

The molecular basis for the development of adult T-cell leukemia/lymphoma in patients with an IRF4^{K59R} mutation

Srinivasan Sundararaj¹  | Sandali Seneviratne¹  | Simon J Williams²  | Anselm Enders^{3,4}  | Marco G. Casarotto¹ 

¹John Curtin School of Medical Research, Australian National University, Canberra, Australia

²Research School of Biology, Australian National University, Canberra, Australia

³Department of Immunology, John Curtin School of Medical Research, Australian National University, Canberra, Australia

⁴Center for Personalised Immunology, John Curtin School of Medical Research, Australian National University, Canberra, Australia

Correspondence

Srinivasan Sundararaj and Marco G. Casarotto, John Curtin School of Medical Research, Australian National University, Canberra, Australia.
Email: srinivasan.sundararaj@anu.edu.au (S. S.) and marco.casarotto@anu.edu.au (M. C.)

Funding information

Australian National University; National Health and Medical Research Council, Grant/Award Number: GNT1079648

Abstract

Interferon regulatory factor 4 (IRF4) is an essential regulator in the development of many immune cells, including B- and T-cells and has been implicated directly in numerous hematological malignancies, including adult T-cell leukemia/lymphoma (ATLL). Recently, an activating mutation in the DNA-binding domain of IRF4 (IRF4^{K59R}) was found as a recurrent somatic mutation in ATLL patients. However, it remains unknown how this mutation gives rise to the observed oncogenic effect. To understand the mode of IRF4^{K59R}-mediated gain of function in ATLL pathogenesis, we have determined the structural and affinity basis of IRF4^{K59R}/DNA homodimer complex using X-ray crystallography and surface plasmon resonance. Our study shows that arginine substitution (R59) results in the reorientation of the side chain, enabling the guanidium group to interact with the phosphate backbone of the DNA helix. This markedly contrasts with the IRF4^{WT} wherein the K59 interacts exclusively with DNA bases. Further, the arginine mutation causes enhanced DNA bending, enabling the IRF4^{K59R} to interact more robustly with known DNA targets, as evidenced by increased binding affinity of the protein–DNA complex. Together, we demonstrate how key structural features underpin the basis for this activating mutation, thereby providing a molecular rationale for IRF4^{K59R}-mediated ATLL development.

KEYWORDS

adult T-cell leukemia/lymphoma, crystal structure, interferon regulatory factor 4 (IRF4), protein–DNA complex, protein–DNA interaction, transcription factor

1 | INTRODUCTION

Adult T-cell leukemia/lymphoma (ATLL) is an aggressive, non-Hodgkin lymphoma characterized by frequent patient relapses and poor survival outcomes. ATLL develops as a result of infection with human T-cell leukemia virus type 1 (HTLV-1), with approximately 5% of

20 million patients infected with of HTLV-1 developing the disease.¹ The genomic integration of HTLV-1, characteristic of ATLL, is central to disease pathogenesis,² with HTLV-1 driving expression of the Tax *trans*-activator/oncoprotein, which in turn, results in activation of the NF- κ B pathway, the critical initiator in ATLL development.^{1,3} Another key HTLV-1 derived protein important for tumorigenesis is the HTLV-1 basic zipper protein (HBZ) whose expression has been shown to induce

Srinivasan Sundararaj and Sandali Seneviratne contributed equally.

proliferation and malignant transformation of CD4 T-cells.⁴ However, from the low rate of HTLV-1 infected people that develop ATLL, it is clear that other factors besides expression of Tax and HBZ are required for the development of ATLL.

Interferon regulatory factor 4 (IRF4), a canonical lymphoid transcription factor (TF) is a key candidate for ATLL development. It co-occupies regulatory elements with NF- κ B and coordinates the expression of several target genes, including *MYC*, *CCR4*, and *BIRC3*, and together with NF- κ B was identified as a master regulator for ATLL development.⁵ Overexpression and genetic alterations of IRF4 have been identified as one of the most common factors underpinning ATLL development and are associated with poor prognosis and treatment outcome. Knockdown of IRF4 expression in ATLL cells results in increased apoptosis, suggesting the requirement of IRF4 for ATLL cell viability.⁶ Furthermore, treatment with either IRF4 antisense oligonucleotides (ASOs) or lenalidomide, an FDA-approved drug for multiple myeloma which indirectly affects IRF4 expression, have been shown to reduce the proliferation of ATLL cells.¹ Collectively, these studies have identified a key role for IRF4 in the pathogenesis of ATLL.

IRF4 is a member of the IRF family of TFs that are associated with the development of immune cells, including B- and T-cells. This family consists of eight other TFs, IRF1 through to IRF9, all of which recognize promoters consisting of the IRF consensus sequence 5'-GAAA-3'.⁷ However, IRF4 is distinct within this family of TFs as it is the only IRF that is not regulated by interferons (IFNs) and it has distinct roles in determining immune cell lineage and fate.⁸ Despite its unique functional role, the structure of IRF4 shares many conserved functional domains with other members of the IRF family.⁹ The general domain arrangement within this family consists of an N-terminal DNA-binding domain (DBD) and a C-terminal IFN association domain (IAD) (Figure 1). The DBD is highly conserved within the IRF family,

comprising of a pentad repeat of tryptophans that forms a type of winged helix–turn–helix structure^{10,11} facilitating its interaction with DNA. Intriguingly, in almost all the cases, mutations are detected in the DNA-binding domain, implying that the IRF4/DNA interaction is a key driver for malignant transformation. IAD is a protein–protein interaction domain that enables its interaction with other TFs. Connecting these two functional domains is a flexible linker. Additionally, the distal C terminal region comprises an autoinhibitory domain, a short sequence of amino acids believed to play a role in the regulation of IRF4 interactions with DNA.¹²

Several DNA regulatory elements were identified as targets for IRF4 interaction.¹³ Notably, IRF4 can bind its target DNA as either a homodimer or heterodimer in association with other TFs. For instance, it engages IFN-stimulated response elements (ISRE) DNA as a homodimer. Conversely, it binds erythroblast transformation specific (Ets) IFN composite elements (EICE), and AP-1-IRF composite elements (AICE1 or 2) as a heterodimer and requires the engagement of PU.1 and BATF TF, respectively, for these interactions.⁸

Due to its critical role in the development of B- and T-cells, IRF4 dysfunction has been implicated in several lymphoid malignancies, including chronic lymphocytic leukemia, multiple myeloma and adult T-cell leukemia. Markedly, the pathogenic mutations are often detected in the DNA-binding domain, implying that the IRF4/DNA interaction is a key driver for malignant transformation. In the case of CLL, a recurrent somatic heterozygous mutation affecting the IRF4 DBD (IRF4 L116R) was identified in 1.5% of CLL patients.¹⁴ Likewise, recurrent somatic mutations (L116R, K123R) in the IRF4 DBD are observed in MM patients.^{15,16}

Molecular analysis of ATLL cells has identified IRF4 as one of the most frequently mutated genes in ATLL development.¹⁷ Notably, mutations that result in the amplification of the IRF4 gene were observed in 1 in 4 patients, while 1 in 7 had activating single-nucleotide

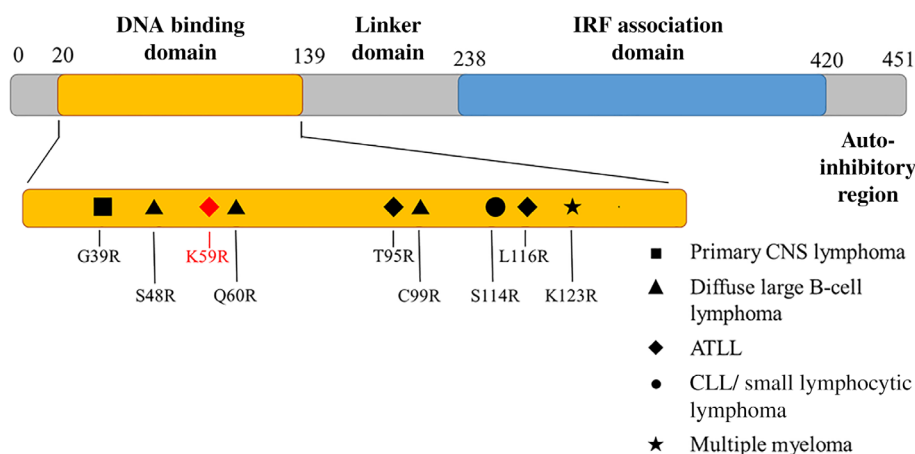


FIGURE 1 Schematic representation of the IRF4 domain structure. The functionally critical DNA binding domain and IRF association domain are depicted in yellow and blue, respectively. Clinically relevant arginine substitution mutations that occur within the IRF4 DNA-binding domain are also presented using individual symbols to represent the relevant disease

variations (SNV).^{1,17} Previous studies have shown IRF4^{K59R} as a recurrent somatic mutation in ATLL. In wildtype, lysine 59 lies in the DBD of IRF4 and is fairly conserved amongst the IRF family of TFs. Interestingly, *Irf4*^{K59R} is almost exclusive to ATLL development, suggesting a very specific effect of the mutation.⁶ Notably, the *Irf4*^{K59R} mutation was shown to enhance nuclear expression and transcriptional activity of IRF4.⁶ Nevertheless, the mechanistic basis of the IRF4^{K59R}-mediated gain of function in ATLL remains unknown.

To understand the molecular basis of IRF4^{K59R} and its implication in ATLL development, we have determined the crystal structure of the IRF4^{K59R} homodimer in complex with the well-characterized ISRE DNA found in the murine *Prdm1* gene (5'-CAACTGAAACCGAGAAAGC-3') comprising two overlapping consensus IRF (GAAA) recognition sequences.¹⁸ Our study has shown that the arginine substitution resulted in increased DNA interactions and greater DNA distortions, thereby providing a rationale for altered transcriptional regulation and oncogenic effects.

2 | RESULTS

2.1 | Arginine substitution increases DNA affinity

To determine how the arginine substitution can influence the affinity of the IRF4/DNA interaction and rationalize

the basis for the observed gain of function in the ATLL patients, we measured the direct binding affinity of this interaction by SPR. Purified IRF4^{K59R} protein was titrated against the known IRF4 DNA target motifs and disassociation constant (K_D) values were calculated on the basis of equilibrium analysis. We observed a varying degree of binding between IRF4^{K59R} and the target motifs, ranging from low nanomolar to low micromolar affinities. IRF4^{K59R} binds robustly to the canonical ISRE DNA sequence with a K_D of $0.09 \pm 0.01 \mu\text{M}$. This indicates a two- to threefold increase in binding affinity compared to the WT ($0.25 \pm 0.15 \mu\text{M}$). A similar trend was also observed for other DNA motifs (EICE and AICE) with the IRF4 mutant binding with a K_D of $0.18 \pm 0.01 \mu\text{M}$ and $0.91 \pm 0.17 \mu\text{M}$, respectively. These binding affinities were consistently higher when compared to their WT counterparts¹⁹ (Figure 2). Together, the affinity study showed that lysine 59 to arginine substitution strengthens the IRF4 DNA-binding domain's interaction with the target DNA, providing a potential rationale linking the oncogenic effects of this mutation to the observed enhanced DNA interaction.

To determine if the IRF4^{K59R} mutation influences the structural stability of the IRF4 protein, thermal stability measurements of both IRF4^{WT} and IRF4^{K59R} in the presence and absence of DNA were examined using DSF. IRF4^{K59R} has a T_m of $57.06 \pm 0.60^\circ\text{C}$ which is lower than the T_m of WT ($T_m = 59.12 \pm 0.25^\circ\text{C}$). Using the same technique, we also measured the structural stability of these proteins in the DNA-bound state. It is apparent that

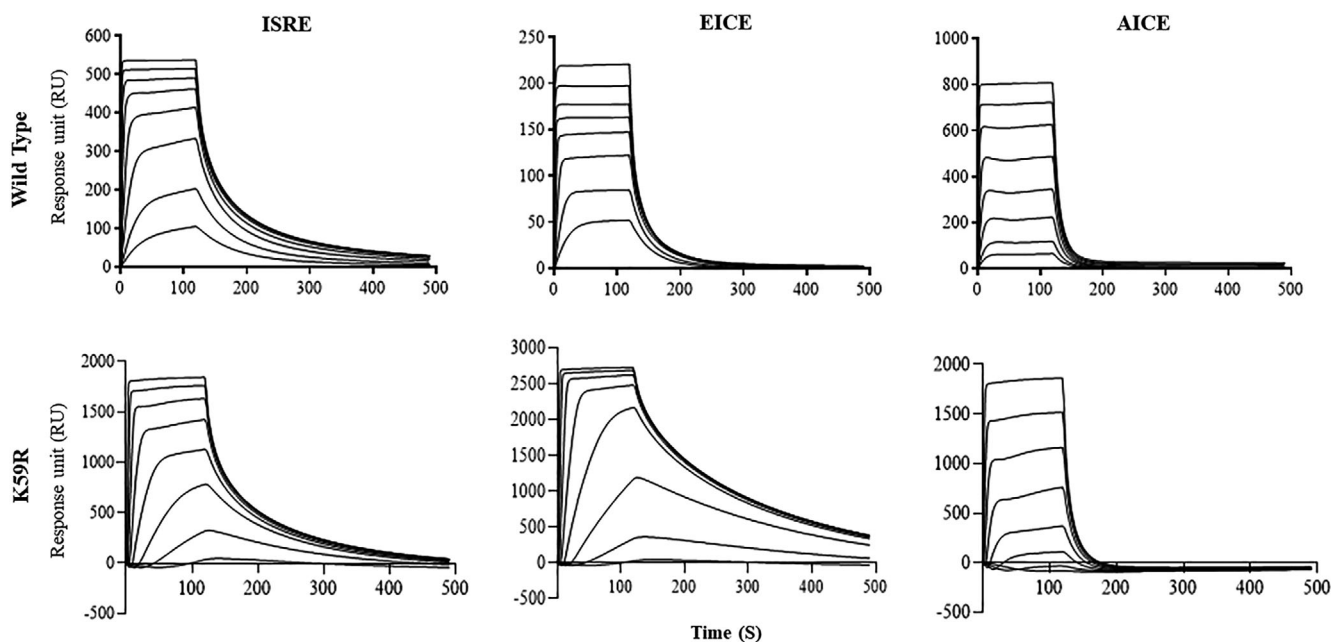


FIGURE 2 Affinity measurement of IRF4/DNA interactions. Top panel corresponds to a typical sensorgram for IRF4^{WT} for specific DNA targets. The bottom panel shows the sensorgram for the same DNA targets for IRF4^{K59R}

when bound to ISRE motif, the thermal stability of both the IRF4^{WT} and IRF4^{K59R} is comparable (Figure 3a,b). Although both DBDs show a significant increase in stability upon ISRE binding, it is of particular note that the stability of the IRF4^{K59R} DBD was affected more than that of the IRF4^{WT}. While the IRF4^{WT} DBD showed an average increase of 3.38°C upon complexation, the T_m for IRF4^{K59R} DBD increased by nearly double that of the IRF4^{WT} (increase of 6.14°C). We further assessed the possible implication of reduced stability of the unbound IRF4^{K59R} on the secondary structural elements by testing WT and mutant IRF4 proteins using Far-UV CD spectroscopy. Globally, as expected the spectrum for both IRF4^{WT} and IRF4^{K59R} were comparable (Figure 3c), suggesting that mutation has no impact on the folding of the protein.

2.2 | Structural determination

To characterize the structural basis of the observed IRF4^{K59R} gain of function, we purified and co-complexed IRF4 and DNA using size exclusion chromatography (Figure 4a–c) and determined the crystal structure of the IRF4^{K59R}-ISRE DNA homodimer. The complex crystallized in the space group of $P3_1 2 1$ with the unit cell dimension of $a = 117.8$, $b = 117.8$, $c = 154.58$ and $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 120^\circ$. The co-crystals were grown by hanging drop vapor diffusion from 1:2 M ratio (protein: precipitant) of well solution containing 4% PEG 4000 and 0.1 M Na acetate pH 4.8. The structure was determined by molecular replacement and refined to the final R_{work} and R_{free} of 20.3% and 24%, respectively

(Table 1). The final model comprised of four IRF4 DBDs namely IRF4-A (aa 22–129), IRF4-B (aa 21–129), IRF4-G (aa 21–129), and IRF4-H (aa 21–129), and two DNA

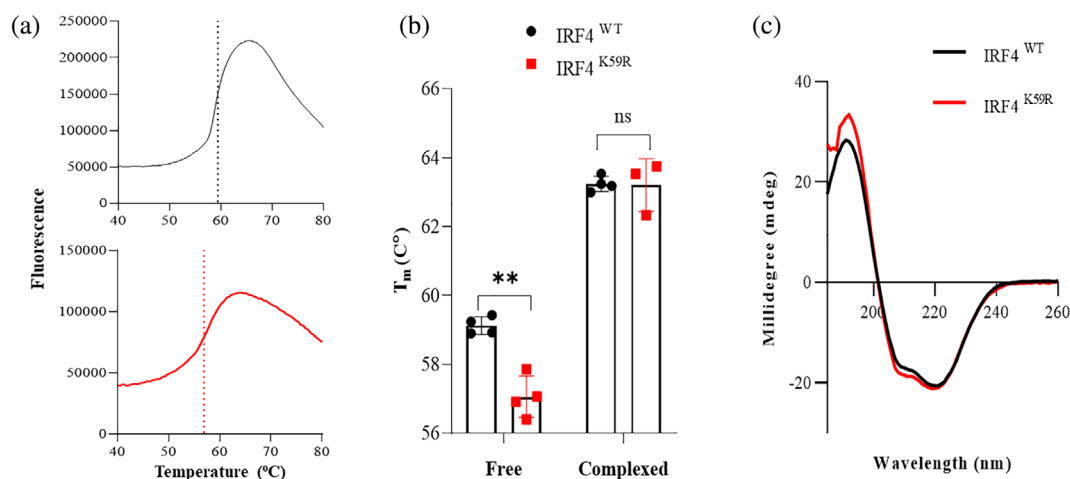


FIGURE 3 IRF4 protein structural and thermal stability of IRF4^{WT} (black) and IRF4^{K59R} (red). (a) Representative melting curves of IRF4^{WT} (upper panel) and IRF4^{K59R} (lower panel) DBD proteins showing T_m on curve. (b) Comparison of the T_m of IRF4 proteins between unbound and DNA-bound complexes. A two-tailed T test was performed to determine degree of statistical difference between conditions. (c) CD spectrum of IRF4^{WT} and IRF4^{K59R}

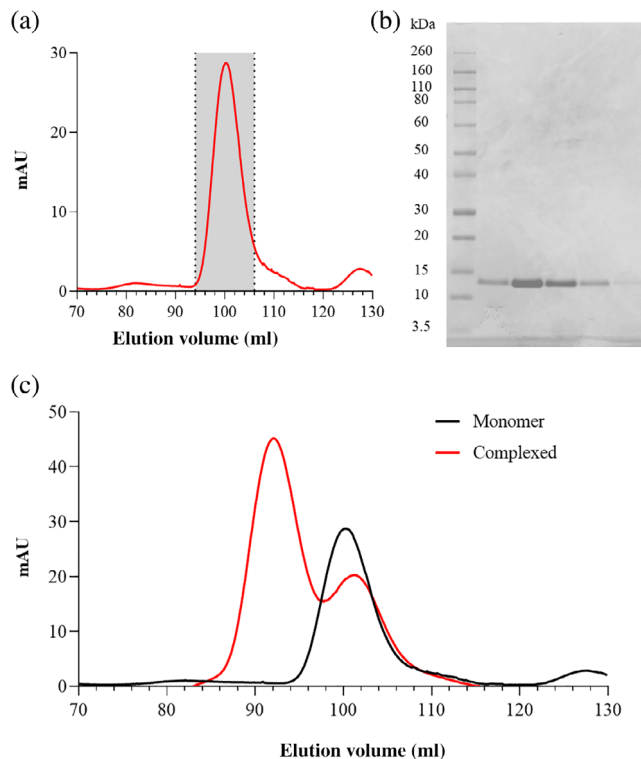


FIGURE 4 Purification of IRF4^{K59R} protein: (a) Size exclusion elution profile of IRF4^{K59R} protein eluted from S200 16/600 gel filtration column. (b) 15% SDS PAGE showing the purity of the SEC-eluted IRF4^{K59R} protein. (c) SEC elution profile for IRF4^{K59R} protein and IRF4^{K59R}/ISRE DNA complex eluted from S200 16/600 gel filtration column

TABLE 1 Data collection and refinement statistics

IRF4 ^{K59R} homodimer complex	
Wavelength	
Resolution range	41.84–2.475 (2.563–2.475)
Space group	P 31 2 1
Unit cell	114.874 114.874 154.788 90 90 120
Total reflections	432,755 (44,104)
Unique reflections	42,586 (4,077)
Multiplicity	10.2 (10.2)
Completeness (%)	99.68 (97.30)
Mean I/sigma(I)	15.4 (1.7)
Wilson B-factor	59.80
R-merge	0.072 (0.850)
R-meas	0.076 (0.894)
R-pim	0.024 (0.275)
CC1/2	0.998 (0.832)
Reflections used in refinement	42,579 (4,077)
Reflections used for R-free	2,001 (191)
R-work	0.2035 (0.2874)
R-free	0.2400 (0.3567)
Number of non-hydrogen atoms	5,209
Macromolecules	5,196
Ligands	0
Solvent	13
Protein residues	435
RMS(bonds)	0.011
RMS(angles)	1.29
Ramachandran favored (%)	96.25
Ramachandran allowed (%)	3.75
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.03
Clashscore	3.42
Average B-factor	64.87
Macromolecules	66.14
Solvent	51.40
Number of TLS groups	31

Note: Statistics for the highest-resolution shell are shown in parentheses.

duplexes, namely DNA-1 (chains D and E) and DNA-2 (chains C and F). Together, these represented two homodimer complexes in the asymmetric unit. Superimposition of wild-type and mutant homodimer complexes gave a root-mean-square (RMS) deviation of 0.183 Å for all atoms, indicating both complexes were very similar but not identical. For subsequent structural analysis, the IRF4^{K59R} homodimer structure consisting of IRF4-A, IRF4-B, and DNA-1 (chains D and E) was used and were compared with the respective chains in the IRF4^{WT}

homodimer complex (PDB 7JM4), unless otherwise mentioned.

2.3 | IRF4^{K59R}/DNA homodimer complex structure

The overall structure revealed the IRF4 DBD binds to the opposite face of DNA in a head to tail orientation (Figure 5a–f). Both domains adopted a conserved α/β

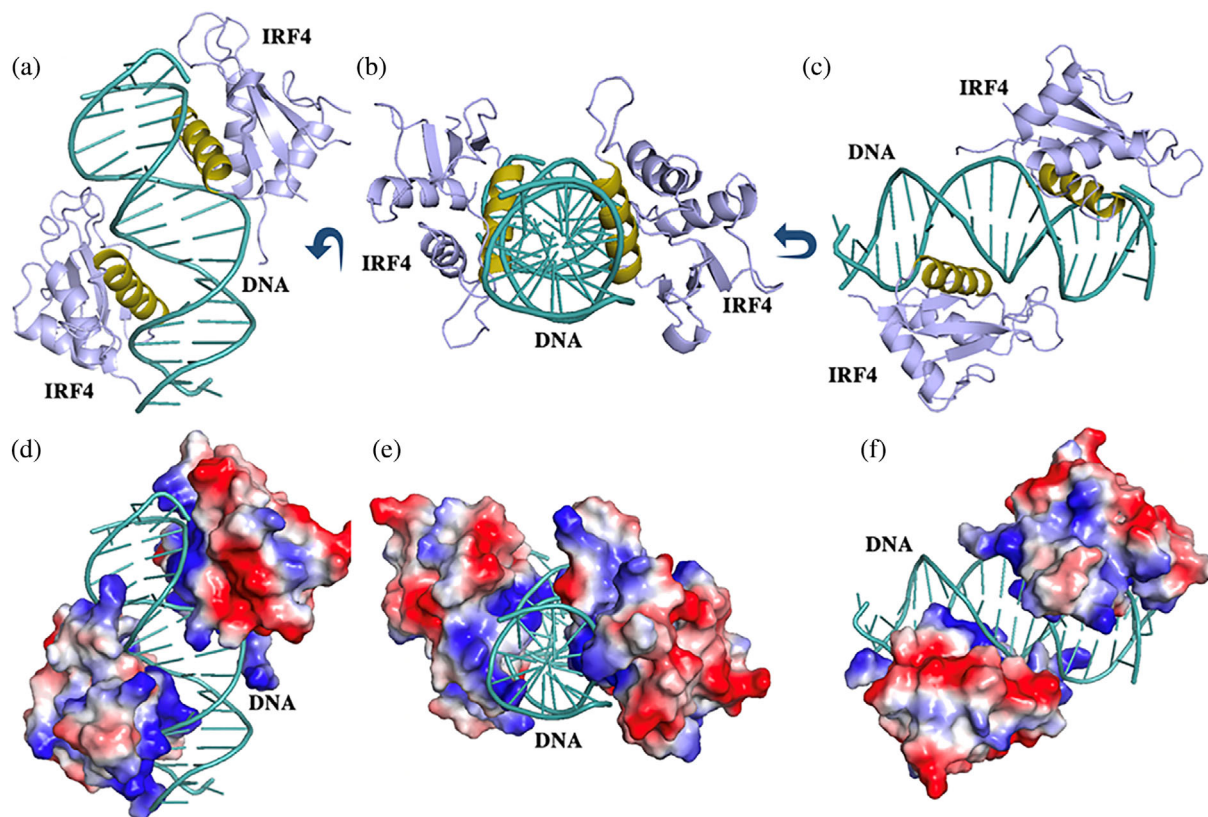


FIGURE 5 Overall structure of the IRF4^{K59R}/ISRE DNA complex. (a) Standard view, (b) top view, (c) side view. IRF4^{K59R}, lightblue α 3-recognition helix, olive; DNA, light teal, respectively. (d) and (e) represent the electrostatic view of the IRF4^{K59R} protein in standard, top, and side view, respectively

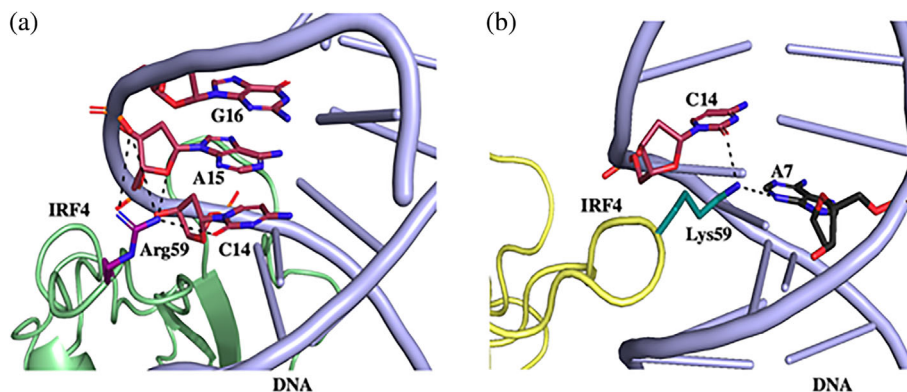
structural topology and were comprised of three α -helices (α 1– α 3) with four standard antiparallel β -sheets (β 1– β 4) and three long loops (L1–L3). The overall structural features were comparable to the previously determined IRF4^{WT}/DNA homodimer complex structure.¹⁹ Since IRF4^{K59R} substitution occurs in the connecting IRF4 loop L1, we assessed the conformational rearrangement of this region. Superimposition of both the IRF4^{WT} and IRF4^{K59R} structures showed minimal structural rearrangement in this loop region (L1) with an RMS deviation of 0.308 Å for all atoms, indicating that the mutation has little effect on the overall conformation of L1. However, closer inspection of the mutation site in the IRF4 DBD (IRF4-A) revealed considerable repositioning of the arginine side-chain. Specifically, in the case of the IRF4^{WT}, Lys 59 is positioned more centrally in the minor groove of the DNA, enabling it to contact both the DNA strands (chains D and E). Namely, the positioning has resulted in a hydrogen-bonded contact with the adenine (A7) (underlined) of the GAAA consensus sequence (chain D). Furthermore, a second hydrogen bond forms with the cytosine (C14) of the chain E. Conversely, in the case of IRF4^{K59R} complex, the side chain of Arg 59 has moved \sim 4 Å away from its wildtype counterpart and interacts exclusively with the chain E of the DNA.

Notably, this distinctive change in orientation has resulted in an obvious difference in the DNA contact. Specifically, the repositioning has led to the formation of two hydrogen bonds, one between N^{H2} of the arginine with the O4 of the adenine (A15) and the other with the O2 of the cytosine (C14). Additionally, the repositioning of the side chain has allowed the guanidinium group to interact intimately with the backbone of guanine (G16) through two salt bridges, which are further enhanced by the van der Waal's (VDW) interactions (Figure 6a,b, S2). For IRF4-B, both the residues in the equivalent structures sit at the edge of the DNA and while a similar pattern was observed in side-chain positioning, no direct DNA contact was evident in both WT and IRF4^{K59R} (cutoff $<$ 5 Å, figure not shown). Together, it shows that arginine substitution is confined to changes in the side-chain orientation and affects the local environment of the DNA-binding domain thereby resulting in an overall increase in the DNA interaction.

2.4 | DNA structure deformation

We and others have shown that IRF4 engages its target DNA by distorting its structure.^{19,20} This distortion and

FIGURE 6 Cartoon representation of IRF4^{WT} and IRF4^{K59R} structure highlighting the arginine and lysine interaction. (a) IRF4^{K59R}–DNA interaction. (b) IRF4^{WT}–DNA interaction. IRF4^{WT}, yellow; IRF4^{K59R}, green; DNA, lightblue; Arg59, pink; Lys59, deep teal; hydrogen bond and electrostatic interactions are represented by black-dashed lines



changes in the related conformational parameters have been identified as one of the prerequisites for the optimal engagement of the interacting IRF4 monomer. The availability of the IRF4^{WT}/ISRE homodimer complex structure has enabled us to compare the extent of the DNA distortion and conformational changes upon arginine substitution using the nucleic acid-based analysis package, Curves+.^{19,21} The analysis showed that mutation increases the overall bending of the DNA, with a bending angle of approximately 21° (WT 15°). Differences were also observed in other helical parameters, including an average propeller twist accounting to −10.3° (WT −11.7°) and base pair tilt of 1 Å (compared to 1.5 Å for the WT). However, other parameters such as mean axial rise per turn (3.38 Å) and average helical twist (30.8°) were comparable to the WT (Figures S3A and S3B). Collectively, the IRF4^{K59R}/DNA complex structure shows that the arginine substitution is not only associated with changes in the side-chain orientation and the local DNA-binding environment, but extends to have a more global structural effect.

3 | DISCUSSION

IRF4 is a key immune regulator that has been identified to have a direct impact on the development of lymphoid malignancies, including ATLL. Several recurrent mutations have been identified in the DBD of IRF4 and were experimentally shown to have an oncogenic effect in the malignant transformation of the target cell.⁶ Of particular relevance is the IRF4^{K59R} mutation, which is exclusive to ATLL development and is the most frequent mutation in the IRF4 DBD sited in the COSMIC database. Lys 59 has been identified as a highly conserved residue across evolutionary diverse species and we have recently shown that it interacts directly with the IRF consensus sequence in the formation of the IRF4 homodimer complex.^{6,19} Although previous studies have recognized the importance of this

mutation in disease development, the underlying molecular basis has not been explored. In the current study, determination of the IRF4^{K59R}/ISRE DNA homodimer crystal structure has provided a rationale for the observed gain of function in the IRF4^{K59R} mutant.

Our affinity data revealed that the arginine substitution causes a notable impact on DNA-binding affinity. We found that the replacement of lysine with arginine increased the overall binding strength of the IRF4–DNA interactions to all DNA targets included in the study. The conclusions from our affinity study are consistent with a previous finding which indicated that the IRF4^{K59R} leads to increased nuclear expression and transcriptional potency compared to the IRF4^{WT},⁶ linking disease development to direct IRF4–DNA interaction. Further, we evaluated the impact of the arginine substitution on the overall structural stability of the IRF4 proteins which showed that although arginine substitution led to a moderate decrease in protein stability in the unbound state, this is more than compensated for by the increased stability of the protein–DNA complex. This observed DNA-induced structural stabilization is a feature commonly observed in other transcription factors.^{22,23}

Several prominent structural differences were also observed in the homodimer structure upon arginine substitution. At a local level, a seemingly conservative mutation from a lysine to an arginine residue resulted in the reorientation of the arginine sidechain so that it now participates in hydrogen bond and salt bridged interactions with the opposing DNA strand, compared with the WT lysine residue. Globally, a major transformation lies in the extent of DNA deformation, in particular DNA bending. Comparison of IRF4^{WT} and IRF4^{K59R} has identified a significant increase in the DNA bend upon arginine mutation. It is well-known that DNA bending plays an essential role in transcriptional regulation by bringing promoters and enhancers into close proximity, thereby facilitating co-transcriptional regulation.²⁴ In addition, distortion of the DNA may also mediate interactions

between the transcription factors by wrapping the DNA into large protein–DNA complexes.²⁴ Based on the structure, it is likely that the increased bending may promote altered binding of other TFs potentially targeting regulatory elements which may affect the oncogenic outcome.

The fact that a single realignment of a positively charged sidechain can exert such global structural effects on the protein–DNA complex points to the precise role the DNA-binding region of IRF4 plays in the recognition of its DNA-binding partners. In the context of other disease-causing mutations in the IRF4 DBD, this region represents a mutant hotspot for lymphocytic leukemias. Notably, arginine substitution is more recurrent in several other lymphoid malignancies, including chronic lymphocytic leukemia (L116R)¹⁴ and multiple myeloma (K123R).¹⁶ On the basis of the IRF4^{WT} homodimer crystal structure, it is apparent that Leu 116 and Lys 123 are in close proximity to the DNA backbone. Although at this stage, the exact structural effects of these arginine substitutions are not known, we hypothesize these arginine mutations would likely increase the local electropositive charge and potentially enhance the interaction with the phosphate backbone in a manner similar to IRF4^{K59R}. This raises the possibility that the oncogenic effects attributed to these mutations could be driven by the increase in the electrostatic contact.

The most common genetic mutations (>50%) in human cancers occur in the P53 gene²⁵ with most of the mutations presenting in the DNA-binding core of the protein.²⁶ The P53 mutation R280K is a loss of function mutant and involves a loss in transcriptional activity and has been related to different cancer types, including ureter, gallbladder, bladder, and prostate cancers (<http://www-p53.iarc.fr/>). From a structural viewpoint, this loss of function mutation can be rationalized by the inability of the lysine mutation to maintain two hydrogen bonds with DNA as depicted in (Figure S4). This mutation represents a mirror image of the situation encountered by the IRF4^{K59R} mutant and highlights the structural and functional effects that may be imposed from a seemingly conservative mutation.

Collectively, we have shown that the gain of function incurred due to the arginine substitution in IRF4^{K59R} can be attributed to its enhanced DNA interaction coupled with its local and global structural changes. This has provided a molecular rationale for the IRF4^{K59R}-mediated oncogenic effect in ATLL pathogenesis. It will be interesting to map the structural changes of other clinically relevant IRF4 arginine mutations, which will clearly underpin the molecular basis for their observed gain of function and could provide a basis for the development of IRF4-centric therapeutics.

4 | MATERIALS AND METHODS

4.1 | Expression and purification

Codon-optimized IRF4WT and IRF4K59R gene constructs were cloned into pJ411KanR (ATUM) and overexpressed as an N-terminal His6-tag fusion protein in *Escherichia coli* BL21(DE3) (Novagen) at 18°C following induction with 0.5 mM IPTG. The cells were resuspended in 50 mM sodium phosphate buffer (pH 7.0), 500 mM NaCl, 3 mM β -mercaptoethanol, 0.5% Triton X-100, 4 mM MgCl₂, and protein inhibitor cocktail (Roche), lysed by French press at 1,500 psi and cleared by centrifugation, and IRF4 proteins were purified using a 5 ml HisTrap column (GE Healthcare) in 50 mM sodium phosphate buffer (pH 7.0), 500 mM NaCl and 20–500 mM imidazole gradient. To cleave the His6-tag, the eluted fractions were pooled and subjected to an overnight HRV3c protease digestion at 4°C in 50 mM sodium phosphate buffer (pH 7.0), 500 mM NaCl and 20 mM imidazole. The His6-tag cleaved IRF4 proteins were subsequently loaded and purified by passing through a 5 ml HisTrap column (GE Healthcare). The flow-through was later dialyzed into 10 mM Tris buffer (pH 7.0), 150 mM NaCl and 1 mM TCEP and purified subsequently using size exclusion chromatography. The co-complexation of IRF4 and DNA was performed as described previously.¹⁹

4.2 | Microplate thermal shift assay

Thermal shift data was acquired via differential scanning fluorimetry (DSF) performed using a Quant Studio 3 qPCR machine (Life technologies). The SEC-purified IRF4WT and IRF4K59R proteins, both unbound and bound to ISRE DNA, were concentrated to 0.3 mg/ml and mixed with Sypro Orange 1X Protein Thermal Shift dye (Life technologies) in 20 mM Tris buffer (pH 7.4), 150 mM NaCl, 1 mM TCEP buffer. Twenty microliter of the protein–dye mixture was then distributed into wells of a 96-well FAST-block optical plate (Life technologies) in four technical replicates. Emissions at 610 nm were monitored as the temperature was increased from 25 to 95°C, at a rate of 0.05°C/s. For DSF data of protein–DNA complexes, IRF4WT and IRF4K59R DBD constructs were incubated with ISRE motif (Integrated DNA technologies) in a 1:0.5 M ratio overnight at 4°C to ensure full complex formation. Other procedures for obtaining thermal shift data for the protein–DNA complexes are as above. The data was analyzed using Protein Thermal Shift software (ThermoFisher) with GraphPad Prism

Version 8.0 used for data presentation and the melting temperature, T_m calculated.

4.3 | Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were carried out at 20°C on a Biacore 8,000 instrument (Cytiva) using HBS buffer (10 mM HEPES-HCl, pH 7.4, and 150 mM NaCl) with 3 mM EDTA and 0.05% P20 as a running buffer. The biotinylated DNA motifs (ISRE, EICE, AICE) (Integrated DNA technologies) (Table S1) were coupled (up to ~2,000 RU) onto a streptavidin (SA) chip. Measurement of IRF4/DNA interaction affinity was performed by injecting incrementally increasing concentrations of IRF4K59R (up to 5 μ M) at 30 μ l/min flow rate. The final response unit was calculated by subtracting the response unit of the reference flow cell. The steady-state multi cycle affinity data were fitted using the Biacore 8 K BIAevaluation software. GraphPad Prism Version 8.0 was used for data presentation.

4.4 | Far-UV circular dichroism (CD) spectroscopy

CD spectra were acquired using a Chirascan spectrometer (Applied Photophysics) with a circulating water bath at 20°C using a cuvette with a 0.1-mm optical path length. For CD experiments, IRF4 proteins were prepared in 10 mM potassium phosphate and 50 mM sodium fluoride (pH 6.8) at a concentration of 0.15 mg/ml. Samples were prefiltered and spectra recorded from 180 to 260 nm using a scanning speed of 15 nm/min. Similarly, spectra were acquired with buffer alone which and subtracted from IRF4 spectra. A total of four independent scans were recorded and the mean residue ellipticity was calculated from the protein absorbance at 280 nm, residue number, and molecular weight.

4.5 | Crystallization and structural determination

The homodimer isolated from SEC were concentrated up to 7 mg/ml and initially screened using 200 nanoliter sitting-drop format with commercially available 96-well screens at 19°C. The initial crystallization hits were subsequently optimized and the complexes were crystallized using a hanging drop vapor diffusion method in 3–9% PEG 4000, 0.1 M Na acetate pH 4.8 at 18°C. Single diffraction quality crystals were cryoprotected using the mother liquor plus 25% PEG 4000, and flash-frozen in liquid nitrogen. Diffraction of the crystals was carried out at

the MX2 beamline (Australian synchrotron). The dataset was processed with the XDS software package and scaled using Aimless^{27,28} in the CCP4 suite.²⁹ The crystal structure of the IRF4K59R-ISRE homodimer complex was determined by molecular replacement using the Phaser-MR program with IRF4 Protein Data Bank (PDB) ID code 7JM4 and ideal B DNA as a separate search model. Iterative model building and subsequent refinement cycles were performed with the program COOT and Phenix refine, respectively.^{30,31} The quality of the structure was validated at the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank Validation and Deposition Services. All presentations of molecular graphics were created with the programme PyMOL.

ACKNOWLEDGEMENTS

This research utilized the MX2 beamline at the Australian Synchrotron (ANSTO) as well as the Australian Cancer Research Foundation (ACRF) detector. A.E was supported by National Health and Medical Research Council grant (GNT1079648). We also thank Dr. Emmalene Bartlett and Professor Philip Board (JCSMR, Australian National University) for critically proofreading the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Srinivasan Sundararaj: Conceptualization (equal); formal analysis (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Sandali Seneviratne:** Investigation (equal); methodology (equal); writing – original draft (equal); writing – review and editing (equal). **Simon J Williams:** Formal analysis (equal); investigation (equal); writing – review and editing (supporting). **Anselm Enders:** Resources (equal); writing – original draft (supporting); writing – review and editing (supporting). **Marco G Casarotto:** Conceptualization (equal); formal analysis (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT


Atomic coordinates and structure factors for the reported crystal structure have been deposited with the Protein Data bank under accession number 7RH2.

ORCID

Srinivasan Sundararaj  <https://orcid.org/0000-0002-9112-2454>

Sandali Seneviratne  <https://orcid.org/0000-0001-5626-7968>

Simon J Williams  <https://orcid.org/0000-0003-4781-6261>

Anselm Enders  <https://orcid.org/0000-0001-5933-6463>
Marco G. Casarotto  <https://orcid.org/0000-0002-0571-7671>

REFERENCES

1. Rauch DA, Olson SL, Harding JC, et al. Interferon regulatory factor 4 as a therapeutic target in adult T-cell leukemia lymphoma. *Retrovirology*. 2020;17:27.
2. Matutes E. Adult T-cell leukaemia/lymphoma. *J Clin Pathol*. 2007;60:1373–1377.
3. Shudofsky AMD, Giam CZ. Cells of adult T-cell leukemia evade HTLV-1 Tax/NF-kappaB hyperactivation-induced senescence. *Blood Adv*. 2019;3:564–569.
4. Satou Y, Yasunaga JI, Zhao T, et al. HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. *PLoS Pathog*. 2011;7:e1001274.
5. Wong RWJ, Tan TK, Amanda S, et al. Feed-forward regulatory loop driven by IRF4 and NF-kappaB in adult T-cell leukemia/lymphoma. *Blood*. 2020;135:934–947.
6. Cherian MA, Olson S, Sundaramoorthi H, et al. An activating mutation of interferon regulatory factor 4 (IRF4) in adult T-cell leukemia. *J Biol Chem*. 2018;293:6844–6858.
7. Biswas PS, Bhagat G, Pernis AB. IRF4 and its regulators: Evolving insights into the pathogenesis of inflammatory arthritis? *Immunol Rev*. 2010;233:79–96.
8. Nam S, Lim JS. Essential role of interferon regulatory factor 4 (IRF4) in immune cell development. *Arch Pharm Res*. 2016;39:1548–1555.
9. Remesh SG, Santosh V, Escalante CR. Structural studies of IRF4 reveal a flexible autoinhibitory region and a compact linker domain. *J Biol Chem*. 2015;290:27779–27790.
10. Brennan RG, Matthews BW. The helix-turn-helix DNA binding motif. *J Biol Chem*. 1989;264:1903–1906.
11. Yanai H, Negishi H, Taniguchi T. The IRF family of transcription factors: Inception, impact and implications in oncogenesis. *Onco Targets Ther*. 2012;1:1376–1386.
12. Taniguchi T, Ogasawara K, Takaoka A, Tanaka N. IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol*. 2001;19:623–655.
13. Shukla V, Lu R. IRF4 and IRF8: Governing the virtues of B lymphocytes. *Front Biol (Beijing)*. 2014;9:269–282.
14. Havelange V, Pekarsky Y, Nakamura T, et al. IRF4 mutations in chronic lymphocytic leukemia. *Blood*. 2011;118:2827–2829.
15. Chapman MA, Lawrence MS, Keats JJ, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011;471:467–472.
16. Lohr JG, Stojanov P, Carter SL, et al. Widespread genetic heterogeneity in multiple myeloma: Implications for targeted therapy. *Cancer Cell*. 2014;25:91–101.
17. Kataoka K, Nagata Y, Kitanaka A, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet*. 2015;47:1304–1315.
18. Ochiai K, Maienschein-Cline M, Simonetti G, et al. Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity*. 2013;38:918–929.
19. Sundararaj S, Seneviratne S, Williams SJ, Enders A, Casarotto MG. Structural determinants of the IRF4/DNA homodimeric complex. *Nucleic Acids Res*. 2021;49:2255–2265.
20. Escalante CR, Brass AL, Pongubala JMR, et al. Crystal structure of PU.1/IRF-4/DNA ternary complex. *Mol Cell*. 2002;10:1097–1105.
21. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res*. 2009;37:5917–5929.
22. Mathiasen L, Valentini E, Boivin S, et al. The flexibility of a homeodomain transcription factor heterodimer and its allosteric regulation by DNA binding. *FEBS J*. 2016;283:3134–3154.
23. Leitner PD, Vietor I, Huber LA, Valovka T. Fluorescent thermal shift-based method for detection of NF-kappaB binding to double-stranded DNA. *Sci Rep*. 2021;11:2331.
24. van der Vliet PC, Verrijzer CP. Bending of DNA by transcription factors. *Bioessays*. 1993;15:25–32.
25. Joerger AC, Fersht AR. Structure-function-rescue: The diverse nature of common p53 cancer mutants. *Oncogene*. 2007;26:2226–2242.
26. Hamroun D, Kato S, Ishioka C, Claustres M, Bérout C, Soussi T. The UMD TP53 database and website: Update and revisions. *Hum Mutat*. 2006;27:14–20.
27. Kabsch W. XDS. *Acta Cryst D*. 2010;66:125–132.
28. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Cryst D*. 2013;69:1204–1214.
29. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Cryst D*. 2011;67:235–242.
30. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Cryst D*. 2010;66:213–221.
31. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst D*. 2010;66:486–501.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sundararaj S, Seneviratne S, Williams SJ, Enders A, Casarotto MG. The molecular basis for the development of adult T-cell leukemia/lymphoma in patients with an IRF4^{K59R} mutation. *Protein Science*. 2022;31:787–96. <https://doi.org/10.1002/pro.4260>