calculated from those of the ^1H - ^{15}N HSQC spectra. An asterisk (*) shows the position of the mutation. (C-D) Heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE values of wild type Sp1-QBc in the absence and presence of TAF4-Q12 and Sp1-QBc-WA mutant protein in the absence of TAF4-Q12 were shown by black, blue and red bars, respectively. Note that, the top is negative for better understanding. (E) Changes in the ^{13}C chemical shift values of ^{15}N , ^{13}C -Sp1-QBc in the absence and presence of an equimolar amount of unlabeled TAF4-Q12. This is the expanded figure of **Fig. 2-11A**.

3-3-4. Investigation of conformation changes in Sp1

The above observations emphasized that the “negative” correlations between

Sp1-QBc-WA mutant protein and the complex formed by wild type with TAF4-Q12 in the central region (450-465) of the Sp1-QBc molecule. That is, the direction of the chemical shift change and the mobility of each residue ($\{^1\text{H}\}$ - ^{15}N NOE value) showed the opposite tendency. The author noticed by the careful analysis of these chemical shift data that another region (430-445 in Sp1, called as “A-region”) also showed a different effect by the mutation and TAF4-Q12 binding (represented by light green box in **Fig. 3-4A** and **3-4B**). From the $\{^1\text{H}\}$ - ^{15}N NOE values in this region, the mobility of the complex formed by wild type Sp1-QBc with TAF4-Q12 decreased, while that of the Sp1-QBc-WA mutant protein increased (represented by light green box in **Fig. 3-4C** and **3-4D**). And then, the author reconsidered ^{13}C chemical shift change, which indicates secondary structure propensity. The ^{13}C chemical shift change of Sp1-QBc plotted against the residue number of Sp1 (**Fig. 3-4E**) showed that the values at the region around residue number of 435 increased. In this plot, positive value indicates α -helix propensity and negative value does β -strand propensity. That is, α -helical propensity of the A-region in wild type Sp1-QBc increased by the addition of TAF4-Q12.

3-4. Discussion

3-4-1. Re-examination of the interaction of Sp1-QBc mutants to TAF4 by NMR

Early molecular biological study identified that W464, L467, L469 and L472 in Sp1 involved in the interaction of Sp1-QBc with TAF4 by yeast two-hybrid assays using Sp1-QBc mutants by Gill *et al*¹⁸. The region containing these residues was slightly different from the region identified in chapter 2. In particular, almost no chemical shift perturbations were observed at the region of the C-terminus from L467 containing L467, L469 and L472 (**Fig. 3-4A**), which suggests that the addition of TAF4-Q12 does not affect the environment of the spins at the region in Sp1-QBc. The results indicates that TAF4-Q12 does not bind to this region in wild type Sp1-QBc. The author attempted to clarify the reason why the interaction region identified in chapter 2 was inconsistent with that reported by the mutational studies by Gill *et al*¹⁸. One of the possible reasons is the difference in the method to detect the interaction, that is, NMR chemical shift perturbations and yeast two-hybrid assays. Therefore, the author re-examined the effect of mutations of Sp1-QBc by the NMR experiments in a similar manner to the detection of the interaction region of wild type Sp1-QBc. In the first step

of experiments, the author prepared mutant proteins used in the reported study by Gill *et al.*¹⁸, that is, Sp1-QBc-WA, Sp1-QBc-L₃A₃, and Sp1-QBc-QNA₃ mutant proteins. The results revealed that Sp1-QBc-WA and Sp1-QBc-L₃A₃ mutants did not bind to TAF4-Q12, and that Sp1-QBc-QNA₃ mutant bound, which was consistent with the previous report using yeast two-hybrid assays¹⁸ (**Fig. 3-1**).

3-4-2. A possible formation of transient hydrophobic core

The two different kinds of method showed the same results that Sp1-QBc-WA and Sp1-QBc-L₃A₃ mutant proteins did not bind to TAF4-Q12, and that Sp1-QBc-QNA₃ mutant protein bound. Therefore, the author considered that the loss of the interaction ability of the Sp1-QBc-WA and Sp1-QBc-L₃A₃ was attributed to the nature of the proteins.

The author analyzed ¹H-¹⁵N HSQC spectra of wild type ¹⁵N-Sp1-QBc in the absence and presence of unlabeled TAF4-Q12, and two kinds of Sp1-QBc mutant proteins (Sp1-QBc-WA and Sp1-QBc-L₃A₃) which lost the binding to TAF4-Q12 (**Fig. 3-3**). As a result, the regions of the relative chemical shift perturbations changed largely in wild type ¹⁵N-Sp1-QBc in the presence of TAF4-Q12 and ¹⁵N-Sp1-QBc-WA in the absence of TAF4-Q12 were overlapped. Moreover, the affected peaks of Sp1-QBc-WA shifted in opposite direction compared with those of wild type Sp1-QBc in the absence of TAF4-Q12, and four kinds of peaks were located on a straight line.

In NMR experiment, when a protein is in equilibrium between two states exchanging fast ($k_{ex} > 10^3 \text{ s}^{-1}$), the peak position is located at the average of the spectra of two states. On the other hand, in the case of slow exchange ($k_{ex} < 10 \text{ s}^{-1}$), two peaks are observed⁶². In addition, NMR peaks reflect relative populations of two states. In the NMR titration experiments of Sp1-QBc in chapter 2, the position of the affected peaks shifted with the concentration of unlabeled TAF4-Q12 (see **Fig. 2-9B**), which represents that the association and dissociation of Sp1-QBc with TAF4 exchange fast. Then, since the affected peaks of wild type Sp1-QBc in the absence and presence of TAF4-Q12, Sp1-QBc-WA alone and Sp1-QBc-L₃A₃ alone were also located on a straight line similar to the superimposed spectrum from the titration experiments of ¹⁵N-Sp1-QBc with unlabeled TAF4-Q12, the author assumes that wild type Sp1-QBc in the absence of TAF4-Q12 is also in fast equilibrium between two states and that one state is similar

to Sp1-QBc-WA and the other is similar to the complex formation. Here, all mutated residues (W464, L467, L469 and L472) are hydrophobic and mutually close at least in the primary sequence. It is important to note that these two kinds of mutants where the hydrophobic residues were replaced by Ala, also lost the ability to bind to TAF4-Q12. This suggests that a possible transient formation of the hydrophobic core by W464, L467, L469 and L472 might be important to interact with TAF4-Q12. The author calls one state having transient hydrophobic core as a “residual state” and the other state similar to Sp1-QBc-WA as a “random state”.

This hypothesis is consistent with the results of heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE experiments. The results of the analyses of the NOE values showed that the mobility at the interaction region in Sp1-QBc-WA mutant protein increased and that in wild type Sp1-QBc decreased compared with that in wild type Sp1-QBc without TAF4-Q12 (**Fig. 3-4C** and **3-4D**). On the other hand, it is supposed that the mobility of the “residual state” and the “random state” are low and high, respectively. That is, it is accountable that the conformations of the “residual state” and the “random state” at the interaction region are similar to those of the complex formed by wild type Sp1-QBc and Sp1-QBc-WA, respectively.

3-4-3. The conformational changes introduced to the region distant from the binding site

Compared the apparent chemical shift changes of ^{15}N -Sp1-QBc in the presence of TAF4-Q12 with that of ^{15}N -Sp1-QBc-WA mutant protein in the absence of TAF4-Q12, different tendency was observed around residue number of 440, and then the author herein calls this region as “A-region” (light green boxes in **Fig. 3-4A** and **3-4B**).

The relative chemical shift perturbations at the A-region in the complex formed by wild type ^{15}N -Sp1QBc with unlabeled TAF4-Q12 are larger than those in ^{15}N -Sp1-QBc-WA alone in comparison with ^{15}N -Sp1-QBc alone, while the extent of changes at the A-region upon the complex was one third of those observed at the interaction region (**Fig. 3-4A**), suggesting that the A-region did not appear to be the direct binding site. On the other hand, the results of $\{^1\text{H}\}$ - ^{15}N NOEs showed that the mobility of the A-region in the complex formed by wild type Sp1-QBc decreased and that in the Sp1-QBc-WA mutant protein increased (**Fig. 3-4C** and **3-4D**). Moreover, the

^{13}C chemical shift changes of the complex formed by wild type Sp1-QBc at the A-region indicated relatively large positive values (**Fig. 3-4E**). That means the A-region gets α -helical propensity with the addition of TAF4-Q12. Therefore, the author supposed that some conformational changes might occur in the A-region (**Fig. 3-5**). Although it is not possible at this stage to distinguish these conformational changes occur sequentially or simultaneously, the two kinds of conformational changes, that is, a transient formation of α -helical structures at the A-region and the clustering of W and L residues, might be caused by the binding of TAF4-Q12.

The author considers the conformational change at the A-region is classified as allosteric effect. Although allosteric effects are often found in the interaction between folded proteins and are related to many important biological functions of them, the effect in an IDP has never been found. The allosteric effect of Sp1 suggests the possibility that the A-region in Sp1-QBc interacts with other proteins including transcription factors. In future, developments of molecules that bind to the A-region in Sp1 bring better understanding about physicochemical properties of transcription regulation through Sp1.

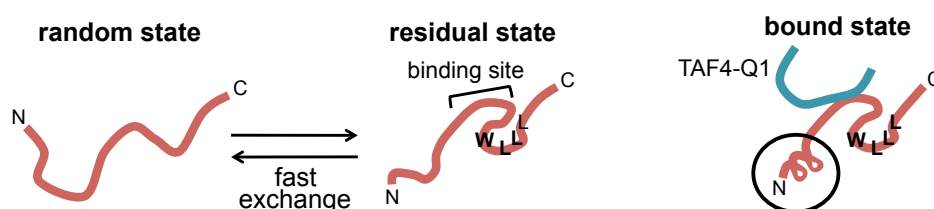


Figure 3-5. Schematic representation of equilibrium of wild type Sp1-QBc. Sp1-QBc and TAF4-Q1 are colored by red and blue, respectively. In Sp1-QBc, N-terminus and C-terminus are shown in figure. “W” and “L” represent the residues which are substituted by A in this study. Wild type Sp1-QBc is assumed to be in fast equilibrium between the “random state” and the “residual state” when the interaction partner, TAF4-Q12, does not exist. Furthermore, if TAF4-Q12 exists, α -helicity of N-terminal region of Sp1-QBc (“A-region”, represented by black circle) increases.

Conclusions

The transcription factor Sp1 and the general transcription factor TAF4 have two and four Q-domains, respectively. Although Q-domains have been reported to be important for interactions by previous genetic and biochemical studies, the molecular structures of these proteins remain unclear, and quantitative analyses on this interaction are limited. In chapter 1, the author examined the structure of the Q-domains in Sp1 and TAF4 by several physicochemical methods including NMR, CD and SAXS experiments, and found that they were IDPs.

Most IDPs interact with other molecular with some secondary structure changes. This mechanism is known as “coupled folding and binding”, and most of IDPs interact on this mechanism. In chapter 2, the author investigated the interaction between Q-domains in Sp1 and TAF4 and identified the interaction regions of Sp1 and TAF4 at residue level using high-resolution NMR experiments. Moreover, in order to determine conformational changes, the author evaluated ^{13}C chemical shift changes, which are useful for identifying the conformation changes of main chain. However, the results showed that any secondary structure changes did not occur in the interaction between Q-domains in Sp1 and TAF4. That is, it is revealed that the interaction mode between Sp1 and TAF4 is not well-known “coupled folding binding mechanism”.

In Chapter 3, in order to examine the discrepancy between the binding site reported previously and that determined in chapter 2, the author executed NMR experiments using ^{15}N -labeled Sp1-QBc-WA, Sp1-QBc-L₃A₃ and Sp1-QBc-QNA₃ mutant proteins in addition to wild type ^{15}N -Sp1-QBc in the absence and presence of unlabeled TAF4-Q12. Through several NMR experiments, the author assumed that Sp1-QBc was in equilibrium between the “random state” and the “residual state” in the absence of the interaction partner TAF4-Q12. Moreover, the author supposed that the association of TAF4-Q12 affected the region away from the binding site in Sp1-QBc, which was like allosteric effect.

This research will shed light on the elucidation of not only the interaction between Sp1 and TAF4, but also the interactions through IDPs.

Acknowledgements

The author would like to express her sincere gratitude to Professor Katsumi Matsuzaki (Graduate School of Pharmaceutical Sciences, Kyoto University) and Associate Professor Masaru Hoshino (Graduate School of Pharmaceutical Sciences, Kyoto University) for giving constant advices throughout this study. The author is also grateful to Senior Lecturer Yoshiaki Yano (Graduate School of Pharmaceutical Sciences, Kyoto University) for his helpful discussions and assistances.

The author wishes to express appreciation to Professor Masaaki Sugiyama (Kyoto University Research Reactor Institute) and Rintaro Inoue (Kyoto University Research Reactor Institute) for technical supports on SAXS and advices.

The author would like to thank Professor Jun Kuwahara (Doshisha Women's College of Liberal Arts) for providing a plasmid encoding Sp1-QA, Sp1-QB, Sp1-QBn and Sp1-QBc.

The author is also grateful to Dr. Takahiro Yamaguchi, Dr. Kenichi Kawano, Dr. So Ishino and the members of the Department of Analytical Chemistry, Graduate School of Pharmaceutical Sciences, Kyoto University.

Finally, the author deeply thanks to her parents, sister and grandparents.

References

1. Hampsey M (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**, 465–503.
2. Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327–335.
3. Albright SR, Tjian R (2000) TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**, 1–13.
4. Blackwood EM, Kadonaga JT (1998) Going the distance: a current view of enhancer action. *Science* **281**, 60–63.
5. Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13**, 2465–2477.
6. Hoey T, Dynlacht BD, Peterson MG, Pugh BF, Tjian R (1990) Isolation and characterization of the *Drosophila* gene encoding the TATA box binding protein, TFIID. *Cell* **61**, 1179–1186.
7. Verrijzer CP, Chen JL, Yokomori K, Tjian R (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**, 1115–1125.
8. Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev* **11**, 3020–3031.
9. Burley SK, Roeder RG (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* **65**, 769–799.
10. Tanese N, Pugh BF, Tjian R (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev* **5**, 2212–2224.
11. Chen JL, Attardi LD, Verrijzer CP, Yokomori K, Tjian R (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93–105.
12. Dynlacht BD, Hoey T, Tjian R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**, 563–576.
13. Hoey T, Weinzierl RO, Gill G, Chen JL, Dynlacht BD, Tjian R (1993) Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell* **72**, 247–260.

14. Ferreri K, Gill G, Montminy M (1994) The cAMP-regulated transcription factor CREB interacts with a component of the TFIID complex. *Proc Natl Acad Sci USA* **91**, 1210–1213.
15. Tanese N, Saluja D, Vassallo MF, Chen JL, Admon A (1996) Molecular cloning and analysis of two subunits of the human TFIID complex: hTAFII130 and hTAFII100. *Proc Natl Acad Sci USA* **93**, 13611–13616.
16. Mengus G, May M, Carr e L, Chambon P, Davidson I (1997) Human TAF(II)135 potentiates transcriptional activation by the AF-2s of the retinoic acid, vitamin D3, and thyroid hormone receptors in mammalian cells. *Genes Dev* **11**, 1381–1395.
17. Saluja D, Vassallo MF, Tanese N (1998) Distinct subdomains of human TAFII130 are required for interactions with glutamine-rich transcriptional activators. *Mol Cell Biol* **18**, 5734–5743.
18. Gill G, Pascal E, Tseng ZH, Tjian R (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci USA* **91**, 192–196.
19. Pugh BF, Tjian R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187–1197.
20. Kadonaga JT, Jones KA, Tjian R (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends Biochem Sci* **11**, 20–23.
21. Dynan WS, Tjian R (1983) Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**, 669–680.
22. Courey AJ, Tjian R (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**, 887–898.
23. Courey AJ, Holtzman DA, Jackson SP, Tjian R (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**, 827–836.
24. Kadonaga JT, Carner KR, Masiarz FR, Tjian R (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**, 1079–1090.
25. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208.
26. Wilkins RC, Lis JT (1999) DNA distortion and multimerization: novel functions of the glutamine-rich domain of GAGA factor. *J Mol Biol* **285**, 515–525.

27. Reijns MA, Alexander RD, Spiller MP, Beggs JD (2008) A role for Q/N-rich aggregation-prone regions in P-body localization. *J Cell Sci* **121**, 2463–2472.
28. Furukawa T, Tanese N (2000) Assembly of partial TFIID complexes in mammalian cells reveals distinct activities associated with individual TATA box-binding protein-associated factors. *J Biol Chem* **275**(38), 29847–29856.
29. Oka S, Shiraishi Y, Yoshida T, Ohkubo T, Sugiura Y, and Kobayashi Y (2004) NMR Structure of Transcription Factor Sp1 DNA Binding Domain. *Biochemistry* **43**(51), 16027–16035.
30. Wang X, Truckses DM, Takada S, Matsumura T, Tanese N, Jacobson RH (2007) Conserved region I of human coactivator TAF4 binds to a short hydrophobic motif present in transcriptional regulators. *Proc Natl Acad Sci USA* **104**(19), 7839–7844.
31. Werten S, Mitschler A, Romier C, Gangloff YG, Thuault S, Davidson I, Moras D (2002) Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J Biol Chem* **277**(47), 45502–45509.
32. Kohno T, Kusunoki H, Sato K, Wakamatsu K (1998) A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}\text{N}/^{13}\text{C}$. *J Biomol NMR* **12**, 109–121.
33. Kusunoki H, Wakamatsu K, Sato K, Miyazawa T, Kohno T (1998) G protein-bound conformation of mastoparan-X: heteronuclear multidimensional transferred nuclear overhauser effect analysis of peptide uniformly enriched with ^{13}C and ^{15}N . *Biochemistry* **37**, 4782–4790.
34. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293.
35. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* **101**, 12491–12496.
36. Kataoka M, Goto Y (1996) X-ray solution scattering studies of protein folding. *Fold Des* **1**, R107–R114.

37. Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.
38. Hiramatsu N, Hibino E, Matsuzaki K, Kuwahara J, Hoshino M (2012) Interaction between isolated transcriptional activation domains of Sp1 revealed by heteronuclear magnetic resonance. *Protein Sci* **21**, 1481–1488.
39. Saluja D, Vassallo MF, Tanese N (1998) Distinct subdomains of human TAFII130 are required for interactions with glutamine-rich transcriptional activators. *Mol Cell Biol* **18**, 5734–5743.
40. Berlow RB, Dyson HJ, Wright PE (2015) Functional advantages of dynamic protein disorder. *FEBS Lett* **589**, 2433–2440.
41. Uversky VN (2015) The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett* **589**, 2498–2506.
42. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**(11), 2411–2423.
43. Wittekind M, Muller M (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson ser B* **101**, 201–205.
44. Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly ¹³C/¹⁵N-enriched proteins. *J Biomol NMR* **3**, 185–204.
45. Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* **96**, 432–440.
46. Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ¹HN, ¹⁵N, and ¹³C' chemical shifts in ¹⁵N–¹³C-labelled proteins. *J Magn Reson* **97**, 213–217.
47. Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ¹⁵N sequential correlations in (¹³C, ¹⁵N) labeled proteins: Application to unfolded proteins. *J Biomol NMR* **20**, 135–147.
48. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293.

49. Garrett DS, Powers R, Gronenborn AM, Clore GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* **95**, 214–220.
50. Wishart DS, Sykes BD (1994) The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**(2), 171–180.
51. Pascal E, Tjian R (1991) Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev* **5**(9), 1646–1656.
52. Mayr BM, Guzman E, Montminy M (2005) Glutamine rich and basic region/leucine zipper (bZIP) domains stabilize cAMP-response element-binding protein (CREB) binding to chromatin. *J Biol Chem* **280**(15), 15103–15110.
53. Hafsa NE, Wishart DS (2014) CSI 2.0: a significantly improved version of the Chemical Shift Index. *J Biomol NMR* **60**(2–3), 131–146.
54. Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. *Nucleic Acids Res* **43**(W1), W370–377.
55. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**(4), 213–223.
56. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21), 2947–2948.
57. Kuwahara J, Yonezawa A, Futamura M, Sugiura Y (1993) Binding of transcription factor Sp1 to GC box DNA revealed by footprinting analysis: different contact of three zinc fingers and sequence recognition mode. *Biochemistry* **32**, 5994–6001.
58. Duncliffe KN, Rondahl ME, Wells JR (1995) A H1 histone gene-specific AC-box-related element influences transcription from a major chicken H1 promoter. *Gene* **163**(2), 227–232.
58. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**(1), 54–60.
59. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**(23), 8972–8979.

60. Sahu SC, Bhuyan AK, Udgaonkar JB, Hosur RV (2000) Backbone dynamics of free barnase and its complex with barstar determined by ^{15}N NMR relaxation study. *J Biomol NMR* **18**(2), 107–118.
61. Palmer III AG, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **339**, 204–238.