

# Structures of apo IRF-3 and IRF-7 DNA binding domains: effect of loop L1 on DNA binding

Pablo De Ioannes<sup>1</sup>, Carlos R. Escalante<sup>2,\*</sup> and Aneel K. Aggarwal<sup>1,\*</sup>

<sup>1</sup>Department of Structural and Chemical Biology, Mount Sinai School of Medicine, Box 1677, 1425 Madison Avenue, NY 10029 and <sup>2</sup>Department of Physiology and Biophysics, Virginia Commonwealth University, School of Medicine, 1101 East Marshall Street, Box 980551, Richmond, VA 23298, USA

Received February 18, 2011; Revised April 20, 2011; Accepted April 21, 2011

## ABSTRACT

Interferon regulatory factors IRF-3 and IRF-7 are transcription factors essential in the activation of interferon- $\beta$  (IFN- $\beta$ ) gene in response to viral infections. Although, both proteins recognize the same consensus IRF binding site AANNGAAA, they have distinct DNA binding preferences for sites *in vivo*. The X-ray structures of IRF-3 and IRF-7 DNA binding domains (DBDs) bound to IFN- $\beta$  promoter elements revealed flexibility in the loops (L1–L3) and the residues that make contacts with the target sequence. To characterize the conformational changes that occur on DNA binding and how they differ between IRF family members, we have solved the X-ray structures of IRF-3 and IRF-7 DBDs in the absence of DNA. We found that loop L1, carrying the conserved histidine that interacts with the DNA minor groove, is disordered in apo IRF-3 but is ordered in apo IRF-7. This is reflected in differences in DNA binding affinities when the conserved histidine in loop L1 is mutated to alanine in the two proteins. The stability of loop L1 in IRF-7 derives from a unique combination of hydrophobic residues that pack against the protein core. Together, our data show that differences in flexibility of loop L1 are an important determinant of differential IRF-DNA binding.

## INTRODUCTION

The interferon regulatory factor (IRF) family of proteins plays an essential role in the activation and regulation of immune response genes implicated in both innate and adaptive immunity (1–5). In addition, some members of this family have critical roles in the differentiation and development of hematopoietic cells and in the regulation

of apoptosis (6,7). In mammals, nine IRF proteins have been identified so far and all share the same general architecture, with a highly conserved N-terminal domain of approximately 120 residues, which is involved in DNA-specific binding and a variable C-terminal domain or IRF association domain (IAD) that mediates not only homo or hetero-oligomerization among IRF factors, but also mediates association with other transcription factors and co-activators, like CBP/p300 (8,9). Despite the high degree of structural similarity among the different IRF DNA binding domains (DBDs), there are important differences; for instance, IRF-4 contains an N-terminal extension of 20 amino acids that inhibit DNA binding (10) and IRF-7 has a recognition helix  $\alpha$ 3 that is longer than other members of the IRF family (11). Nevertheless, all IRF proteins recognize a DNA element with a consensus sequence AANNGAAA. This sequence is found in a multitude of promoters from interferon response genes, usually containing two or more of these elements (12–14). The crystal structure of IRF-1 DBD bound to DNA (15), revealed that the IRF DBD consists of a modified version of the helix-turn-helix (HTH) motif that includes a four-stranded antiparallel  $\beta$ -sheet and three large loops (L1–L3) connecting the different secondary structural elements. Subsequent structures of IRF-2 (16), IRF-4 (10), IRF-3 (17) and IRF-7 (11) bound to DNA have confirmed that DNA recognition is achieved by (i) conserved residues on helix  $\alpha$ 3 (Arg, Cys, Asn) interacting with GAAA bases (AANNGAAA) in the major groove and (ii) a conserved His on loop L1 protruding into the minor groove and making water-mediated contacts with two consecutive A:T steps upstream of the GAAA core sequence (AANNGAAA). These sets of interactions are further stabilized through a network of hydrogen bonds between the IRF proteins and the DNA phosphate backbone. The recent structures of IRF-3 DBD bound to the PRD1–PRDIII DNA element of the IFN- $\beta$  enhancer and the structure of IRF-3, IRF-7 and NF- $\kappa$ B DBDs bound to the

\*To whom correspondence should be addressed. Tel: +1 804 628 1202; Fax: +1 804 828 6991; Email: cescalante@vcu.edu  
Correspondence may also be addressed to Aneel K. Aggarwal. Tel: +1 212 659 8650; Fax: +1 212 849 2456; Email: aneel.aggarwal@mssm.edu

PRDII–PRDIII element have provided insights on the basis of cooperativity when two or more IRF DBDs bind to natural enhancers and promoters (11,18). From the structures, due to the overlapping nature of the individual sites, the binding of one IRF molecule affects the binding of the second molecule primarily through DNA bending rather than direct contacts with each other. Also, IRF-3 appears to be capable of recognizing both consensus and non-consensus sites and IRF-7 can better accommodate sites with G:C base pairs upstream of the AA element. In order to fully understand the structural and thermodynamic determinants of IRF DNA recognition, it is important to characterize the structural changes that occur upon DNA binding. To address this issue, we have solved the structures of IRF-3 and IRF-7 DBDs in the absence of DNA. Our data shows that differences in the flexibility of loop L1 in IRF proteins play an important role in DNA-specific binding.

## MATERIALS AND METHODS

### Protein purification

The cDNA fragments encoding human IRF-3 DBD (residues 1–113) and mouse IRF-7 DBD (residues 1–134) were generated by PCR and cloned in pET-15b vectors (Novagene). All point mutations were introduced by the Quick-change site-directed mutagenesis kit (Stratagene). Proteins were purified as previously described (18). Briefly, the proteins were expressed in the *Escherichia coli* strain BL21(DE3) pLyS (Stratagene). Cells were grown in LB medium supplemented with ampicillin (100 µg/ml) at 37°C to OD<sub>600</sub> 0.5 and induced with 0.4 mM IPTG for 3 h. Clarified lysates were applied to nickel affinity columns (Qiagen) obtaining protein up to 90% pure as judged by SDS–PAGE. The His-6 affinity tags were removed by thrombin digestion (Sigma). The proteins were then loaded onto a SP-Sepharose column (GE healthcare) equilibrated with 25 mM HEPES pH 7.0, 50 mM NaCl, 1 mM DTT and eluted using a linear gradient of 0–1.0 M NaCl in buffer. Further purification was carried out using a SD75 gel filtration column (GE healthcare) equilibrated with a buffer containing 25 mM Tris–HCl pH 7.5, 300 mM NaCl and 1 mM TCEP. The purified proteins were concentrated and used for crystallization.

### Crystallization and data collection

Selenomethionine (Semet) crystals of IRF-3 DBD were obtained by mixing equal volumes of protein solution (30 mg/ml) and reservoir solution containing 6% PEG 1000, 100 mM sodium acetate (pH = 6.0), 300 mM zinc chloride and 8% cadaverine. Crystals grew at 20°C within 2 days and belonged to space group P3<sub>2</sub>21 with unit cell dimensions of  $a = b = 64.90$  Å,  $c = 157.62$  Å. Multiwavelength anomalous diffraction data (MAD) were measured from a single frozen crystal at the Advance Photon Source (APS, beamline 17ID) of Argonne National laboratory. Data were measured at three wavelengths, corresponding to the inflection (0.9796 Å) and peak (0.9791 Å) of the selenium edge absorption profile plus at remote point (0.9537 Å). IRF-3

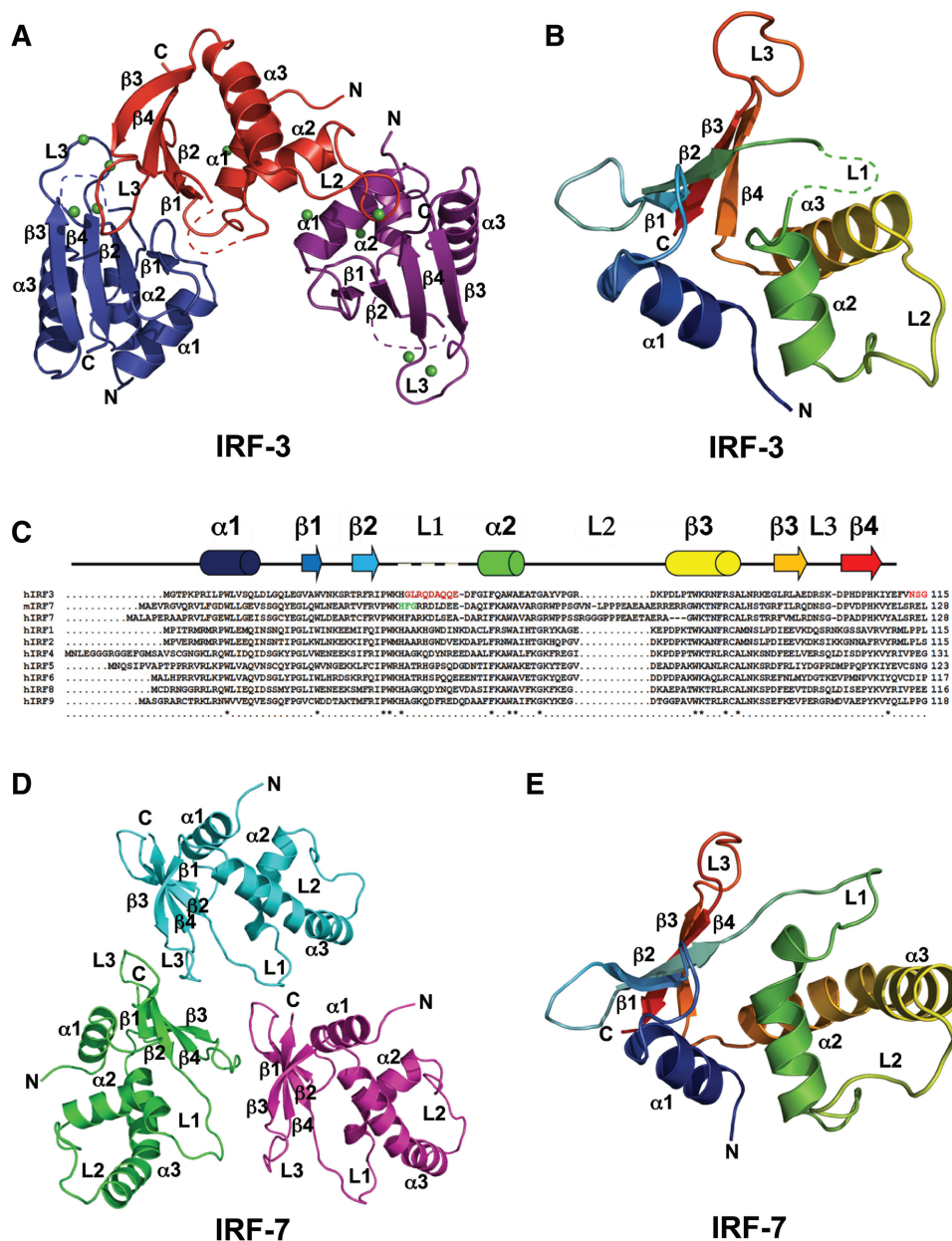
crystals diffracted to 2.1 Å resolution at APS. The data were truncated to 2.3 Å, due to the presence of an ice ring in the 2.1–2.25 Å resolution range.

IRF-7 DBD apo crystals were obtained mixing two volumes of protein solution (46 mg/ml) with one volume of reservoir solution containing 18% PEG 1500. The crystals grew over night and belonged to the trigonal space group P3<sub>1</sub> with unit cell dimension of  $a = b = 68.59$  Å,  $c = 68.61$  Å. Crystals were flash frozen in liquid nitrogen using mother liquor plus 20% glycerol as a cryoprotectant. Diffraction data were collected at National Synchrotron Light Source (NSLS, beam lines X6A and X29) at  $\lambda = 1.0$  Å. The crystals diffracted to 1.3 Å resolution. The data were integrated and reduced using HKL2000 package (19).

### Structure determination

The structure of IRF-3 DBD was solved by molecular replacement (MR) using one of the IRF monomers from the IRF-3/PRDIII–PRDI complex (PDB id 2PI0) (18) as a search model. The initial MR solution was obtained using the program PHASER (20) and included only two molecules in the asymmetric unit. The inspection of density map calculated with this solution suggested the presence of another IRF-3 molecule in the asymmetric unit, which was found by the program MolRep (21). An anomalous difference Fourier map was used to locate the positions of selenium (and zinc) atoms. Structure refinement was carried out by rounds of energy minimization and B-factor refinement using the CNS package (22). Manual model rebuilding was performed with the program COOT (23) using composite omit maps generated by CNS. Final cycles of restrained refinement were performed using the programs REFMAC (24) and PHENIX (25). The final model contains three monomers in the asymmetric unit, named IRF-3A (amino acids 5–40 and 49–110), IRF-3B (amino acids 5–39 and 49–110), IRF-3C (amino acids 5–39 and 49–110); 9 zinc ions; 3 sodium ions; 7 chloride ions and 151 water molecules (Figure 1A). The model presents excellent geometry, with the Ramachandran analysis indicating that 94.6% of the residues are in the allowed regions and 5.4% are in the preferred regions (Table 1).

The structure of mouse IRF-7 DBD apo form was solved by MR using a homology search model from the human IRF-7 DBD bound structure (PDB id 2O61) (11), generated by the program CHAINSAW (26). The initial MR solution was obtained with the program PHASER, which suggested three molecules in the asymmetric unit. After a round of rigid-body refinement, the  $R_{\text{fact}}$  and  $R_{\text{free}}$  values were 47 and 48%, respectively. A significant twin fraction ( $\alpha = 0.36$ ) was detected by the merohedral crystal twinning server (<http://nihserver.mbi.ucla.edu/twinning/>) and Xtriage program in the Phenix suite (25). Reflections for the free set were thus selected using the highest lattice symmetry [P3( $\bar{3}$ )m] in order to avoid a bias in the calculation of  $R_{\text{free}}$ , due to the pseudo-symmetry generated by the twinning. In the final step of refinement, twinning was included, with a significant improvement of the refinement statistics (final  $R_{\text{free}} = 20.0\%$ ).



**Figure 1.** Crystal structures of human IRF-3 DBD and mouse IRF-7 DBD. (A) Ribbon representation of three IRF-3 DBDs in the asymmetric unit. IRF-3A is colored red, IRF-3B is colored blue, and IRF-3C is colored purple. Zinc atoms are colored green. Loop L1 is represented in dashed lines. (B) Ribbon representation of IRF3-A. Chain was colored according to the secondary structure assignment in the alignment (see below). Missing residues in loop L1 are represented in dashed lines. (C) Structural based sequence alignment of IRF DBDs. Missing residues in the apo IRF-3 DBD structure are highlighted in red. The secondary structure elements in IRF-3 DBD are shown on top of the alignment and colored according to B. Missing region in loop L1 is shown in dashed lines. Unique residues in L1 of IRF-7 are highlighted in green. (D) Ribbon representation of three apo IRF-7 DBDs in the asymmetric unit. IRF-7A is colored green, IRF-7B is colored cyan, and IRF-7C is colored purple. (E) Ribbon representation of IRF-7A. The chain was colored according to the secondary structure assignment in the alignment.

Iterative rounds of restrained refinement and building were performed in the same manner as IRF3-DBD apo. The final model contains three monomers in the asymmetric unit, named IRF-7A (amino acids 9–130), IRF-7B (amino acids 9–125) and IRF-7C (amino acids 8–125); 287 water molecules and 3 sodium ions. The model has excellent geometry with 98.0% of the residues in the most favored regions of the Ramachandran plot, and no

residues in the disallowed regions (Table 1). All figures were made with the PyMol software (27).

#### Fluorescence anisotropy measurements

6-Carboxyfluorescein (6FAM) labeled (5'-6FAM-GAG<sub>A</sub>AGTGAAAGTGG-3') and unlabeled complementary DNA oligonucleotides containing the PRDI consensus

**Table 1.** Data collection and refinement statistics

Data collection	hIRF3-DBD	mIRF7-DBD
Space group	P3 <sub>2</sub> 21	P3 <sub>1</sub>
Molecules in AU	3	3
Cell parameters (Å)	$a = b = 64.91,$ $c = 157.62$	$a = b = 68.59,$ $c = 68.61$
Resolution (Å)	38.3–2.1	50.0–1.3
Total reflections	314 455	559 270
Unique reflections	25 580	85 725
Data coverage (%)	98.1 (99.7) <sup>a</sup>	96.7 (94.2) <sup>b</sup>
$R_{\text{sym}}$ (%)	7.6 (35.2) <sup>a</sup>	7.7 (41.8) <sup>b</sup>
Completeness (%)	97.97	96.64
$I/\sigma$	27.6 (6.7) <sup>a</sup>	21.3(3.1) <sup>b</sup>
Refinement statistics		
Resolution range	38.3–2.3	30–1.3
Reflections	17 524	85 702
$R_{\text{cryst}}$ (%)	21.2	16.2
$R_{\text{free}}$ (%)	26.4	20.0
Non-hydrogen atoms	2552	3234
Protein atoms	2382	2944
Solvent atoms	151	287
Metals atoms	19	3
Average B factors (Å <sup>2</sup> )	37.5	23.5
RMSD bonds (Å)	0.009	0.008
RMSD angles (°)	1.198	1.029
Ramachandran quality plot		
Most favored (%)	94.6	98.0
Allowed (%)	5.4	2.0
Disallowed (%)	0.0	0.0

<sup>a</sup>High-resolution shell (2.2–2.1 Å).

<sup>b</sup>High-resolution shell (1.35–1.30 Å).

$R_{\text{sym}} = 100 \times \sum (|I - \langle I \rangle|) / \sum |I|$ .

$R_{\text{fact}} = 100 \times \sum (|F_o| - |F_c|) / \sum (|F_o|)$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors, respectively ( $F_o > 0\sigma$ ).  $R_{\text{free}}$  was determined from ~5% of the data

site were purchased from IDT DNA Technologies (Coralville, IA, USA). Oligos were resuspended in 1 × TE buffer supplemented with 100 mM NaCl. DNA duplex were formed by heating to 95°C a mixture of one equivalent of the 6-FAM-labeled strand with one equivalent of the complementary strand and permitting the sample to cool down to room temperature. Each reaction sample (volume 150 µl) consisted of 5 nM of 5'-Fluorescein-labeled DNA and increasing concentrations of IRF-3 DBD or IRF-7 DBD (0.5 nM–300 µM) in a buffer containing 25 mM HEPES pH 7.0, 50 mM NaCl. Reactions were incubated at 20°C for 15 min prior to measurement. Fluorescence anisotropy intensity data were collected on a Beacon 2000 fluorescence polarization system (Invitrogen) by setting excitation filter at 490 nm and emission filter at 520 nm. The fraction of DNA bound (B) was calculated using the following equation:

$$B = \frac{([A]_X - [A]_{\text{DNA}})}{([A]_{\text{final}} - [A]_{\text{DNA}})}$$

where  $[A]_X$  equals the anisotropy measured at protein concentration X,  $[A]_{\text{DNA}}$  is the anisotropy in absence of protein and  $[A]_{\text{final}}$  is the anisotropy value at saturation. Each data point was calculated from an average of five anisotropy measurements. DNA fraction bound values were then plotted versus domain concentration and the

data were fitted by hyperbolic regression, using Origin 6.1 (OriginLab), to the following equation:

$$B = \frac{B_{\text{max}} * D_0}{K_d + D_0}$$

where  $B$  is the fraction of DNA duplex bound,  $D_0$  is the total concentration of the DBD, and  $K_d$  is the dissociation constant of the complex.

## RESULTS

### Overview

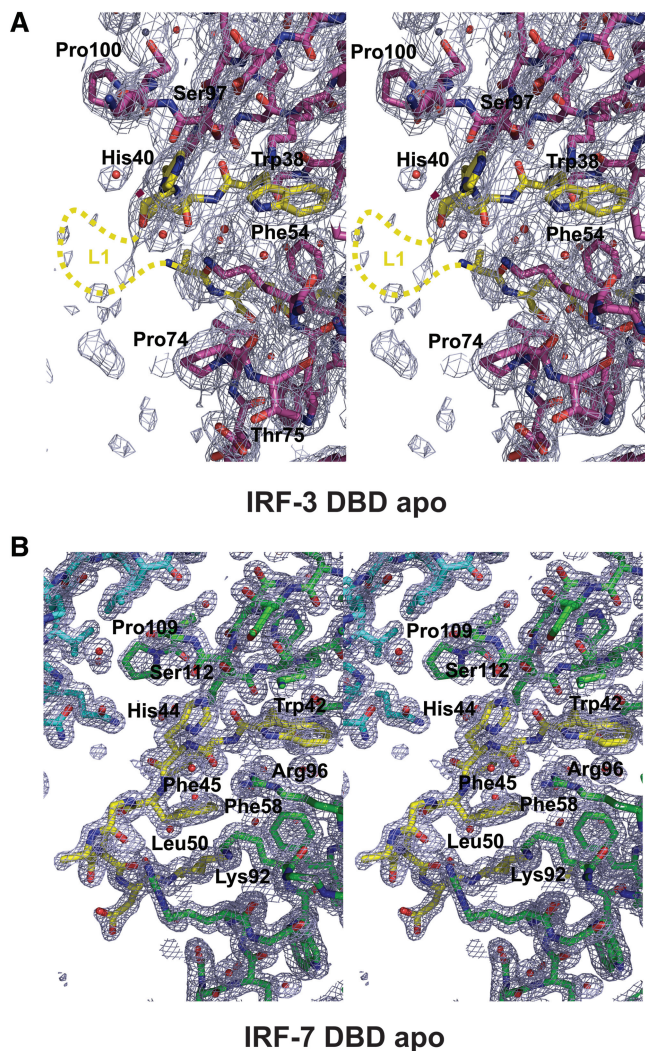
As in previous structures of IRF DBD/DNA complexes, the apo forms of IRF-3 and IRF-7 DBD's retain an  $\alpha/\beta$  architecture consisting of three  $\alpha$  helices ( $\alpha 1$ – $\alpha 3$ ) flanked by a four-stranded antiparallel  $\beta$ -sheet ( $\beta 1$ – $\beta 4$ ), a variant of the helix-turn-helix (HTH) motif found commonly in transcription factors (Figure 1B and E) (28). As is characteristic for this family, three large loops (L1–L3) connect different parts of the secondary structure. Loop L1 connects  $\beta 2$  and  $\alpha 2$ , loop L2 connects  $\alpha 2$  and  $\alpha 3$ , and loop L3 connects  $\beta 3$  and  $\beta 4$ . The electron density for apoIRF-3 DBD is well defined with the exception of 4 residues at the N-terminus and approximately 10 residues from loop L1. In the case of IRF-7 DBD, only the eight residues at the N-terminus are not defined. The apo IRF-3 and IRF-7 DBD structures superimpose with an RMSD of 1.39 Å, but if the superposition is performed with only the three  $\alpha$ -helices and the  $\beta$ -sheet, the RMSD decreases to 1.19 Å. The largest deviations between the two structures occur in loops L1 and L2.

### IRF-3 DBD apo form

The three IRF-3 DBD molecules in the crystallographic asymmetric unit are arranged with their loops (L1–L3) and helix  $\alpha 3$  exposed to the solvent while the rest of the molecules pack against each other. Interestingly, the apo IRF-3 DBD crystals grow only in the presence of  $\text{ZnCl}_2$ . We identified nine  $\text{Zn}^{2+}$  ions via their anomalous signal and together these ions mediate intermolecular interactions between the three symmetry-related molecules by coordinating a set of Asp, Glu and His residues (Figure 1A), as well as chloride ions. The coordination is typically tetrahedral with an average  $\text{Zn}^{2+}$ -ligand distance of 2.1 Å (29). A superposition of the three molecules yields a low RMSD value of 0.5 Å<sup>2</sup> for 98 C $\alpha$ 's. The largest deviation occurs in the linker between helix  $\alpha 3$  and strand  $\beta 3$  (RMSD of 1.47 Å) of molecule C and molecules A/B. A notable feature of the structure is the absence of density for most of loop L1 (amino acids Gly41–Asn48) in all three molecules of the asymmetric unit (Figure 1A).

In the previous IRF-3-DNA complex structures, loop L1 interacts with the DNA minor groove, with residues Trp38, Gly41, Leu42 and Gln44 contacting the DNA backbone and His40 making water-mediated contacts with the two contiguous A:T base pairs upstream of the GAAA core sequence (AANNNGAAA) (11,17,18). The loop is somewhat flexible in the DNA complexes (average B-factor of ~70 Å<sup>2</sup>) but, as we show here, in the absence

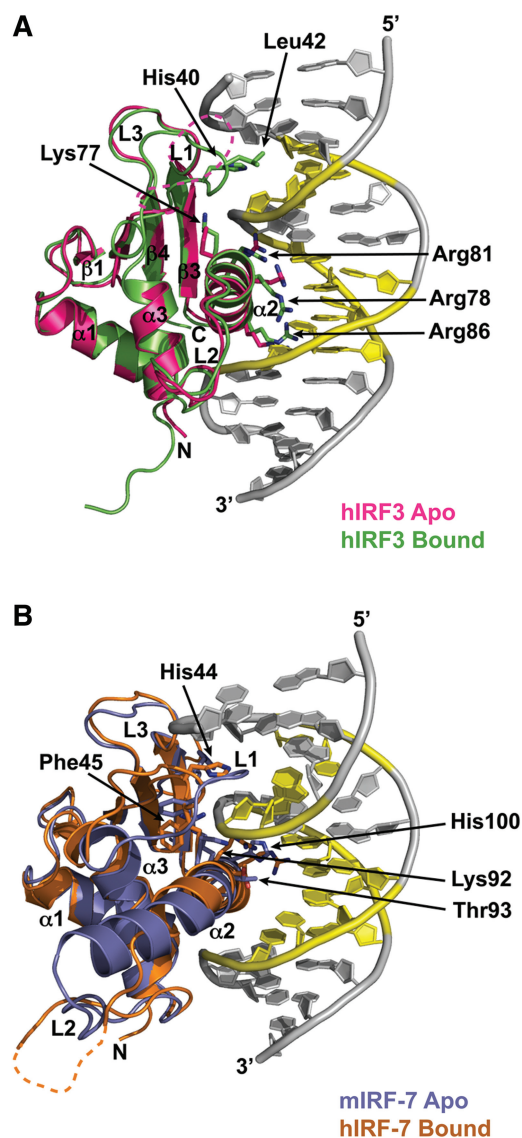
of DNA, it is largely disordered (Figures 1B and 2A). Also, residues that flank the loop deviate substantially in conformation from that observed in complex with DNA. For example, the C $\alpha$  of Phe51 is shifted by  $\sim 10$  Å and the C $\alpha$  of His40 is displaced by  $\sim 3.7$  Å. Also, side chain density of His40 is defined only in one of the three molecules in the asymmetric unit (Figure 2A). With the exception of loop L1 (and the flanking residues), the rest of the apo structure (including loops L2 and L3) superimposes well with the DNA complex, with an RMSD of  $\sim 1.06$  Å for 98 C $\alpha$ 's (Figure 3A). Most of the conformational changes are limited to the local rearrangement of side chains. Most notably, Lys77, Arg78, Arg81 and Arg86 (on helix  $\alpha$ 3) that make in contacts with the DNA in the complex adopt different (or multiple) conformations in the absence of DNA (Figure 3A) (18).



**Figure 2.** Structure of loop L1 in IRF-3 DBD and IRF-7 DBD. Stereo view of the  $2F_o - F_c$  electron density map for loop L1 region of IRF-3A DBD (A, red) and IRF-7A DBD (B, green), contoured at  $1\sigma$ . For clarity L1 is colored in yellow in both structures. IRF-7C DBD is colored cyan (B). Note that loop L1 in IRF-3 DBD lacks density form residues 41–48, whereas in IRF-7 DBD there is good definition for the density.

### IRF-7 DBD apo form

Although the IRF-7 DBD crystallizes in a different space group than the IRF-3 DBD there are again three molecules in the asymmetric unit (Figure 1B). The three molecules superimpose with a maximum RMSD of  $\sim 0.5$  Å for



**Figure 3.** Conformational changes in IRF-3 DBD and IRF-7 DBD upon DNA binding. (A) Structural comparison between apo and DNA-bound forms of IRF3-DBD. Apo IRF-3 DBD superimposes with an RMSD of 1.04 Å on the DNA bound form (PDB ID = 2P10 chain B). Note the disordered region in loop L1, represented by a dashed line in apo IRF-3 DBD. Side chains of His40 and Leu42 on loop L1 and Lys77, Arg78, Arg81 and Arg86 on Helix  $\alpha$ 3 are represented by sticks. DNA molecule from the complex is colored in grey, where the core recognition sequence (AANNAAAA) is highlighted in yellow. (B) Structural comparison between apo and DNA-bound forms of and IRF-7 DBD. Apo mouse IRF-7 DBD superimposes with an RMSD of 1.41 Å on the DNA-bound form of human IRF-7 DBD (PDB ID 2O61). Side chains of His44 and Phe45 on loop L1 and Lys92, Thr93 and His100 on helix  $\alpha$ 3 of mouse IRF-7 (and counterparts on human IRF-7) are represented by stick. Note that helix  $\alpha$ 2 is extended and all three loops (L1–L3) are ordered in apo IRF-7 DBD.

117 C $\alpha$ 's. In contrast to the apo IRF-3 DBD, loop L1 is well defined (Figure 1E and 2B). This ordering of loop L1 in IRF-7 does not appear to be a crystallization artifact, as the conformation of the loop is almost identical in all three molecules in the asymmetric unit despite their different crystal packing environments. Loop L1 in IRF-7 differs from that in other IRF family members in containing a phenylalanine (Phe45 instead of an alanine) after the conserved histidine (His44) (Figure 1C). Phe45 packs against Leu50 (also unique to IRF-7) in loop L1 and Phe58 in helix  $\alpha$ 2, as part of the hydrophobic core of the IRF-7 DBD (Figure 2B). Together, these hydrophobic interactions appear to confer a more restrictive conformation to loop L1, which is reflected in lower average B-factors for this region ( $\sim 25 \text{ \AA}^2$ ) compared with the same region in IRF-3.

Loop L2 is much longer in the IRF-7 DBD than in other IRFs and has been proposed to interact with the p50 subunit of NF- $\kappa$ B when the two transcription factors bind to adjacent sites on the interferon- $\beta$  enhancer (10). Because of its length and the abundance of proline and glycine residues, it is the most flexible segment of the IRF-7 DBD structure. Indeed, in the structure of IRF-7 bound to DNA, loop L2 is partially unstructured (11). In the apo structure, loop L2 is well defined but has the highest average B-factors ( $\sim 29 \text{ \AA}^2$ ) (Figure 3B). Interestingly, residues 71–78 in L2 are unique to IRF-7 and adopt an extended conformation characteristic of poly glycine and poly proline chains (Figure 3B). This extended conformation appears to be important in allowing Arg67 on loop L2 to contact Asn75 on NF- $\kappa$ B.

Intriguingly, the recognition helix  $\alpha$ 3 in the apo structure is longer at the N-terminus by three amino acids when compared to IRF-7 DBD in the presence of DNA (Figure 3B). Among these amino acids is Arg89, which in the DNA complex makes contacts with the DNA backbone but in the absence of DNA is directed toward the interior of the protein and makes contacts with structural water in the protein cavity. Interestingly, Arg89 in apo IRF-7 adopts a similar conformation as Lys92 in DNA complex, pointing to a switch in position between the two residues upon DNA binding. One feature of helix  $\alpha$ 3 that is conserved in the apo and DNA bound forms of IRF-7 is a kink in the helix at Gly90, indicating that this bending is not a consequence of the DNA binding.

We also identify a putative metal ion between helix  $\alpha$ 3 and strand  $\beta$ 3 in the apo structure that is coordinated with carboxylic groups of Leu99, Gly103 and Phe105. The metal (possibly Na<sup>+</sup> from the crystallization mix) appears to compensate for the large dipole moment of long helix  $\alpha$ 3. A similar metal binding site is located in the DNA bound structure of IRF-2 (16).

### Mutations in loop L1

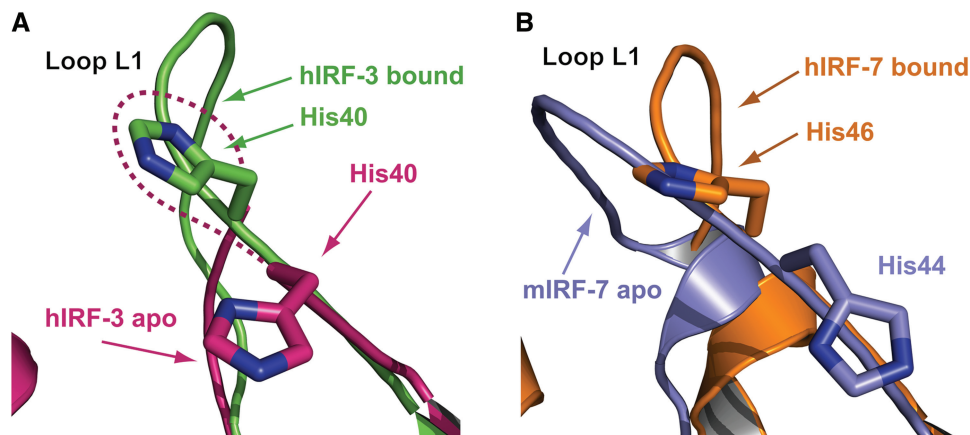
The structures of all IRF/DNA complexes have shown that a conserved histidine on loop L1 (His40 in IRF-3 and His44 in IRF-7) partially enters the DNA minor groove and makes water mediated contacts with A:T

base pairs upstream of the GAAA core sequence (AANNGAAA). The observed difference in flexibility of loop L1 in IRF-3 and IRF-7 DBDs raises the question of the effect of mutating residues in loop L1 on DNA binding. To address this, we constructed two mutants: (i) IRF3-H40A, in which His40 in IRF-3 DBD is mutated to alanine, (ii) IRF7-H44A, in which His44 in IRF-7 DBD is also mutated to alanine. The wild-type (WT) and mutant proteins were tested for DNA binding by fluorescence anisotropy. Figure 5 shows the result obtained using the PRDI DNA element with sequence GAGAAGTGAAAGT. Our experiments show that IRF-3 DBD<sub>wt</sub> binds with a dissociation constant of 486 nM, while IRF-7 DBD<sub>wt</sub> binds with a slightly lower affinity of 630 nM. These values are consistent with previously reported affinities of IRFs proteins for the PRD-I element (30). However, the affinity of IRF-3-H40A decreases by  $\sim 3.6$ -fold ( $K_d = 1706 \text{ nM}$ ), IRF-7-H44A binds with almost the same affinity ( $K_d = 532 \text{ nM}$ ) as the WT protein.

### DISCUSSION

We present here the first crystal structures of IRF-3 and IRF-7 DBDs in the absence of DNA. Superposition of apo and DNA bound IRF-3 DBDs reveals two primary conformational changes on DNA binding: (i) the N-terminus of IRF-3 undergoes a disorder to order transition, with Lys5 making contacts with DNA along the minor groove, (ii) loop L1 becomes ordered and approximately 10 residues that are not visible in the apo IRF3 structure are well defined in the DNA-bound structure. In contrast to IRF-3, the L1 loop is ordered in the apo IRF-7 DBD structure and the major change is a rearrangement of the loop as a rigid body around the minor groove. The inherent flexibility of loop L1 is consistent with the apo NMR structure of IRF-2 DBD, wherein L1 displays a large number of conformations (31). A conserved PWKH motif in loop L1 appears to be important in positioning L1 for its interaction with DNA. That is, the proline (Pro37 in IRF-3) fixes the adjoining tryptophan (Trp38) in an outward conformation for interaction with the DNA backbone. The indole ring of this tryptophan moves by about 1.6  $\text{\AA}$  in order to interact with a DNA phosphate group. At the same time, the main chain carbonyl group of Lys39 is positioned to make a hydrogen bond with Lys77 (on the recognition helix  $\alpha$ 3), providing stability to the overall loop conformation. Notably, the imidazole ring of His40 swings by almost 6  $\text{\AA}$  into the DNA minor groove in order to make water-mediated contacts with A:T base pairs, while the main chain carboxyl group of His40 interacts with the main chain amide group of Arg43 to further stabilize the loop (Figure 4A).

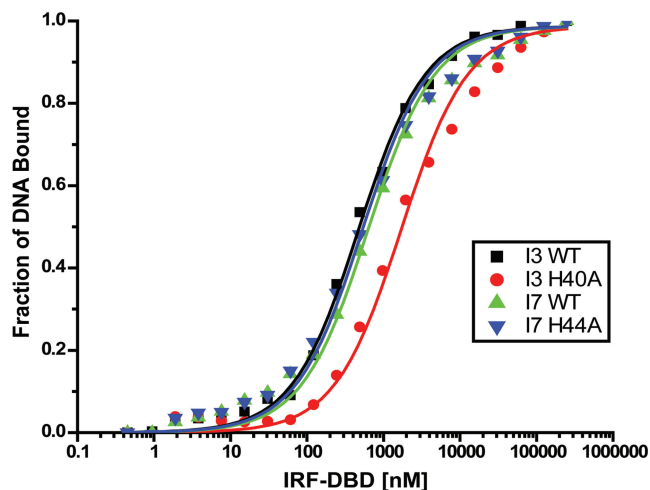
In contrast to IRF-3, the L1 loop is ordered in the apo IRF-7 DBD structure and stabilized in part by two hydrophobic residues (Phe45 and Leu50) that fold back into the core of the protein. Upon DNA binding, the loop undergoes a rigid body transition of  $\sim 2 \text{ \AA}$  that pivots around two points at both ends of the loop. At the



**Figure 4.** Structural comparison of Loop L1 between apo and DNA bound form of IRF DBDs. (A) Close up view of Loop L1 of IRF-3 DBD apo (magenta) and bound (green) forms. Conserved histidine residue (His40) is shown in stick representation. (B) Close up view of loop L1 of IRF-7 DBD apo (blue) and bound (orange) forms. Conserved histidines (His46 on human and His44 on mouse) are shown in stick representation. For clarity, DNA is not shown. Loop L1 and the conserved histidine adopt similar conformations when bound to DNA, but adopt different (IRF-7) or disordered (IRF-3) conformations in the absence of DNA.

N-terminal end, the C $\alpha$  of the conserved residue Trp42 moves by about 1 Å in order to interact with the DNA backbone. Moreover, the displacement of Trp42 pushes against the other pivot site in residue Phe60. Similarly to IRF-3 DBD, the imidazole ring of the conserved Histidine residue (His44) rotates around the chi1 bond by  $\sim 25^\circ$  to position itself to interact with the minor groove. The difference in the inherent flexibility of loop L1 between IRF-3 and IRF-7 seems to be reflected on the effect of His mutations on DNA binding. Thus, while mutation of His40 to alanine significantly decreases DNA binding of IRF-3 by  $\sim 3.6$ -fold, the equivalent mutation in IRF-7 shows no significant change in binding (Figure 5). The decrease in binding with the IRF-3 H40A mutation likely represents the loss of interactions that stabilize loop L1 upon DNA binding, including the interactions between His40 and the DNA minor groove. However, because loop L1 is already ordered in apo IRF-7 and His44 is involved in interactions with Ser112 and several water molecules, it is conceivable that there is less of a gain in enthalpy when His44 (compared to His40 in IRF-3) interacts with DNA. Our data suggest that differences in the inherent flexibility of loop L1 between IRF-3 and IRF-7 have a direct effect on DNA binding and may play a role in the distinct DNA binding specificities observed between the two proteins. Indeed, there is some indication from biochemical studies that IRF-3 and IRF-7 interact differently with DNA. IRF-7, for example, is more tolerant than IRF-3 to changes in the AA sequence upstream of the GAAA core (AANNGAAA) suggesting a somewhat looser interaction with DNA (9,31). To more fully explore the differences between IRF-3 and IRF-7 in DNA selection will require additional mutations and a more detailed thermodynamic analysis on several DNA sites.

Taken together, we show here that the structures apo IRF-3 and IRF-7 DBD's are generally similar to the DNA bound forms, including a kink in the recognition helix  $\alpha 3$



Domain	K <sub>d</sub> (nM)	K <sub>d</sub> Ratio (Mutant/WT)	R <sup>2</sup>
IRF-3 DBD WT	473 ± 21.3	3.6	0.99735
IRF-3 DBD H40A	1706 ± 119		0.99198
IRF-7 DBD WT	630 ± 46	0.844	0.99198
IRF-7 DBD H44A	532 ± 40		0.99163

**Figure 5.** Effect of mutations in loop L1 on DNA binding. Isothermal binding curves over the PRDI consensus DNA site for IRF-3 DBD wild type (black squares), IRF-3 DBD H40A (red dots), IRF-7 DBD wild type (green triangles) and IRF-7 DBD H44A (blue inverted triangles). The calculated K<sub>d</sub>'s are shown in the table.

and ordered loops L2 and L3. The primary difference is in loop L1, which undergoes a disorder to order transition in the IRF-3 DBD and a conformational change in the IRF-7 DBD. The varying intrinsic flexibility of loops and tails in the IRF family may serve as a mechanism to modulate the binding specificity of its members and

to respond to a larger population of diverse promoter sites.

## ACCESSION NUMBERS

Coordinates have been submitted to the RCSB Protein Data Bank with accession codes: 3QU6 for IRF-3 and 3QU3 for IRF-7

## ACKNOWLEDGEMENTS

We thank the staff at APS beamline 17ID and NSLS beamline X6A for help with data collection.

## FUNDING

US National Institutes of Health (R01 AI41706 to A.K.A. and R01 GM092854 to C.R.E.). Funding for open access charge: NIH.

*Conflict of interest statement.* None declared.

## REFERENCES

- Nguyen,H., Hiscott,J. and Pitha,P.M. (1997) The growing family of interferon regulatory factors. *Cytokine Growth Factor Rev.*, **8**, 293–312.
- Vaughan,P.S., van Wijnen,A.J., Stein,J.L. and Stein,G.S. (1997) Interferon regulatory factors: growth control and histone gene regulation—it's not just interferon anymore. *J. Mol. Med.*, **75**, 348–359.
- Mamane,Y., Heylbroeck,C., Genin,P., Algarte,M., Servant,M.J., LePage,C., DeLuca,C., Kwon,H., Lin,R. and Hiscott,J. (1999) Interferon regulatory factors: the next generation. *Gene*, **237**, 1–14.
- Ozato,K., Tailor,P. and Kubota,T. (2007) The interferon regulatory factor family in host defense: mechanism of action. *J. Biol. Chem.*, **282**, 20065–20069.
- Richez,C., Barnette,T., Miceli-Richard,C., Blanco,P., Moreau,J.F., Rifkin,I. and Schaeferbeke,T. (2010) Role for interferon regulatory factors in autoimmunity. *Joint Bone Spine*, **77**, 525–531.
- Battistini,A. (2009) Interferon regulatory factors in hematopoietic cell differentiation and immune regulation. *J. Interferon Cytokine Res.*, **29**, 765–780.
- Hiscott,J., Grandvaux,N., Sharma,S., Tenover,B.R., Servant,M.J. and Lin,R. (2003) Convergence of the NF-kappaB and interferon signaling pathways in the regulation of antiviral defense and apoptosis. *Ann. N. Y. Acad. Sci.*, **1010**, 237–248.
- Taniguchi,T., Ogasawara,K., Takaoka,A. and Tanaka,N. (2001) IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.*, **19**, 623–655.
- Lin,R., Genin,P., Mamane,Y. and Hiscott,J. (2000) Selective DNA binding and association with the CREB binding protein coactivator contribute to differential activation of alpha/beta interferon genes by interferon regulatory factors 3 and 7. *Mol. Cell. Biol.*, **20**, 6342–6353.
- Escalante,C.R., Brass,A.L., Pongubala,J.M., Shatova,E., Shen,L., Singh,H. and Aggarwal,A.K. (2002) Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol. Cell*, **10**, 1097–1105.
- Panne,D., Maniatis,T. and Harrison,S.C. (2007) An atomic model of the interferon-beta enhanceosome. *Cell*, **129**, 1111–1123.
- Goodbourn,S. and Maniatis,T. (1988) Overlapping positive and negative regulatory domains of the human beta-interferon gene. *Proc. Natl Acad. Sci. USA*, **85**, 1447–1451.
- Naf,D., Hardin,S.E. and Weissmann,C. (1991) Multimerization of AAGTGA and GAAAGT generates sequences that mediate virus inducibility by mimicking an interferon promoter element. *Proc. Natl Acad. Sci. USA*, **88**, 1369–1373.
- Schafer,S.L., Lin,R., Moore,P.A., Hiscott,J. and Pitha,P.M. (1998) Regulation of type I interferon gene expression by interferon regulatory factor-3. *J. Biol. Chem.*, **273**, 2714–2720.
- Escalante,C.R., Yie,J., Thanos,D. and Aggarwal,A.K. (1998) Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature*, **391**, 103–106.
- Fujii,Y., Shimizu,T., Kusumoto,M., Kyogoku,Y., Taniguchi,T. and Hakoshima,T. (1999) Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences. *EMBO J.*, **18**, 5028–5041.
- Panne,D., Maniatis,T. and Harrison,S.C. (2004) Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J.*, **23**, 4384–4393.
- Escalante,C.R., Nistal-Villan,E., Shen,L., Garcia-Sastre,A. and Aggarwal,A.K. (2007) Structure of IRF-3 bound to the PRDIII-I regulatory element of the human interferon-beta enhancer. *Mol. Cell*, **26**, 703–716.
- Otwinowski,Z. and Minor,W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.*, **276**, 307–326.
- McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
- Vagin,A.A. and Isupov,M.N. (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 1451–1456.
- Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
- Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
- Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
- Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
- Stein,N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.*, **41**, 641–643.
- The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.
- Harrison,S.C. and Aggarwal,A.K. (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.*, **59**, 933–969.
- Roe,R. and Pang,Y. (1999) Zinc's exclusive tetrahedral coordination governed by its electronic structure. *J. Mol. Model.*, **5**, 134–140.
- Hargreaves,V.V., Makeyeva,E.N., Dragan,A.I. and Privalov,P.L. (2005) Stability and DNA binding ability of the DNA binding domains of interferon regulatory factors 1 and 3. *Biochemistry*, **44**, 14202–14209.
- Furui,J., Uegaki,K., Yamazaki,T., Shirakawa,M., Swindells,M.B., Harada,H., Taniguchi,T. and Kyogoku,Y. (1998) Solution structure of the IRF-2 DNA-binding domain: a novel subgroup of the winged helix-turn-helix family. *Structure*, **6**, 491–500.