Published in final edited form as:
*Proteins*. 2009 August 15; 76(3): 655–664. doi:10.1002/prot.22379.

# Formation of the RFX gene regulatory complex induces folding of RFXAP

**LaTese Briggs**[1], **Kholiswa Laird**[1], **Jeremy M. Boss**[2], and **Colin W. Garvie**[1,*]

1*Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250*

2*Department of Microbiology and Immunology, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322*

## Summary

Major Histocompatibility Complex Class II (MHCII) molecules have a central role in the mammalian adaptive immune response against infection. The level of the immune response is directly related to the concentration of MHCII molecules in the cell, which have a central role in initiating the immune response. MHCII molecules are therefore a potential target for the development of immunosuppressant drugs for the treatment of organ transplant rejection and autoimmune disease. The expression of MHCII molecules is regulated by a cell specific multi-protein complex termed the MHCII enhanceosome. The RFX complex is the key DNA binding component of the MHCII enhanceosome. The RFX complex is comprised of three proteins – RFX5, RFXAP, and RFXB – all of which are required for activation of expression of the MHCII genes. We have identified the precise regions of RFX5, RFXAP, and RFXB, that interact to form the RFX complex and have characterized the individual domains and the complexes they form. In addition, we have shown that the interaction between RFX5 and RFXAP promotes folding of RFXAP and this increases the affinity RFXAP has for RFXB.

### Keywords

RFX; MHCII enhanceosome; gene regulation; biophysical analysis; NMR

## Introduction

Major histocompatibility class II (MHCII) molecules are heterodimeric glycoproteins that bind to antigenic peptides. The MHCII-peptide complex is then presented on the surface of antigen presenting cells, where it interacts with the T-cell receptor of helper T-cells. This interaction leads to a cascade of events that ultimately results in the inflammatory response [1]. Ideally the peptides are derived from bacterial, fungal or viral infections. In certain situations, the antigenic peptides are obtained from host proteins, which can result in autoimmune disorders. The antigens can also be derived from organ or tissue transplantation, which can result in rejection of the transplant. The ability to manipulate the level of MHCII molecules in the cell is therefore of significant interest with regards to treating autoimmune disorders and to prevent organ and tissue transplantation rejection.

*Corresponding author: Colin Garvie, Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250. Telephone: (410) 455-2512; Fax: (410) 455 2608; E-mail: garvie@umbc.edu

The expression of the MHCII genes is regulated by a cell-specific multi-protein complex, termed the MHCII enhanceosome [2]. The key DNA binding component of the MHCII enhanceosome is the RFX complex. The RFX complex is comprised of three proteins – RFX5, RFXAP, and RFXB – all of which are required for expression of MHCII genes [3; 4; 5; 6]. Previous studies have suggested that RFX5 and RFXAP assemble in the cytoplasm before translocating to the nucleus, where they form the full RFX complex with RFXB [7]. The RFX complex then binds to the MHCII promoter, where it promotes assembly of the other components of the MHCII enhanceosome and initiates expression of the MHCII genes [8; 9; 10; 11; 12]. Deletions or point mutations in the RFX proteins have been shown to abrogate expression of the MHCII genes, and directly suppress the inflammatory response [13].

In our previous studies, we expressed and purified RFXB, RFXAP, and a series of fragment of RFX5 (14 and unpublished data}. We revealed that RFX5 dimerizes through two domains – the oligomerization domain and the dimerization domain – that flank the RFX DNA binding domain (Figure 1). In addition, we showed that a dimer of RFX5 interacts with a monomer of RFXAP through the oligomerization domain of RFX5, and this complex binds to a monomer of RFXB to form the full $RFX5_2$•RFXAP•RFXB complex. In our current study we have identified the precise regions of RFX5, RFXAP, and RFXB, that interact to form the RFX complex and have characterized the individual domains and the complexes they form. We show that the region of RFXAP involved in forming a complex with RFX5 and RFXB adopts an unfolded state that becomes folded upon binding to RFX5. These results provide a direct explanation as to the low affinity between RFXAP and RFXB when RFXAP is not bound to RFX5.

## Results

### RFXB and RFXAP bind to distinct regions of RFX5

In our previous studies we reported that we saw no interaction between RFXB and either RFX5 (1-170) or RFX5(168-330), and concluded that RFX5 does not interact with RFXB in the absence of RFXAP [14]. However, recent DNA binding studies suggested that RFXB could interact with RFX5(1-330) (unpublished data). To investigate this, we combined RFX5(1-330) and RFXB in a 2:1 molar ratio and passed them over a Superose 6 10/30 size exclusion column (GE Healthcare) (Figure 2a). Both proteins eluted together in a peak that had a lower elution volume than RFX5(1-330) alone, confirming that the two proteins formed a complex. The ability of RFXB to interact with RFX5(1-330) but not RFX5(1-170) or RFX5(168-330), suggested that both the RFX domain and the dimerization domain are required for the interaction between RFX5 and RFXB. We therefore repeated the SEC experiment using RFX5 (88-330), which encompasses both the RFX domain and the dimerization domain. Both RFX5 (88-330) and RFXB co-eluted in a single peak with an elution volume lower than either protein alone, confirming that RFX5(88-330) and RFXB formed a complex (Figure 2b). RFX5 (88-330) was not observed to interact with RFXAP (data not shown), but RFXAP did co-elute with RFX5(88-330) and RFXB when all three were combined together and passed over the Superose 6 column (Figure 2b). This demonstrated that RFXB can bind to both RFX5(88-330) and RFXAP simultaneously.

In our previous studies we observed that RFX5(1-90) was sufficient to interact with RFXAP but not RFXB. Since RFX5(88-330) can form a complex with RFXB, and RFXAP can form a complex with RFX5(1-90), we hypothesized that the oligomerization domain, the RFX domain, and the dimerization domain of RFX5 are linked through contacts made between RFXAP and RFXB in the RFX complex. To address this hypothesis, we combined RFX5 (1-90), RFX5(88-330), RFXAP, and RFXB in a 2:2:1:1 complex and passed the mixture over a Superose 6 column. All four proteins co-eluted in a single peak, confirming that they formed a complex (Figure 2b and 2c).

### The first 25 amino acids of RFX5 are dispensable for RFX complex formation

Previous studies have suggested that the N-terminal 19 amino acids are dispensable for the function of RFX5 [15]. In addition, secondary structure predictions of the oligomerization domain suggests that residues 1 – 26 are unstructured (Figure 3a). We therefore expressed and purified RFX5(25-90) and compared its properties to RFX5(1-90). RFX5(25-90) was passed over a Superdex 200 10/30 size exclusion column (SEC) and static light scattering (SLS) measurements were collected to determine its mass. The SEC-SLS estimated mass of RFX5 (25-90) was 22.9 ± 0.3, which is between the mass of a dimer (15 kDa) and a tetramer (30 kDa). Previous SEC-SLS analysis of RFX5(1-90) showed the exact same result, with the estimated mass being between a dimer and a tetramer [14]. Subsequent chemical crosslinking and sedimentation equilibrium experiments of RFX5(1-90) confirmed that it was a tetramer in solution. Exposure of RFX5(25-90) to the chemical crosslinking agent EDC gave four crosslinked species on an SDS-PAGE gel (Figure 3b). Considering the electrophoretic mobility of the denatured non-crosslinked RFX5(25-90) subunit, the other bands likely correspond to two, three and four crosslinked subunits. This confirmed that RFX5(25-90) was a tetramer in solution. To determine any effects on structure of deleting the first 24 amino acids, CD spectra in the Far-UV region of RFX5(1-90) and RFX5(25-90) were compared (Figure 3c). Both constructs gave similar spectra, with a maxima at 192 nm and minima at approximately 208 nm and 223 nm, which suggests that they have similar secondary structure. To ascertain whether the majority of RFX5(25-90) was folded, we expressed and purified $^{15}$N-labeled RFX5 (25-90) and collected a $^{15}$N-$^{1}$H HSQC NMR spectrum. The spectrum gave well resolved, widely dispersed peaks, suggesting that the majority of RFX5(25-90) was folded (Figure 3d).

Shorter constructs with deletions at the C-terminal end were also expressed and purified, but they were observed to be significantly less stable than RFX5(25-90) (data not shown). This suggested that although the last 10 – 20 amino acids of the oligomerization domain are predicted to be unstructured, they may have a role in stabilizing tertiary structure.

### The C-terminal domain of RFXAP is sufficient for complex formation

Previous studies have shown that the last 58 amino acids of RFXAP can complement full length RFXAP in the expression of at least one MHCII isotype [16]. This coincides with a helical region that is predicted by secondary structure algorithms to occur between residues 224 – 262 (Figure 4A). We therefore expressed and purified RFXAP(215-272) and compared it to full length RFXAP. The point mutation, F217Y was introduced into this construct in order to follow the protein spectroscopically. RFXAP(215-272) eluted as a single species from a size exclusion column. The small mass of this construct together with a low UV absorbance at 280 nm made it difficult to obtain accurate static light measurements to determine its mass. Since RFXAP (215-272) was expressed as a fusion protein with the chitin binding domain (CBD) fused at its N-terminus, we used the CBD-RFXAP(215-272) fusion protein to determine the mass of RFXAP(215-272). The SEC-SLS estimated mass of the CBD-RFXAP(215-272) was 17.1 ± 0.2 kDa, which is similar to the mass of a monomer (15 kDa). To compare the effect of the N-terminal truncations on the structure of RFXAP, CD spectra were collected in the far-UV region for RFXAP, and RFXAP(215-272) (Figure 4a). Both proteins gave minima at around 207 nm and 224 nm, although RFXAP(215-272) gave a significantly stronger signal at these wavelengths. This is presumably due to contributions from secondary structure in the preceding 214 amino acids of RFXAP. This suggests that RFXAP(215-272) is folded and is likely to contains some helical content. To determine whether the majority of RFXAP(215-272) is folded, we expressed and purified $^{15}$N-labeled protein and collected a 2D $^{15}$N-$^{1}$H HSQC NMR spectrum (Figure 4c). In contrast to the CD analysis, the HSQC spectrum of RFXAP(215-272) showed a cluster of peaks resonating between 8 – 8.5 ppm, which indicates that the protein is unfolded. To ensure that the unfolded state shown by the NMR spectrum was not due to this data being collected at 35 °C, as compared to 25 °C for the CD spectrum, we repeated the

2D $^{15}$N-$^{1}$H HSQC NMR at 25 °C. While this spectrum showed a few additional peaks ,it was essentially identical to the 35 °C spectrum (data not shown). To confirm that the features of the CD spectrum represented absorption from secondary structure, we analyzed the effect of an increasing concentration of urea on RFXAP(215-272) (Figure 4b). RFXAP(215-272) showed a significant loss in the absorption at both 207 nm and 224 nm at 1M urea, which was further reduced as the urea concentration was increased. This confirms that RFXAP(215-272) does contain secondary structure. The loss of secondary structure by addition of 1 M urea also suggests that the secondary structure can be readily unfolded. As a comparison, we repeated the denaturation experiment with RFX5(25-90) (data not shown). RFX5(25-90) required 2 M urea and higher to observe a significant change in the CD spectrum, suggesting it was forming a more stable structure than RFXAP(215-272). Together, these suggests that residues 215 to 272 of RFXAP adopts an unfolded state, but regions of the polypeptide chain have a propensity to form helical structure.

## The ankyrin repeats of RFXB are sufficient for RFX complex formation

Mutations and deletions in the ankyrin repeats of RFXB, which occur between resides 90 – 252, have suggested that they have a critical role in the interaction between RFXB and RFXAP and RFX5 [10; 17; 18]. To determine whether the ankyrin repeats are sufficient for RFX complex formation, we expressed and purified RFXB(88-260), and compared its behavior to full length RFXB. Size exclusion chromatography gave a single species with an SEC-SLS estimated mass of 21.3 ± 1.5 kDa, which is essentially identical to the predicted monomer mass of 19.3 kDa. The CD spectra of both full length RFXB and RFXB(88-260) gave a maxima at 195 nm and two minima at 208 nm and 223 nm (Figure 5a). Apart from a slightly stronger absorption at 195 nm and at 223 nm for RFXB(88-260), the two spectra are similar. This suggests that RFXB and RFXB(88-260) have similar secondary structure. A 2D $^{1}$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled RFXB(88-260) revealed well resolved and widely dispersed peaks, which indicates that the majority of RFXB(88-260) is folded (Figure 5b).

## The minimal regions of the RFX proteins are sufficient for complex formation

To determine whether RFX5(25-90) can form a complex with RFXAP(215-272), the two proteins were combined in a 2:1 ratio and passed over a Superdex 200 10/30 column (Figure 6a). Both proteins eluted together at an elution volume larger than for RFX5(25-90) alone, suggesting that the complex had a smaller hydrodynamic radius than the RFX5(25-90) tetramer. SEC-SLS gave a mass of 19.3 ± 0.2 kDa, which agrees with the predicted mass of the RFX5(25-90)$_2$•RFXAP(215-272) complex (21.9 kDa). Exposure of the RFX5(25-90) + RFXAP(215-272) complex to the EDC crosslinking solution showed five bands on as SDS-PAGE gel (Figure 6b). Considering the electrophoretic mobility of the individual proteins, the bands could be correlated to the following species (going from lowest molecular weight band to highest): one subunit of RFXAP(215-272), one subunit of RFX5(25-90), one subunit of RFX5(25-90) crosslinked to one subunit of RFXAP(215-272), two crosslinked RFX5(25-90) subunits, and two crosslinked RFX5(25-90) crosslinked to one subunit of RFXAP(215-272). This supported formation of the RFX5(25-90)$_2$•RFXAP(215-272) complex. Combining RFX5 (25-90), RFXAP(215-272), and RFXB(88-260) together in a 2:1:1 ratio, respectively, and passing them over the Superdex 200 10/30 column, resulted in all three proteins co-eluting with a SEC-SLS estimated mass of 40.1 ± 3.8 kDa (Figure 6a). This is essentially identical to the predicted mass of the RFX5(25-90)$_2$•RFXAP(215-272)•RFXB(88-260) complex (41.2 kDa). Combining RFX5(25-90), RFXAP(215-272), RFXB(88-260), and RFX5(88-330) in a 2:1:1:2 ratio, respectively, and passing them over a Superose 6 10/30 column resulted in all four proteins eluting simultaneously (Figure 6a). This confirmed that the minimal regions of each RFX protein were sufficient to form the same set of interactions as the longer length proteins.

### RFX5 increases the affinity of RFXAP for RFXB

When RFXAP(215-272) and RFXB(88-260) were combined in a 1:1 ratio and passed over a Superdex 75 10/30 size exclusion column they were observed to partially dissociate (Figure 7a). In contrast, when RFXB(88-260) was combined with the RFX5(25-90)$_2$•RFXAP (215-272) complex, no dissociation of RFXB(88-260) was observed (Figure 6a). This suggested that the interaction between RFX5(25-90) and RFXAP(215-272) significantly increased the affinity of RFXAP(215-272) for RFXB(88-260). To place this on a more quantitative level, we compared the binding of RFXB(88-260) to RFXAP(215-272) and to the RFX5(25-90)$_2$•RFXAP(215-272) complex using isothermal titration calorimetry (ITC). Titration of RFXB(88-260) onto RFXAP(215-272) showed a clear change in enthalpy for each titration, indicating that the two proteins were interacting as predicted (Figure 7b). The interaction, however, was very weak and complete formation of the complex could not be obtained. A reliable dissociation constant could not therefore be fit to the integrated data. In contrast, titration of RFXB(88-260) onto the RFX5(25-90)$_2$•RFXAP(215-272) complex gave a $K_d$ of 75.5 ± 5.5 nM, confirming that the complex has a higher affinity for RFXB(88-260) compared to RFXAP(215-272) alone (Figure 7c).

### The binding of RFX5 to RFXAP induces folding of the C-terminal region of RFXAP

One cause for the low affinity between RFXAP(215-272) and RFXB(88-260) could be due to the unfolded state of RFXAP(215-272). We hypothesized that RFXAP(25-90) induces folding of RFXAP(215-272) when they form the RFX5(25-90)$_2$•RFXAP(215-272) complex, and this folded form of RFXAP(215-272) would have a higher affinity for RFXB(88-260). To determine the effect of RFX5(25-90) on the folding of RFXAP(215-272), we titrated unlabeled RFX5(25-90) onto $^{15}$N-labeled RFXAP(215-272) (Figure 6c). The resultant spectra gave well resolved peaks that are widely dispersed in chemical shifts, suggesting that RFXAP(215-272) is folded upon binding to RFX5(25-90).

### A 78 amino acid region is sufficient to promote dimerization of the helical domain of RFX5

Secondary structure prediction of the dimerization domain of RFX5, residues 168-330, suggested there is a structured domain between residues 197 to 262 that is primarily helical in nature (Figure 8a). To determine if this region was sufficient to promote dimerization of RFX5 we expressed and purified RFX5(187-265) and compared its properties to RFX5(168-330). This was the shortest construct that gave sufficient protein expression for analysis. RFX5 (187-265) eluted as a single species from a Superdex 200 10/30 column with a SLS-estimated mass of 16.8 kDa ± 0.2 kDa, which is close to the predicted mass of a dimer (17.8 kDa). Exposure of RFX5(187-265) to the EDC crosslinking solution resulted in two bands being evident on an SDS-PAGE gel, which is indicative of two cross-linked monomers (Figure 8b). Both the SEC-SLS and the chemical crosslinking therefore support the proposal that residues 187 to 265 comprise the dimerization domain of RFX5. Comparison of the CD spectra in the far-UV of RFX5(168-330) and RFX5(187-265) showed that they are similar, but not identical (Figure 8c). The CD spectrum of RFX5(187-265) gave characteristic features of helical proteins, exhibiting a maxima at 190 nm, and minima at 208 nm and 222 nm, which agrees with the secondary structure prediction. The CD spectrum of RFX5(168-330) gave a similar spectrum, except the deepest minima is closer to 205 nm. This suggests that the region between residues 168 to 187 and/or the region between 266 to 330 adopts some form of secondary structure that contributes to the CD signal. Low yields of RFX5(187-265) in $^{15}$N-enriched media prevented a 2D $^{15}$N-$^{1}$H HSQC being collected. In addition, while RFX5(187-265) gave high expression levels, neither RFX5(1-265) nor RFX5(88-265) gave sufficient yields to determine if they were sufficient to form a complex with RFXB.

## Discussion

We have shown that RFXAP interacts simultaneously with the oligomerization domain of RFX5 and RFXB, and RFXB in turn interacts simultaneously with RFXAP and a region of RFX5 that encompasses the RFX domain and the dimerization domain. We have also identified the precise regions of the three RFX proteins that are responsible for these interactions. The oligomerization domain and dimerization domain of RFX5 can be localized to residues 25 to 90 and 187 to 265, respectively. The region of RFXAP that is sufficient for interaction with RFX5 and RFXB are contained within residues 215 to 272. Finally, the region of RFXB that is sufficient for interaction with both RFXAP and RFX5 is contained within residues 88 to 260, which essentially comprises the four ankyrin repeats. The interactions between the three subunits also appear to be cooperative, as the individual interactions made by RFXB and RFXAP with RFX5 directly increase the affinity between RFXB and RFXAP.

Secondary structure prediction of the oligomerization domain suggests that it is primarily helical. This prediction agrees with the CD spectrum, which is characteristic of a helical protein. The similarity between the CD spectra of RFX5(1-90) and RFX5(25-90) suggests that deletion of the first 24 amino acids has little effect on the structure of the oligomerization domain. Ankyrin repeats are predominantly helical and therefore it is not surprising that the CD spectrum of RFXB(88-260) is also characteristic of a helical protein. Comparison of the CD spectra for RFXB and RFXB(88-260) also suggests that deletion of the first 87 amino acids has little effect on the structure of the ankyrin repeat region. Analysis of RFX5(25-90) and RFXB(88-260) by $^{15}$N-$^{1}$H HSQC NMR show that they are both predominantly folded. Similar to the oligomerization domain of RFX5 and the ankyrin repeat region of RFXB, RFXAP (215-272) showed a CD spectrum characteristic of a helical protein, which agrees with the secondary structure prediction for this region. The secondary structure of this region, however, is susceptible to low concentrations of denaturant (urea), which suggests that it is easily unfolded. This was confirmed by a $^{15}$N-$^{1}$H HSQC NMR spectrum of RFXAP(215-272), which gave a spectrum that is strongly indicative of an unfolded protein. One model that could explain both the CD and the NMR spectra would be if different regions of RFXAP(215-272) existed in a dynamic equilibrium of unfolded and helical states.

The lack of a folded structure of RFXAP may explain why RFXB interacts weakly with RFXAP. If RFXB requires a binding site formed by the folding of RFXAP into a distinct secondary or tertiary structure, then RFXB will have a low affinity for the unfolded state of RFXAP. If the contact region is small then RFXB alone would have difficulty promoting the folding of RFXAP to form the binding site. When RFXAP is bound to the oligomerization domain of RFX5, the affinity for RFXB increases dramatically. We have shown that the interaction between the oligomerization domain of RFX5 and RFXAP promotes folding of residues 215 to 272 of RFXAP. This may promote formation of the RFXB binding site, and thus explain why the affinity is increased for RFXB. The ability of the oligomerization domain of RFX5 to induce the folding of RFXAP suggests that it contacts a significant region of RFXAP, which would explain the stability of the $RFX5_2$•RFXAP complex. The unfolded state of residues 215 to 272 of RFXAP may also explain why it is extremely susceptible to proteolytic cleavage during cell lysis when not in a complex with RFX5. The binding of RFX5 causes it to fold into a tertiary structure that could prevent access by proteases. It is interesting to speculate that this could be a method for regulation of the activity of RFXAP in the cell as a result of degradation of this key region of RFXAP.

The RFX complex is an attractive target for the development of immunosuppressant drugs as it is specific for MHCII genes and genes related to the function of MHCII molecules. Targeting the interaction between RFX5 and RFXAP is particularly attractive as they are proposed to assemble first in the cytoplasm and then translocate into the nucleus to form a complex with

RFXB [7]. The NMR observed induced folding of RFXAP by RFX5 could serve as a screening method to identify drugs that prevent association of RFX5 and RFXAP. After addition of the drug(s) to the NMR tube, an $^{15}N$-$^{1}H$ HSQC would be collected and the spectra analyzed to determine if RFXAP went from the folded state (Figure 6c) to the unfolded state (Figure 4c). If the unfolded state was induced, one or more of the drugs must be preventing assembly of the RFX5•RFXAP complex.

## Materials and Methods

### Construction of the RFX5, RFXAP, and RFXB expression vectors

PCR was preformed on vectors containing full length RFX5, RFXAP and RFXB [17]. The primers used to perform the PCR, together with the restriction enzymes that were used to cleave the resultant PCR product and vector, are summarized in Table 1. RFX5(1-90), RFX5 (168-330), RFX5(1-330), RFXAP, and RFXB were constructed as described previously [14]. The PCR products for RFX5(25-90) and RFX5(187-265) were inserted into the pProExHT vector (Invitrogen) to produce the pHTT-RFX5(25-90) and pHTT-RFX5(187-265) expression vectors, respectively. The pProExHT vector appends a 20 amino acid N-terminal affinity tag (HTT) onto the expressed protein. The affinity tag contains six histidines and a tobacco etch virus (TEV) protease cleavage site. The PCR products for RFX5(88-330) and RFXB(88-260) were inserted into the pHTCBD vector. This vector attaches an N-terminal affinity tag that contains six histidines, the 50 amino acid chitin binding domain (CBD), and a TEV protease cleavage site. The PCR product for RFXAP(215-272) was inserted into the pCBD vector, which is similar to the pHTCBD vector but does not contain a polyhistidine sequence, to give the vector pCBD-RFXAP(215-272). The primers used also introduced the point mutation F217Y, which was necessary to follow the protein spectroscopically. All vectors described were sequenced to ensure correct insertion of PCR products (Center for Biosystems Research at the University of Maryland College Park, MD).

### Site-directed mutagenesis of HTT-RFX5(25-90)

The pHTT-RFX5(25-90) expression vector was modified by site-directed mutagenesis to give a second vector, the pHTΔT-RFX5(25-90). In this vector the TEV protease recognition site, which codes for the amino acid sequence ENLYFQG, was mutated to code for ENHGFQG. Mutation of these two residues prevents TEV protease from cleaving the affinity tag from RFX5(25-90). The primers used to perform the site-directed mutagenesis are given in Table 1. The mutations were introduced using the QuikChange site-directed mutagenesis protocol (Stratagene). The mutation also introduced an NcoI restriction enzyme site, which was used to confirm the presence of the mutation during screening of colonies. Sequencing of the vectors confirmed the presence of the mutations.

### Construction of the polycistronic expression vectors

A polycistronic expression vector was constructed that expressed RFXAP(215-272,F217Y) with the CBD affinity tag and RFX5(25-90) with the HTΔT affinity tag. The vector was constructed using the pET3aTr/pST39 polycistronic expression system [22]. Briefly, the region of pHTΔT-RFX5(25-90) that codes for RFX5(25-90) and the HTΔT affinity tag was PCR amplified and inserted into the pET3aTr vector using the NdeI and BamHI restriction sites. Similarly, the region of pCBD-RFXAP(215-272,F217Y) that codes for RFXAP (215-272,F217Y) and the CBD affinity tag were PCR amplified and inserted into the pET3aTr vector using the NdeI and HindIII restriction sites. The pET3aTr-CBD-RFXAP (215-272,F217Y) vector was cut with XbaI and BamHI and the purified insert was ligated into the pST39 vector, which was cut with the same restriction enzymes. The pET3aTr-HTΔT-RFX5(25-90) insert was cut out with EcoRI and HindIII and ligated into the pST39-CBD-

RFXAP(215-272,F217Y) vector, which was cut with the same restriction enzymes, to give the final pCBD-RFXAP(215-272,F217Y)-HTΔT-RFX5(25-90) vector.

### Expression and purification of unlabeled RFX proteins

RFX5(1-90), RFX5(168-330), RFXB, RFXAP, and RFXAP(135-272) were expressed as described previously [14]. While RFX5(1-330) was purified in a similar manner as described previously, we found that stricter selection of the size exclusion peak corresponding to the dimeric form of RFX5(1-330) resulted in the majority of the stock of RFX5(1-330) being dimeric at 20 μM. This is in contrast to the stock used in our previous experiments, which contained a greater amount of tetramer at 20 μM. This suggested that some percentage of RFX5 (1-330) forms an irreversible tetramer, while the remainder exists in an equilibrium between the dimeric and tetrameric form, with the dimeric form predominating at low μM concentrations. RFX5(25-90), RFX5(187-265), and RFXB(88-260) were expressed and purified in the same manner as RFX5(1-90), RFX5(168-330), and RFXB, respectively. RFX5 (88-330) was expressed and purified in a similar manner to RFX5(1-330). Cleavage of the affinity tag from RFX5(88-330) required the use of excess TEV protease over a period of two days. RFXAP(215-272) was expressed using a modified approach to that used for full length RFXAP. Cells containing expressed CBD-RFXAP(215-272) were lysed with cells containing HTΔT-RFX5(25-90), which contains a non-TEV cleavable polyhistidine sequence. The complex was initially purified on a nickel-charged HisTrap column as described for RFXAP. TEV protease was added to the complex, resulting in the CBD being cleaved from RFXAP (215-272,F217Y). The CBD affinity tag was removed by passing the complex back over the nickel-charged HisTrap column. The two RFX proteins were separated by passing the complex over the nickel charged HisTrap column under denaturing conditions (6M urea). RFXAP (215-272) was refolded by dialyzing directly into 20 mM phosphate pH 7.0, 150 mM NaCl, and 1 mM EDTA overnight, and further purified using a Superdex 75 size exclusion column (GE Healthcare). All proteins were stored in 20 mM phosphate pH 7.0, 150 mM NaCl, 1mM EDTA, and 1 mM DTT at −20 °C until use.

### Expression and purification of $^{15}$N-labeled RFX proteins

To produce $^{15}$N-labeled RFX5(25-90) for NMR analysis two liters of BL21 Codonplus (DE3) RP cells containing the pHTT-RFX5(25-90) expression vector were grown at 37 °C in LB media containing 100 μg/mL ampicillin until the OD (600 nm) was between 0.5 – 0.6 absorbance units. The cells were spun down at 6000 rpm for 15 minutes, washed with 100 mLs of 1x minimal salt (M9) solution, centrifuged, and resuspended in a final volume of 50 mLs of 1x minimal media. The resuspended cells were transferred to 1 liter of 1x minimal media containing $^{15}$N-labeled ammonium chloride (Cambridge Isotopes Laboratory) and 100 μg/mL ampicillin. The cells were grown at 37 °C for 1 hour to allow clearance of unlabeled metabolites and growth recovery, before inducing with 1 mL of 1 M IPTG for 3 hours at 37 °C. The cells containing the expressed proteins were spun down at 6,000 rpm for 20 minutes and stored at −20 °C until use. The cell pellets were resuspended in 50 mM Tris pH 8.0, 500 mM NaCl, and 50 mM Imidazole, and lysed with a microfluidizer (Microfluidics Corp., Newton, MA). In contrast to the LB-derived RFX5(25-90), the minimal media-derived RFX5(25-90) was insoluble and had to be solubilized using 6 M urea. RFX5(25-90) was initially purified on a nickel charged HisTrap column under denaturing conditions and then refolded by reducing the urea concentration to 0 M over a period of 2 days. After refolding, the affinity tag was cleaved with TEV protease and removed by passing the protein over a nickel-charged HisTrap column. RFX5(25-90) was further purified over a Superdex 75 column. $^{15}$N-labeled RFXB(88-260) was expressed as described for RFX5(25-90), except the temperature was reduced to 30 °C before inducing with IPTG. Purification of $^{15}$N-labeled RFXB(88-260) was identical to the unlabeled protein. $^{15}$N-labeled RFXAP(215-272,F217Y) was produced using the *p*CBD-RFXAP(215-272,F217Y)-HTΔT-RFX5(25-90) polycistronic expression vector. The $^{15}$N-

labeled complex was expressed in an identical manner to RFX5(25-90). RFXAP(215-272) was extracted from the complex in an identical manner described for the LB-derived protein. Electrospray mass spectrometry was used to evaluate the percentage of $^{15}N$ incorporation into each of the protein. All of them were in the range 96 – 98%.

### Secondary structure prediction of RFX proteins

The secondary structure of each protein was predicted using three algorithms: PredictProtein [23], JPred [24], and PSIPRED [24]. The output of the three algorithms were aligned in order to compare the agreement between the different algorithms.

### SEC-SLS analysis of the proteins and complexes

Size exclusion chromatography (SEC) experiments were performed using an AKTA purifier (GE Healthcare) equipped with the following columns: 10/30 Superdex 75, 10/30 Superdex 200, and 10/30 Superose 6. The RFX5 + RFXAP complexes were combined in a 2:1 molar ratio of RFX5 and RFXAP, respectively. Similarly, the RFX5 + RFXB complexes were combined in a 2:1 molar ratio of RFX5 and RFXB, respectively. The RFX5 + RFXAP + RFXB complexes were combined in a 2:1:1 molar ratio of RFX5, RFXAP, and RFXB respectively. The RFX5(1-90) + RFXAP + RFXB + RFX5(88-330) complex was combined in a 2:1:1:2 complex ratio, respectively. Similarly, the RFX5(25-90) + RFXAP(215-272,F217Y) + RFXB (88-260) + RFX5(88-330) complex was combined in a 2:1:1:2 complex ratio, respectively. The proteins were incubated at 4°C for 30 min. 400 μL of protein was loaded onto the size exclusion column after filtering with a 0.2 μm centrifugal spin filter. Size exclusion was performed at 0.3 mL/min with a running buffer of 20 mM phosphate, pH 7.0, 150 mM NaCl, 1 mM DTT, and 1 mM EDTA. SDS-PAGE was used to analyze which proteins eluted in each peak. All the interaction studies were performed at a final protein/complex concentration of 5 μM and 20 μM. Size exclusion chromatography and static light scattering (SEC-SLS) experiments were performed in an identical manner described previously, using a miniDAWN tristar light scattering detector (Wyatt Technologies, Santa Barbara, CA) [14].

### Chemical crosslinking of RFX proteins and complexes

Cross-linking experiments were carried out using the zero-length crosslinking agent 1-ethyl-3-(3 diethylaminopropyl)carbodiimide hydrochloride (EDC) and N-hydroxy succinimide (NHS), as described previously [14]. Briefly, the protein was exposed to 20 mM EDC and 20 mM NHS at room temperature. Samples were taken out at 0, 20, 40, and 60 minutes and the crosslinking reaction was stopped prior to loading the samples onto an SDS-PAGE gel. Crosslinking of the individual proteins was performed with 10 μM, and 100 μM monomer protein concentrations. Crosslinking of the RFX5(25-90)$_2$•RFXAP(214-272) complex was performed with 25 μM complex concentration.

### Analysis of RFX proteins and complexes by c ircular dichroism spectra

All circular dichroism (CD) spectra were measured in the far-UV range between 190-250 nm using a Jasco spectropolarimeter, model J-715 (Jasco, Tokyo, Japan). The instrument was run in a continuous scan mode of 5 nm/min at a pitch of 0.1 nm and a response time of 4.0 seconds. The proteins were in 20 mM phosphate pH 7.4, and, 100 mM NaF. Scans were performed with 0.1 – 0.2 mg/mL protein concentrations and using a 1 mm path length. For each protein spectra the contribution of the solvent was subtracted by collecting a scan of the solvent in the absence of protein. Each spectrum shown represents the average of five independent scans. For the denaturation experiments, aliquots of each protein were transferred to solutions containing 1, 2, 4, and 6 M urea in 20 mM phosphate pH 7.0, and 100 mM NaF. The volume of the aliquots were chosen to cause minimal change in the final concentration of urea and to give 0.2 mg/mL protein. The scans were performed as described, but only collecting between 200 – 250 nm.

### Collection of 2D $^{15}$N-$^{1}$H HSQC NMR spectra

The 2D $^{15}$N–$^{1}$H NMR spectra for RFX5(25-90), RFXB(88-260), and RFXAP (215-272,F217Y) were collected on a Bruker 600 MHz spectrometer equipped with a 35 °C cryoprobe, using protein samples of 0.3 – 1 mM in 20 mM phosphate, pH 7.0, 150 mM NaCl, 1 mM EDTA, and 10% $D_2O$. NMR data were processed using NMRPipe [25] and analyzed using NMRView [26]. The RFXAP-RFX5 titration experiments were performed with a 500 μL solution of 10% $D_2O$, 0.42 mM $^{15}$N-labeled RFXAP(215-272,F217Y), and aliquots of 5.25 mM unlabeled RFX5(25-90) to give 1:0.25, 1:0.5, 1:1, 1:2, and 1:4 ratios of RFXAP to RFX5, respectively.

### Isothermal titration calorimetry

Isothermal titration calorimetry experiments were performed with a Microcal VP-ITC (MicroCal Inc. Northampton, MA). RFX5(25-90), RFXB(88-260), and RFXAP(215-272) were dialyzed into 10 mM phosphate pH 7.0, 150 mM NaCl, and 1 mM EDTA. All experiments were performed at 298 K. To examine the interaction between RFXB(88-260) and RFXAP (215-272), 10 μL injections of 300 μM RFXB(88-260) were titrated into 30 μM RFXAP (215-272). To examine the interaction between RFXB(88-260) and the RFX5(25-90)$_2$•RFXAP (215-272) complex, 10 μL injections of 200 μM RFXB(88-260) were titrated into 20 μM RFX5 (25-90)$_2$•RFXAP(215-272). The heat of dilution of RFXB(88-260) injected into the buffer were subtracted from the binding data before calculating the equilibrium dissociation constants. The final values reported are the average of three independent experiments.

## Acknowledgements

The authors would like to thank Mike Summers, Jamil Saad, Victoria D'Souza, and Robert Edwards for help in collecting and processing NMR data. The authors would also like to thank David Shaw King for mass spectrometry analysis of the labeled proteins. This work was supported by internal grants from the University of Maryland Baltimore County. Financial support for L.L. Briggs and K.M. Laird was provided by an NIGMS Initiative for Minority Student Development Grant (R25-GM55036).

## References

1. Ting JP, Trowsdale J. Genetic control of MHC class II expression. Cell 2002;109(Suppl):S21–33. [PubMed: 11983150]
2. Boss JM, Jensen PE. Transcriptional regulation of the MHC class II antigen presentation pathway. Curr Opin Immunol 2003;15:105–11. [PubMed: 12495741]
3. Durand B, Kobr M, Reith W, Mach B. Functional complementation of major histocompatibility complex class II regulatory mutants by the purified X-box-binding protein RFX. Mol Cell Biol 1994;14:6839–47. [PubMed: 7935401]
4. Hasegawa SL, Boss JM. Two B cell factors bind the HLA-DRA X box region and recognize different subsets of HLA class II promoters. Nucleic Acids Res 1991;19:6269–76. [PubMed: 1956787]
5. Jabrane-Ferrat N, Fontes JD, Boss JM, Peterlin BM. Complex architecture of major histocompatibility complex class II promoters: reiterated motifs and conserved protein-protein interactions. Mol Cell Biol 1996;16:4683–90. [PubMed: 8756625]
6. Reith W, Satola S, Sanchez CH, Amaldi I, Lisowska-Grospierre B, Griscelli C, Hadam MR, Mach B. Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RFX. Cell 1988;53:897–906. [PubMed: 3133120]
7. Nagarajan UM, Long AB, Harreman MT, Corbett AH, Boss JM. A hierarchy of nuclear localization signals governs the import of the regulatory factor X complex subunits and MHC class II expression. J Immunol 2004;173:410–9. [PubMed: 15210800]
8. Jabrane-Ferrat N, Nekrep N, Tosi G, Esserman L, Peterlin BM. MHC class II enhanceosome: how is the class II transactivator recruited to DNA-bound activators? Int Immunol 2003;15:467–75. [PubMed: 12663676]

9. Masternak K, Muhlethaler-Mottet A, Villard J, Peretti M, Reith W. Molecular genetics of the Bare lymphocyte syndrome. Rev Immunogenet 2000;2:267–82. [PubMed: 11258423]
10. Nekrep N, Geyer M, Jabrane-Ferrat N, Peterlin BM. Analysis of ankyrin repeats reveals how a single point mutation in RFXANK results in bare lymphocyte syndrome. Mol Cell Biol 2001;21:5566–76. [PubMed: 11463838]
11. Scholl T, Mahanta SK, Strominger JL. Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5. Proc Natl Acad Sci U S A 1997;94:6330–4. [PubMed: 9177217]
12. Zhu XS, Linhoff MW, Li G, Chin KC, Maity SN, Ting JP. Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. Mol Cell Biol 2000;20:6051–61. [PubMed: 10913187]
13. Reith W, Mach B. The bare lymphocyte syndrome and the regulation of MHC expression. Annu Rev Immunol 2001;19:331–73. [PubMed: 11244040]
14. Garvie CW, Stagno JR, Reid S, Singh A, Harrington E, Boss JM. Characterization of the RFX complex and the RFX5(L66A) mutant: implications for the regulation of MHC class II gene expression. Biochemistry 2007;46:1597–611. [PubMed: 17279624]
15. Villard J, Peretti M, Masternak K, Barras E, Caretti G, Mantovani R, Reith W. A functionally essential domain of RFX5 mediates activation of major histocompatibility complex class II promoters by promoting cooperative binding between RFX and NF-Y. Mol Cell Biol 2000;20:3364–76. [PubMed: 10779326]
16. Peretti M, Villard J, Barras E, Zufferey M, Reith W. Expression of the three human major histocompatibility complex class II isotypes exhibits a differential dependence on the transcription factor RFXAP. Mol Cell Biol 2001;21:5699–709. [PubMed: 11486010]
17. DeSandro AM, Nagarajan UM, Boss JM. Associations and interactions between bare lymphocyte syndrome factors. Mol Cell Biol 2000;20:6587–99. [PubMed: 10938133]
18. Wiszniewski W, Fondaneche MC, Louise-Plence P, Prochnicka-Chalufour A, Selz F, Picard C, Le Deist F, Eliaou JF, Fischer A, Lisowska-Grospierre B. Novel mutations in the RFXANK gene: RFX complex containing in-vitro-generated RFXANK mutant binds the promoter without transactivating MHC II. Immunogenetics 2003;54:747–55. [PubMed: 12618906]
19. Fry DC. Protein-protein interactions as targets for small molecule drug discovery. Biopolymers 2006;84:535–52. [PubMed: 17009316]
20. Zhao L, Chmielewski J. Inhibiting protein-protein interactions using designed molecules. Curr Opin Struct Biol 2005;15:31–4. [PubMed: 15718130]
21. Klages J, Coles M, Kessler H. NMR-based screening: a powerful tool in fragment-based drug discovery. Analyst 2007;132:693–705. [PubMed: 17657909]
22. Tan S. A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. Protein Expr Purif 2001;21:224–34. [PubMed: 11162410]
23. Rost B, Yachdav G, Liu J. The PredictProtein server. Nucleic Acids Res 2004;32:W321–6. [PubMed: 15215403]
24. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. Bioinformatics 1998;14:892–3. [PubMed: 9927721]
25. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–93. [PubMed: 8520220]
26. Johnson BA, Blevins RA. A computer program for the visualization and analysis for NMR data. J. Biomol. NMR 1994;4:603–614.

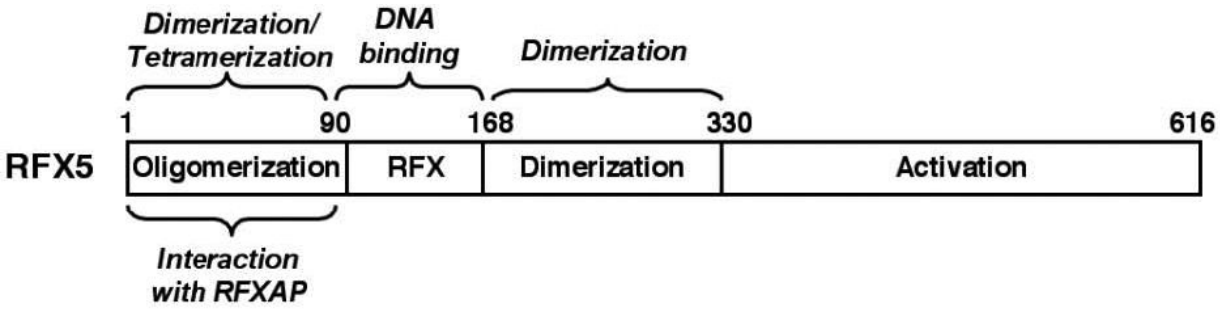


**Figure 1.**
Cartoon summarizing the key domains of RFX5 and their function.

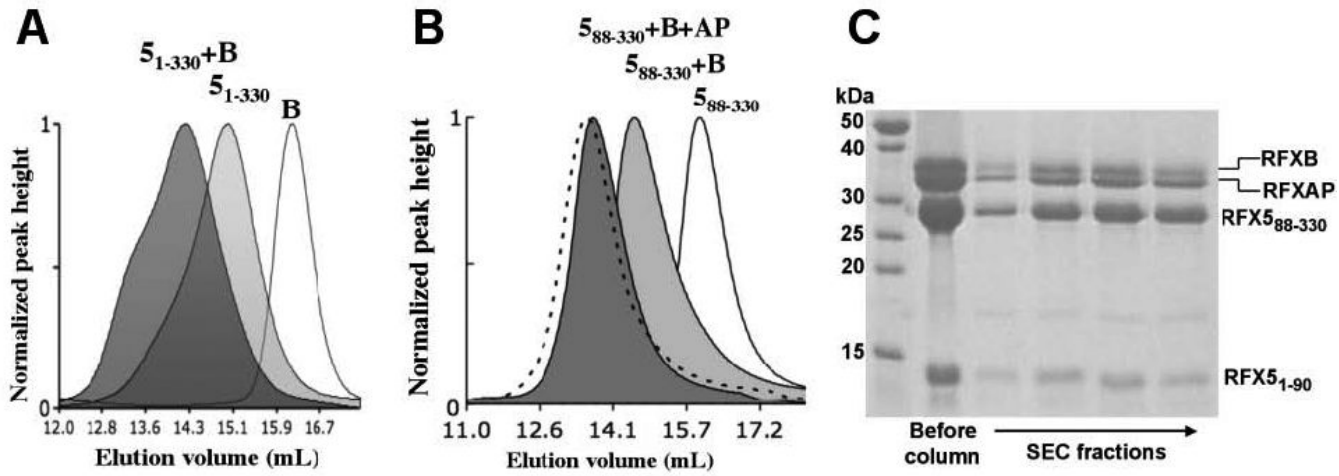


**Figure 2.**

Interactions between the RFX5 and RFXB. (A) Summary of size exclusion chromatograms of 20 μM RFXB (white), $RFX5(1-330)_2$ (light gray), and RFXB + $RFX5(1-330)_2$ (dark gray). The identity of each chromatogram is indicated above the peaks. The peak heights have been normalized for ease of comparison of the elution volumes. (B) Summary of size exclusion chromatograms of 20 μM $RFX5(88-330)_2$ (white), $RFX5(88-330)_2$ + RFXB (light gray) , $RFX5(88-330)_2$ + RFXB + RFXAP (dark gray), and $RFX5(88-330)_2$ + RFXB + RFXAP + $RFX5(1-90)_2$ (dashed line). (C) Coomassie stained SDS-PAGE gel of the $RFX5(88-330)_2$ + RFXB + RFXAP + $RFX5(1-90)_2$ complex before size exclusion, and of the peak that eluted from the size exclusion column. The identity of the bands is indicated on the right of the gel.

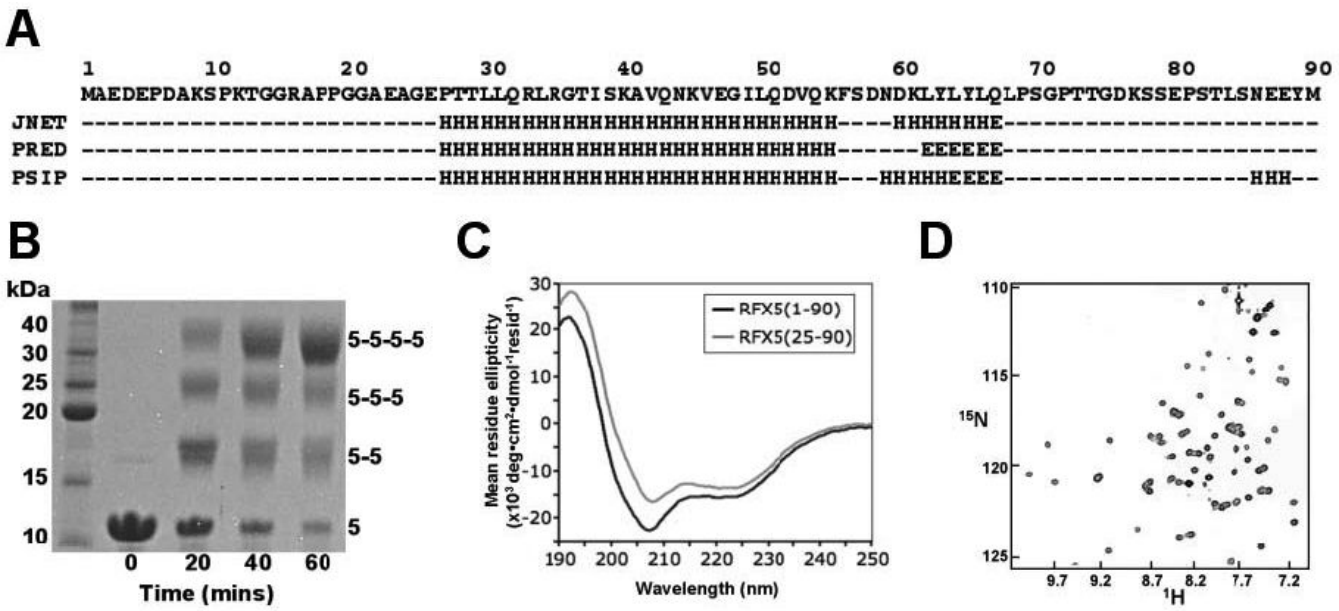


**Figure 3.**

Biophysical analysis of the oligomerization domain of RFX5. (A) Alignment of secondary structure prediction of RFX5(1-90) from three algorithms: JPred (JNET), PredictProtein (PRED), and PSIPRED (PRED). In the alignment, "H" refers to helix, "E" to a β-strand, and "–" to random coil. (B) Coomassie-stained SDS-PAGE gel showing time course crosslinking of 100 μM RFX5(25-90) by EDC. The probably identity of the crosslinked species is indicated on the right of the gel. (C) Far-UV CD spectra of RFX5(1-90) (black) and RFX5(25-90) (gray). (D) $^{15}$N-$^{1}$H HSQC NMR spectrum of RFX5(25-90).

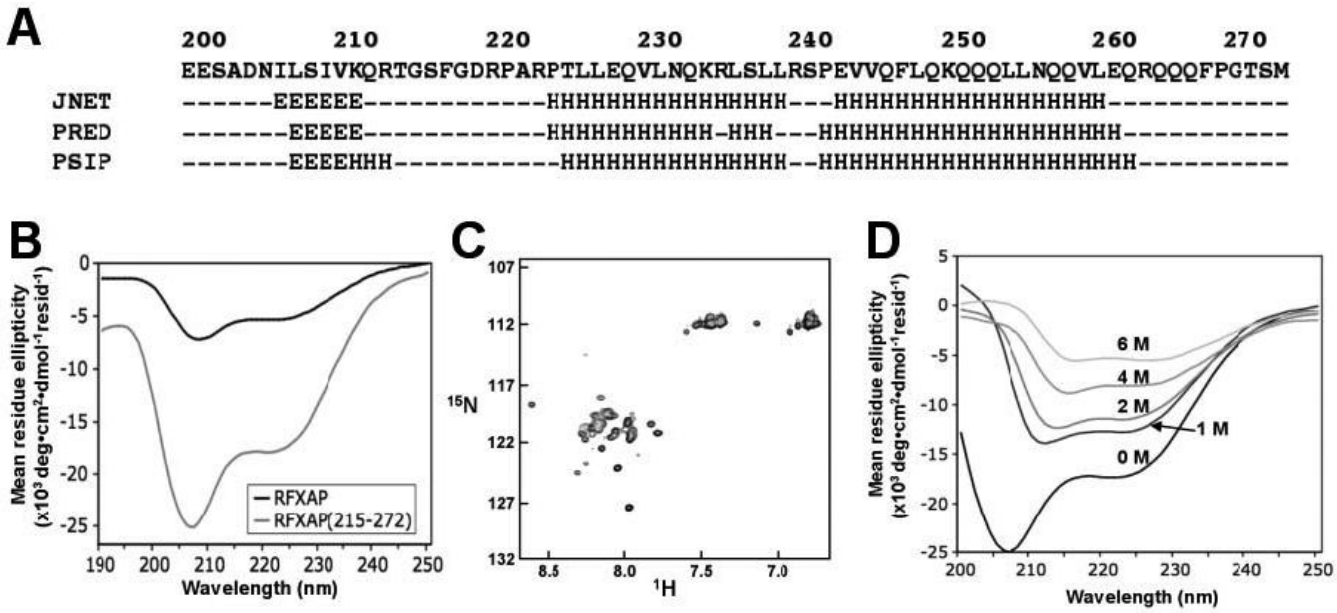


**Figure 4.**

Biophysical analysis of RFXAP(215-272). (A) Alignment of secondary structure prediction of RFXAP(214-272) as described in Figure 3. (B) Far-UV CD spectra of RFXAP (black) and RFXAP(215-272) (gray). (C) $^{15}N$-$^{1}H$ HSQC NMR spectrum of RFXAP(215-272). (D) Far-UV CD spectra of RFXAP(215-272) in 0 M, 1 M, 2 M, 4 M, and 6 M urea. The concentrations of urea are indicated above each spectrum.

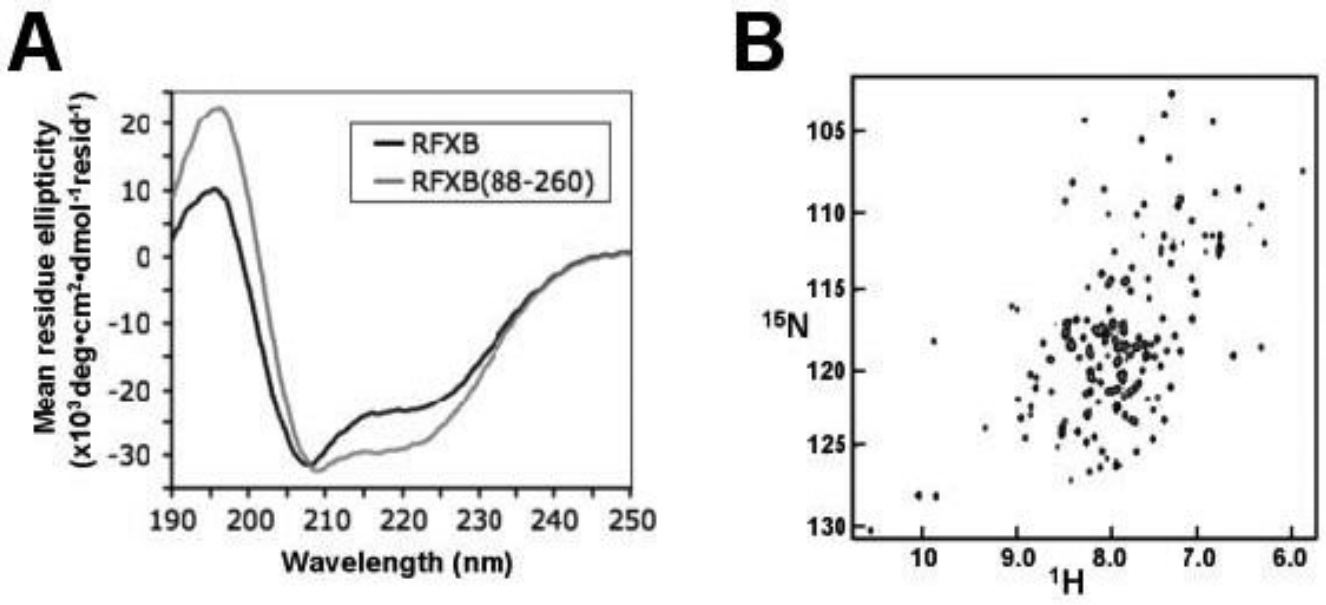


**Figure 5.**

Biophysical analysis of the ankyrin repeat region of RFXB. (A) far-UV CD spectra of RFXB (black) and RFXB(88-260) (gray). (B) $^{15}$N-$^{1}$H HSQC NMR spectrum of RFXB(88-260).

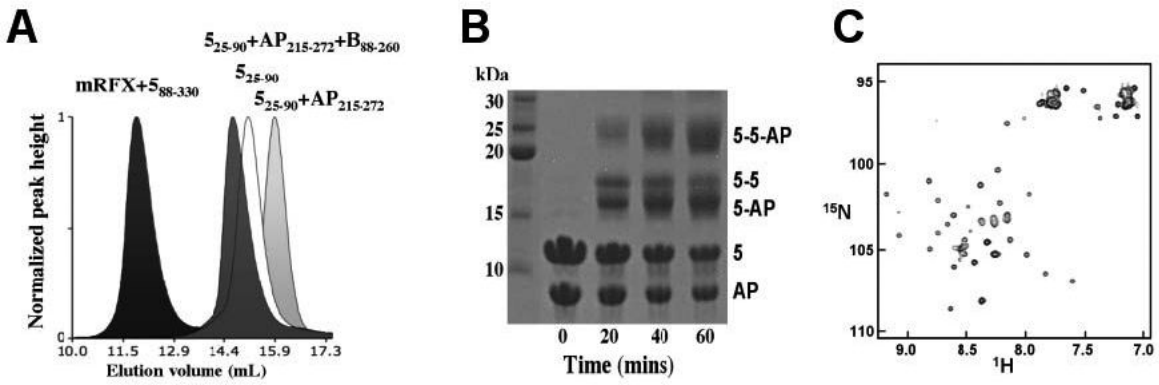


**Figure 6.**

Analysis of complex formation of the truncated RFX proteins. (A) Summary of size exclusion chromatograms of 20 μM RFX5(25-90) (white), RFX5(25-90) + RFXAP(215-272) (light gray), RFX5(25-90) + RFXAP(215-272) + RFXB(88-260) (dark gray), and RFX5(25-90) + RFXAP(215-272) + RFXB(88-260) +RFX5(88-330) (black). The identity of the peaks is indicated above each peak. For clarity, the RFX5(25-90) + RFXAP(215-272) + RFXB(88-260) complex is denoted as "mRFX" for minimal RFX complex. (B) Coomassie-stained SDS-PAGE gel showing time course crosslinking of 25 μM RFX5(25-90) + RFXAP(215-272) by EDC. The probably identity of the crosslinked species is indicated on the right of the gel. (C) $^{15}$N-$^{1}$H HSQC NMR spectrum of $^{15}$N-labeled RFXAP(215-272) titrated with unlabeled RFX5(25-90). The spectrum shown represents titration of a 1:2 ratio of RFXAP:RFX5.

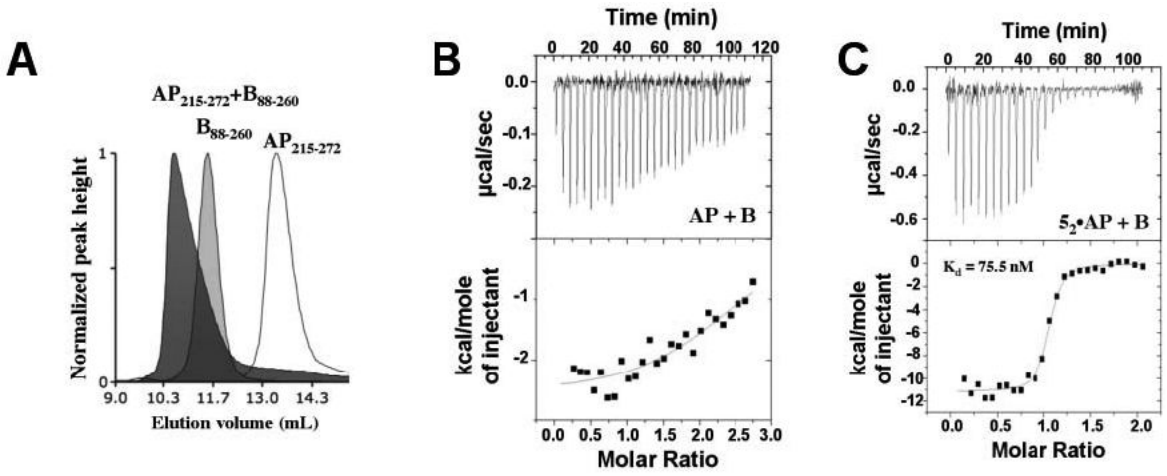


**Figure 7.**

Binding assays for the assembly of the minimal RFX complex. (A) Summary of size exclusion chromatograms of 20 μM RFXAP(215-272) (white), RFXB(88-260) (light gray), and RFXAP (215-272) + RFXB(88-260) (dark gray). (B) Isothermal titration calorimetry (ITC) analysis of the interaction between RFXB(88-260) titrated into RFXAP(214-272). The raw data is shown in the top diagram, while the curve fitting to the data is shown in the lower diagram. Accurate curve fitting could not be obtained for this data (see text). (C) ITC analysis of the interaction between RFXB(88-260) titrated into RFX5(25-90)$_2$•RFXAP(215-272). The top diagram shows the raw data, while the curve fitting to the data is shown in the lower diagram. The data fit well to a single binding site model with a N-value of 1.02 and a $K_d$ of 75.5 ± 5.5 nM.

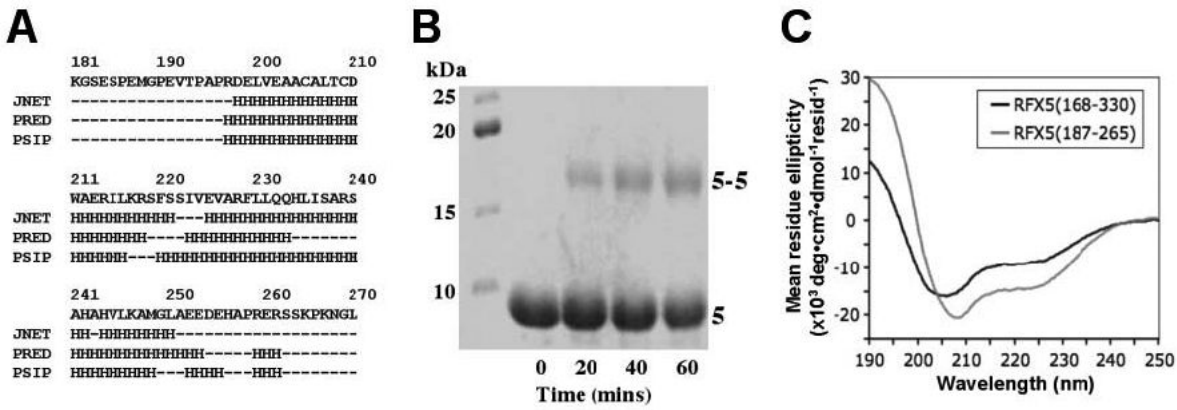


**Figure 8.**

Analysis of the dimerization domain of RFX5. (A) Alignment of secondary structure prediction of RFXAP(214-272) as described in Figure 3. (B) Coomassie-stained SDS-PAGE gel showing time course crosslinking of 100 μM RFX5(187-265) by EDC. The probably identity of the crosslinked species is indicated on the right of the gel. (C) Far-UV CD spectra of RFX5 (168-330) (black) and RFX5(187-265) (gray).

**Table 1**

Primers used to construct expression vectors.

| Primer name | Sense/anti-sense | Primer sequence (5′-3′) |
|---|---|---|
| RFXB-88-NCOI | Sense | TACAGTCCATGGAACCCTAGACTCCCTGTCC |
| RFXB-260-BAMHI | Anti-sense | GGCACGGAGGGATCCTCACTCAGGGTCAGCGGGCACCAG |
| RFXAP-215 -NCOI | Sense | AGTCACCATGGATCTTATGGGGATCGTCCTGCAAG |
| RFXAP-272-BAMHI | Anti-sense | AGTGACGGATCCTCACATTGATGTTCCTGGAAACTG |
| RFX5-88-NCOI | Sense | CAGTTACCATGGAGAGTACATGTATGCCTAT |
| RFX5-330-HINDIII | Anti-sense | CACTGGAAAGCTTTTAAGGCAGCCGAGCCACTAGGGC |
| RFX5-187-SFOI | Sense | GGCACTGGAGGCGCCGGGGAACCTACCACCCTTCTT |
| RFX5-265-HINDIII | Anti-sense | CACTGGAAAGCTTTTATTTAGATGACCGTTCCCGAGG |
| RFX5-SDMTEV-1 | Sense | CCAACGACCGAAAACCATGGTTTTCAGGGCGCCGGG |
| RFX5-SDMTEV-2 | Anti-sense | CCCGGCGCCCTGAAAACCATGGTTTTCGGTCGTTGG |