Targeting the IRF4 Transcriptional Network to Subvert Multiple Myeloma

Alessandro Agnarelli

University of Sussex

Thesis submitted for the degree of

PhD in Biochemistry

September 2021

Abstract

Multiple Myeloma (MM) is an incurable hematologic malignancy characterised by abnormal proliferation of plasma cells. Interferon Regulatory Factor 4 (IRF4), a transcription factor essential for immune system regulation and plasma cell differentiation that exerts its action by binding to specific DNA sequences called interferon sequence response element (ISRE), has emerged as the master regulator of an aberrant gene expression programme in MM. Overexpression of IRF4 is found in MM patients' derived cells, where it is key to their survival. Accumulating evidence suggests that IRF4 and MYC regulate each other in MM cell lines, creating a positive regulatory loop resulting in the proliferation of MM cells. Despite its major role, IRF4 has not been targeted for therapeutic drug discovery programmes. Furthermore, key elements of the mechanism of action of IRF4 in the context of MM, including its ISRE binding strategies, have not been clearly elucidated. The aim of this thesis is to lay the groundwork towards the targeting of IRF4 to subvert MM. To that scope, we pursued several approaches including indirect targeting through IRF4's upstream epigenetic regulators and IRF4 crystal structural studies to understand the details of DNA binding which are going to be key findings towards IRF4 direct targeting. Given the positive auto regulation loop between MYC and IRF4, we examined the effect of the combination of IRF4 and MYC inhibitors on MM cells. Together with transcription factor network modelling of MM, the results point at additional and yet uncovered regulatory interactions within the IRF4 network. To dissect the mechanism of ISRE binding in MM, we solved the structure of the IRF4 DNA binding domain (DBD) in complex with various ISRE sequences. These data provide key insights into the ISRE binding specificity and affinity in the context of MM and are central to developing a small-molecules drug discovery programme to target IRF4.

Declaration

The thesis conforms to an 'article format' in which the middle chapters consist of discrete articles written in a style that is appropriate for publication in peer-reviewed journals in the field. The first and final chapters present synthetic overviews and discussions of the field and the research undertaken.

Chapter 2 is published in Leukemia Research as:

Agnarelli, A., T. Chevassut, and E.J. Mancini, *IRF4 in multiple myeloma-Biology, disease and therapeutic target.* Leuk Res, 2018. 72: p. 52-58.

doi: 10.1016/j.leukres.2018.07.025

The author contributions are as follows: Agnarelli, A. and E.J Mancini wrote the paperT. Chevassut helped in the drafting of the manuscript.

Chapter 3 is written in the style of an article appropriate for submission to a journal.

Alessandro Agnarelli, Simon Mitchell, David Wood, Leanne Milton-Harris, Timothy Chevassut, Michelle J West and Erika J Mancini. "Disrupting the IRF4-MYC oncogenic loop in Multiple Myeloma"

The author contributions are as follows: Alessandro Agnarelli was responsible for experimental procedures in laboratory, data analysis and wrote the manuscript. Simon Mitchell was responsible for the network modelling analysis. David Wood, Leanne Milton-Harris provided assistance in the laboratory. Timothy Chevassut and Michelle J West assisted in the planning of the project. Erika J Mancini planned the experiments, assisted with data analysis, supervised the project and wrote the manuscript.

Chapter 4 is published in the Acta Crystallographica Section F as follows:

Alessandro Agnarelli, Kamel El Omari, Ramona Duman, Armin Wagner, and Erika J. Mancini. Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4. Acta Crystallogr F Struct Biol Commun, 2021. 77(Pt 7): p. 202-207.

doi: 10.1107/S2053230X21006506

The author contributions are as follows: Alessandro Agnarelli purified and crystallised the protein and wrote the manuscript; Armin Wagner and Ramona Duman assisted with the experiments at the Diamond Light Source; Kamel El Omari assisted with the experiments at the Diamond Light Source, solved the crystal structure and wrote the manuscript; Erika Mancini planned the experiments, assisted with data analysis , supervised the project and wrote the manuscript.

Chapter 5 is written in the style of an article appropriate for submission to a journal.

Alessandro Agnarelli, Kamel El Omari, Aaron Alt, Leanne Milton-Harris, Daniel Adrian Epuran, David Wood, Timothy Chevassut, Michelle J West and Erika J Mancini "Investigating the binding mechanism of interferon regulatory factor 4 to DNA in the context of Multiple Myeloma"

The author contributions are as follows: Alessandro Agnarelli performed of all experimental procedures, solved the crystal structures and wrote the paper; Alessandro Agnarelli, Kamel El Omari and Aaron Alt assisted with structure refinement; David Wood and Leanne Milton-Harris provided assistance in the laboratory. Timothy Chevassut and Michelle J West assisted in the planning of the project. Erika J Mancini planned the experiments, assisted with data analysis and supervised the project.

WORK NOT SUBMITTED ELSEWHERE FOR EXAMINATION

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature: Alessandro Agnarelli

Table of Contents

Title page

Abstract

Declaration

Acknowledgments

Firstly, I would like to thank my parents who have supported me over the 4 years of the PhD studies.

I would like to thank everyone at the School of Life Science Department, for making it a very nice working environment over the last 4 years. Particularly I would like to thank David Wood and Megan Payne that with their conversations have made the laboratory a lovely place where to work.

Many thanks to Prof. Michelle West for the huge help with cell-culture experiments and general support in the laboratory.

Many thanks to my co-supervisor Dr. Timothy Chevassut for the support and help during my PhD studies.

A very big thanks to my supervisor Dr. Erika J Mancini who has had to deal with all of my questions as I taught myself biomolecular crystallography, for her huge availability during the most difficult moments of the PhD and for her help in all aspects related and not related to this project.

Abbreviations

ABC	activated B cell-like type
AD	activation domain
AICE	AP-1-IRF composite elements
AR	autoinhibitory region
ATL	adult T cell leukemia
BCMA	B cell maturation antigen
BET	bromodomain and extraterminal domain
bHLH	basic helix-loop-helix
bZIP	basic leucine zipper
CAR	chimeric antigen receptor
cHL	classical Hodgkin lymphoma
CLL	chronic lymphocytic leukemia
CREs	cis-regulatory elements
cryo-EM	cryogenic electron microscopy
CTM	carboxyterminal motif
DBD	DNA binding domain
DLBCL	diffuse large B cell lymphoma
EBV	epstein-barr virus
EICE	interferon composite elements
EMSA	electrophoretic mobility shift assay
ET	extraterminal
Ets	erythroblast transformation specific
FBDD	fragment-based drug discovery
GC	germinal center
GEFs	guanine-nucleotide exchange factors
GSK	GlaxoSmithKline
HAT	histone acetyltransferases
HD	homeodomain
HMG	high-mobility group box
HSCs	hematopoietic stem cells
HTLV1	human T cell leukemia virus-1

HTX	helix-turn-helix
IAD	IRF activation domain
IFN	interferon
IRE	interferon regulatory element
IRF4	interferon regulatory factor 4
IRF4 FL	IRF4 full length
IRFs	interferon regulatory factors
ISRE	interferon sequence response element
KAT	lysine acetyltransferase
LK	linker
MM	multiple myeloma
MR	molecular replacement
MST	microscale thermophoresis
NK	natural killer
NMR	nuclear magnetic resonance
ORR	overall response rate
PC	plasma cells
PEL	primary effusion lymphoma
PFS	progression-free survival
PRD I	positive regulatory domain I
RRMM	relapsed/refractory multiple myeloma
SAD	single-wavelength anomalous dispersion
SAR	structure-activity relationship
SGC	Structural Genomics Consortium
TFs	Transcription factors
TNFRSF17	tumor necrosis factor receptor superfamily member 17
WH	winged helix
XPO1	exportin-1

1. Chapter 1 Introduction

1.1. Transcription Factors

Gene expression is a complex process that is essential for living organism [1]. Transcription factors (TFs) are proteins that can bind a specific DNA sequence and they are the primary regulators of gene expression [1]. TFs regulate gene expression by binding to specific DNA sequences or cis-regulatory elements (CREs) in upstream, intron, or downstream regions of target genes. TFs can also exert their function by interacting with other genome locations that can be distant to the primary DNA sequence and they are called gene regulatory regions [2]. CREs include promoters and sequences called enhancers in cases of transcriptional activation, and silencers in cases of transcriptional repression [3]. The specific domain of TFs that can bind DNA is called the DNA binding domain (DBD). There are a variety of DNA binding structural motifs that can recognize the target sequences, which include homeodomain (HD), helix-turn-helix (HTH), high-mobility group box (HMG), zinc finger, basic leucine zipper (bZIP), winged helix (WH), basic helix-loop-helix (bHLH), Wor3 domain and OB-fold domain [4] [5] [6] [7] [8] [9] [10]. TF function involves the ability to bind specific short sequences of DNA within regulatory regions and the ability to recruit or bind proteins that participate in transcriptional regulation [11]. In many cases, TFs are limited to binding sites that lie within DNA regions that are more accessible in the genome [12]. These more accessible regions presumably are a consequence of earlier acting pioneer TFs, which have the ability to bind to nucleosome-coated DNA and alter chromatin structure, thereby allowing other TFs access to their binding sites [13]. Affinity is another parameter that influences TF binding site selection, which for many TFs can vary more than three orders of magnitude for different DNA sequences [14]. TF binding site selection problem can be solved in part by TFs choosing only the highest affinity binding sites [14]. In contrast, many other TFs and TF complexes bind *in vivo* to sub-optimal or low affinity binding sites that have poor matches to optimal consensus sites [15]. Zeiske *et al.* reported the structures of four heterodimers of the Hox protein Abdominal-B bound with its cofactor Extradenticle to four target DNA molecules [14]. In particular they showed that although the overall ternary structures are very similar, affinity depends on the predicted shape of

the DNA binding site prior to protein binding [14]. Indeed, binding sites that must structurally adapt upon protein binding showed a lower affinity than binding sites that are optimally pre-formed for protein binding [14]. These observations support a general model in which TF-DNA affinity is sensitive to differences in intrinsic DNA shape [14]. Regulatory element sequences have high intrinsic affinity for histone octamers, creating a strong barrier for access of TFs to the underlying DNA [16]. In order to overcome these obstacles, organisms have developed co-operative recognition of DNA by multiple TFs [17]. TFs can co-operate through a variety of mechanisms that include protein-protein interaction and indirect co-operation (Fig. 1) [17].

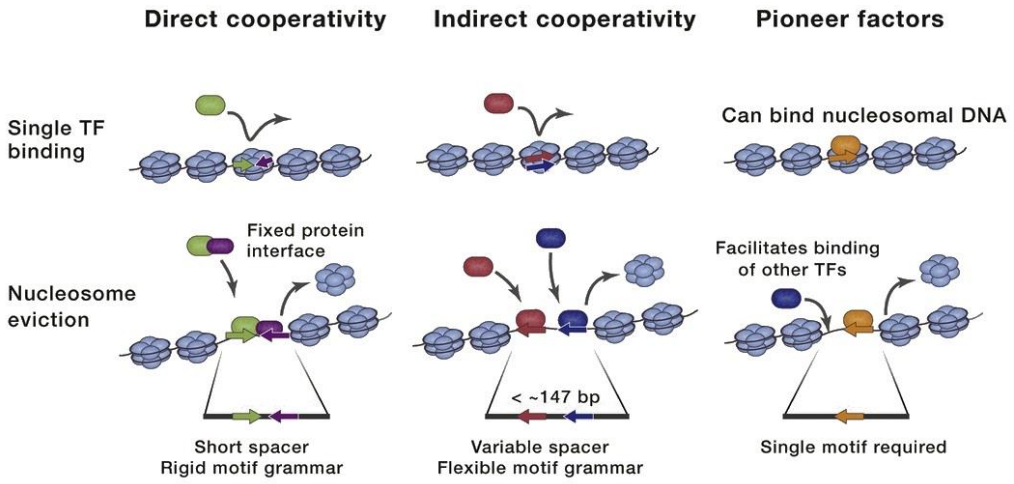


Fig. 1. TF mechanisms for overcoming low affinities to DNA. Adapted from [18].

The formation of functional dimers is an example of protein-protein interactions among TFs (Fig. 1). Indeed, some eukaryotic TFs are not able to bind DNA sequences as monomers but require the physical interactions with an identical molecule or one within the same family to form functional dimers able to bind specific DNA motifs [19]. The co-operative binding of TFs to DNA can also occur without direct protein-protein interactions. This co-operation is achieved through a process known as indirect co-operativity or collaborative competition, in which a cohort of TFs collectively competes with the same histone octamer for access to a specific DNA sequence (Fig. 1) [18]. In addition, at some developmental enhancers there is evidence for step-wise licensing by lineage-determining TFs, also known as master regulators, or pioneer factors which directly bind nucleosomal DNA to prime enhancers for activation (Fig. 1) [18]. These factors can recruit chromatin remodeling activities, which then facilitate post-translational modification of histones

meaning that subsequent TF and coactivator binding is less strictly dependent on direct competition between TFs and nucleosomes [20]. Hematopoietic differentiation is controlled by key TFs, which regulate stem cell functions and differentiation [21]. During hematopoiesis mature blood cells are generated from hematopoietic stem cells (HSCs), which reside in the bone marrow [22]. HSCs generate progeny of intermediate repopulation potential, the progenitor cells, and they are capable of self-renewal with the production of additional HSCs [22]. Hematopoietic progenitor cells undergo further differentiation into mature cells of the lymphoid and myeloid branches [22]. The lymphoid lineage consists of T, B and natural killer (NK) cells, while the myeloid lineage comprises a number of distinct morphological, phenotypical and functional cell types such as different subclasses of granulocytes (neutrophils, eosinophils and basophils), monocytes-macrophages, erythrocytes, megakaryocytes and mast cells (Fig. 2) [23].

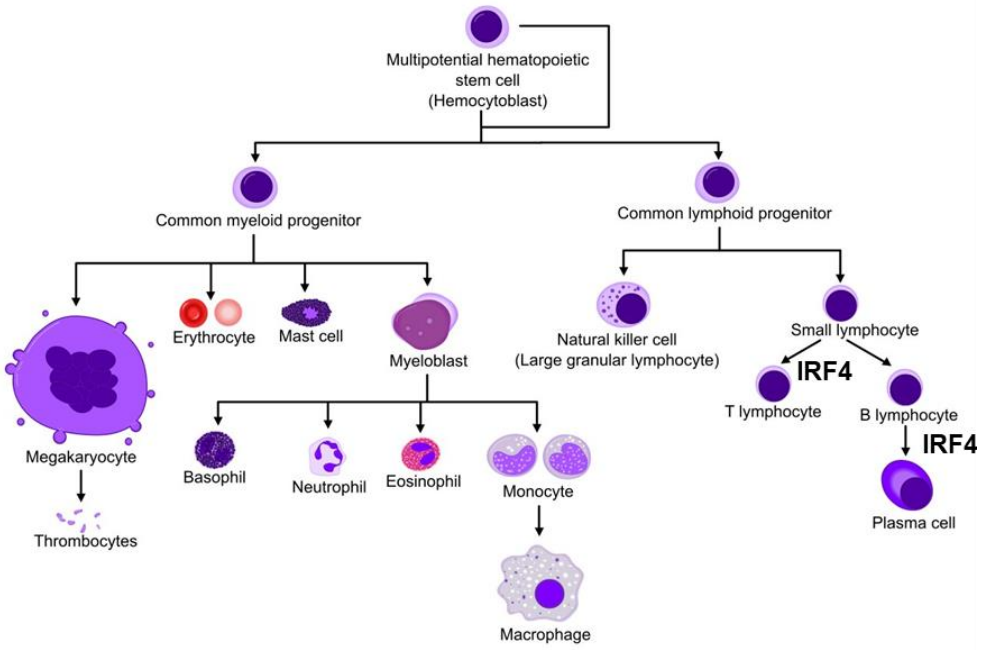


Fig. 2. Development of different blood cells from hematopoietic stem cells to mature cells. Adapted from [24]. IRF4 is implicated in B, T cell development and differentiation to plasma cell [25] [26].

This differentiation process is regulated by internal and external signalling events and by the activity of TFs at the endpoint of these signalling pathways [22]. In particular, the TF interferon regulatory factor 4 (IRF4) is required during the immune response for lymphocyte activation and the generation of immunoglobulin-secreting plasma cells (Fig. 2) [27].

1.2. Interferon regulatory factors

Interferon regulatory factors (IRFs) are TFs that mediate a multitude of functions including the differentiation and development of haematopoietic cells, regulation of apoptosis and host defence against pathogens [28] [29]. These TFs typically exist in their inactive monomeric form into the cytoplasm of an unstimulated cell [30]. The recruitment of adaptor molecules by the induction of the different signaling cascades promotes IRF activation and nuclear translocation [30]. This process ultimately causes the downstream production of cytokines, chemokines and other transcription factors that regulate innate and adaptive immune responses [30]. There are nine known mammalian IRFs (IRF1-IRF9). All IRF family members are composed of a multi-domain structure, which consists of: N-terminal DNA binding domain (DBD), a peptide linker (LK) and IRF activation domain (IAD) within the C-terminal activation domain (AD) [30]. A subset of IRF proteins (IRF3, 4, 5, and 7) contains an auto-inhibitory region (AR) in their structure that regulates their activity via different mechanisms involving conformational changes dependent or independent from phosphorylation events [31]. IRFs have an overall sequence identity of 2.68% but they have sequence identity of 13.6 % over 115 amino acids of the DBD domain (Fig. 3) [30].

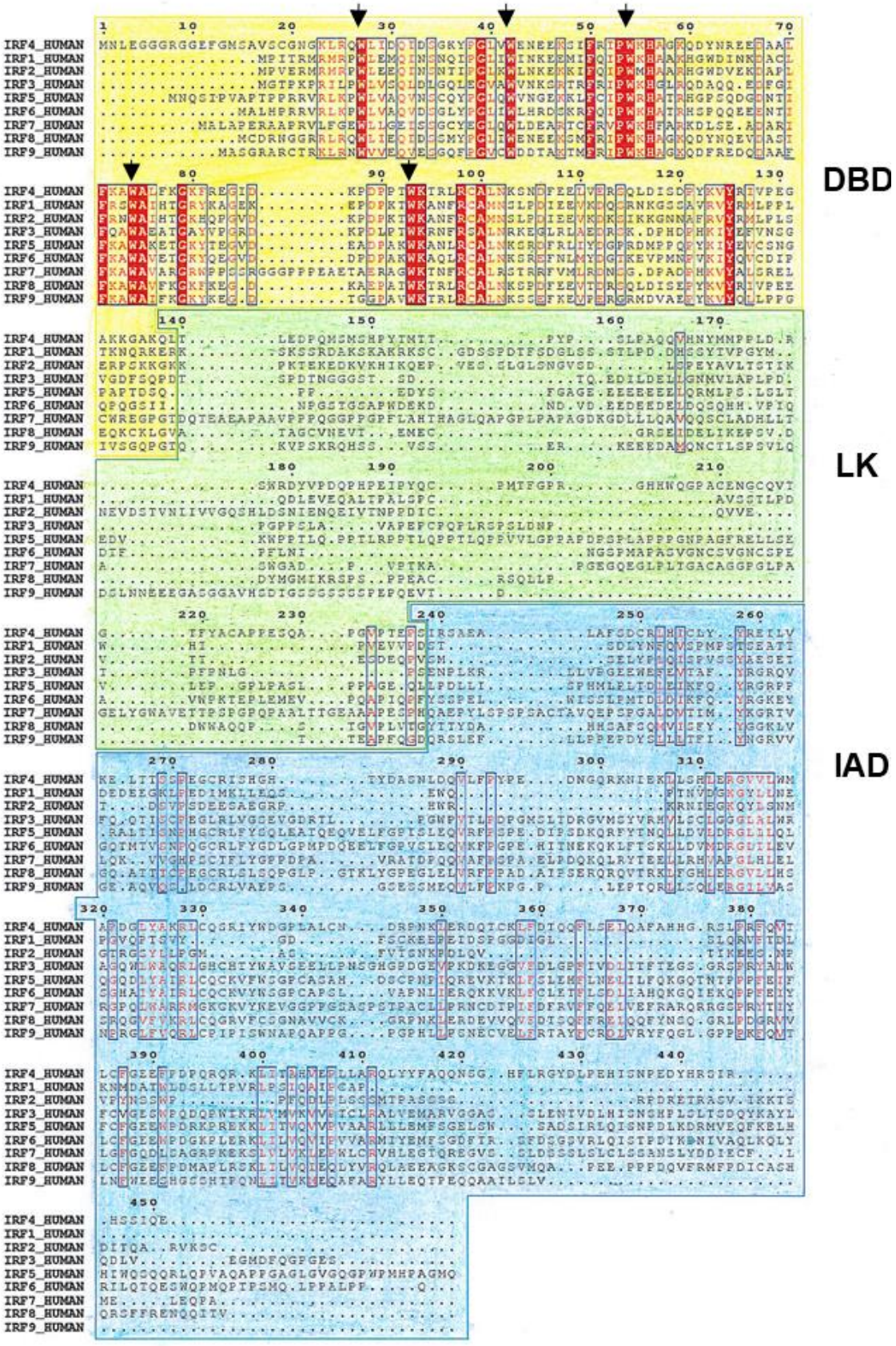

DBD
LK
IAD

Fig 3. Multiple sequence alignment of IRF proteins. IRF DBD domains are coloured in yellow, linker domains (LK) in green, interferon activation domains (IAD) in cyan. Black arrows indicate the five highly conserved tryptophan residues.

The DBD contains a highly conserved tryptophan repeat forming a helix-turn-helix motif that recognizes DNA sequences referred to as interferon (IFN) regulatory element (IRE, NAANNGAAA) and IFN sequence response elements (ISRE) (A/GNGAAANNGAAACT) (Fig. 3) [30]. The C-terminal region, on the other hand, shows diversity across all IRFs, which supports their distinct function(s), and could be potentially used for therapeutic inhibition that would provide specificity to each family member (Fig. 3) [30]. Activated IRFs might bind the ISRE as homo- and heterodimers. It has been reported that each of the IRFs forming a dimer, bind the ISRE half-site on opposite sides of the DNA, in a proximal orientation (Fig. 4) [32].

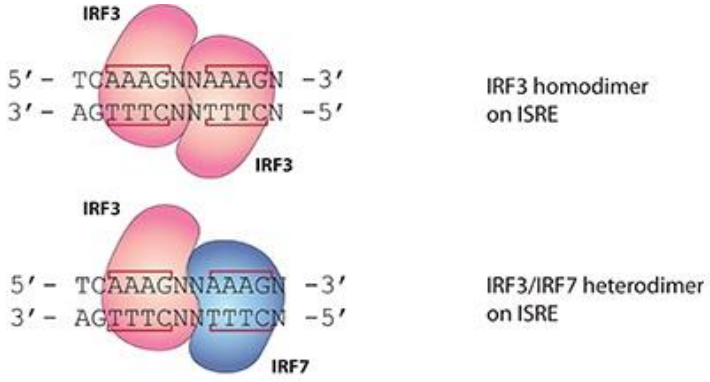


Fig. 4. Example of IRF3 homodimerization and IRF3-IRF7 heterodimerization to ISRE DNA sequences. Adapted from [33].

The crystal structure of IRF1 DBD in complex with the natural positive regulatory domain I (PRD I) DNA element (GAGAAGTGAAAGT) containing a GAAA core sequence showed that topologically the IRF1 DNA binding region is similar to a HTH DBD, and includes a α/β architecture and three large loops (L1-L3) connecting the different secondary structure elements (Fig. 5) [34].

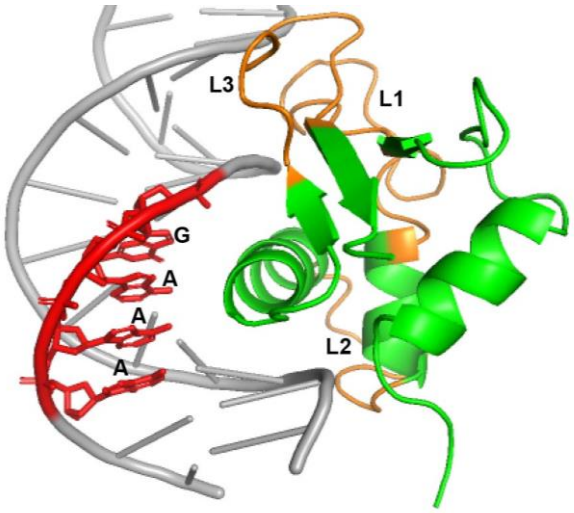


Fig. 5. Crystal structure of IRF1 DBD-PRD1 complex. IRF1 DBD region is characterised by α/β architecture interrupted by multiple loops, labelled as L1, L2 and L3. IRF1 interacts with the GAAA sequence of PRD I element. IRF1 is shown in green with L1, L2 and L3 loops show in orange. GAAA core bases are shown as sticks in red. Adapted from [34].

IRF1 binds to the PRD I element as a monomer and directly interacts with the GAAA core sequence (Fig. 5) [34].The IRF2 DBD-DNA complex reveals a very similar spatial structure. This could be explained by 80% sequence identity with IRF1 within the first 113aa, responsible for DNA binding [35] [36]. However, despite the high similarity between IRFs DBD structures and the fact that they all recognize the same consensus DNA binding site, there are significant differences in DNA binding affinities between family members [37]. IRF3 and IRF7 DBDs crystal structure analysis showed that this can be explained by differences in flexibility and conformational changes in the loops, in particular L1 [37]. In IRF3 this loop is disordered in the apo form and once it binds to DNA, it becomes ordered [37]. In contrast to IRF3, IRF7 L1 is ordered and stabilised by two hydrophobic residues (Phe45 and Leu50) that fold back into the core of the protein in the apo-form and during DNA binding a 2Å rigid body transition is observed [37]. Variable intrinsic loop flexibility of IRFs can cause their binding specificity and differences in DNA binding affinities [37]. Similar to homodimers, IRFs form heterodimers with one IRF on each side of the DNA helix, both contacting the full length ISRE (Fig. 4) [33]. An example of IRFs heterodimerization to ISRE motifs is represented by IRF3/IRF7. In particular, at specific stages during inflammation, IRF3 and IRF7

physically interact [38] [39] . In human fibroblast cell lines viral infection caused the activation of IRF7 and consequently upregulated MAP3K8, a kinase that inhibits IRF3 dimerization and promotes the formation of IRF3-IRF7 heterodimers. These heterodimers were necessary for limiting viral replication in vitro [40].

1.3. Interferon regulatory factor 4

IRF4 is a transcription factor with essential roles in in lymphocytes, where it regulates the development, affinity maturation, and terminal differentiation of B cells and has critical roles in diverse effector T cells [41]. IRF4 is characterised by a sequence identity of 38.74% with IRF8 that is another critical regulator of the immune system [42]. Due to its versatile function, IRF4 interacts with different DNA targets [42]. IRF4 binds to ISRE motifs as a homodimer in order to activate genes related to PC differentiation [43]. Conversely, it engages erythroblast transformation specific (Ets), interferon composite elements (EICE) and AP-1-IRF composite elements (AICE) as a heterodimer and requires PU.1, SPIB or BATF TFs in order to activate genes related to GC B cells [44] [45]. I refer the reader to chapter 2 of this thesis for further details about IRF4 structure and biology. Previous studies showed the crystal structure of IRF4 interacting with different DNA sequences [46] [47]. In particular, Escalante *et al.* reported the crystal structure of the ternary complex formed with the DNA binding domains of PU.1 and IRF4 on a composite DNA element from the IgL λ gene enhancer (5'-AAAAGGAAGTGAAACCA-3') containing overlapping Ets (GGAA) and IRF (AAxxGAAA) sites [46]. The DNA in the complex adopts an unusual S shape that juxtaposes PU.1 and IRF4 for selective electrostatic and hydrophobic interactions across the central minor groove [46]. The crystal structure of the IRF4-ISRE homodimeric complex showed that the complex formation is aided by a substantial DNA deformation with co-operative binding achieved exclusively through protein-DNA interaction [47]. Unlike the heterodimeric complex, no intermolecular interactions were observed between the interacting DBDs [47]. X-ray and small angle X-ray scattering studies showed that IRF4 IAD domain is composed of a set of β-sheets and loops that serve as the binding site for PU.1 [31]. Moreover, unlike the other IRF members, IRF4 is characterised by a flexible AR which is not folded into the IAD [31]. A key event prior to IRF

activation and nuclear translocation is post-translational modification that leads to conformational changes allowing for protein-protein interactions. In the case of IRFs post-translational modification causes the disruption of intramolecular association of the AR with the N-terminal DBD and IAD [48]. These conformational changes enable the IRFs to homo- or heterodimerize with each other or another molecule, thus allowing them to translocate to the nucleus and bind to DNA (with other cofactors), resulting in the regulation of gene transcription [49]. Specific phosphorylated residues in the C-terminus, referred to as the serine rich region (SRR), contribute to the stabilization of IRF dimers and interaction with DNA [30]. In particular, phosphorylation causes structural changes, including removal of the AR, liberation of the IAD and exposure of the C-terminus for further modification(s) and/or protein interaction [30]. IRF4 IAD presents two conservative serine residues, S446 and S447, that can be used as autoinhibition mechanism for IRF4 activity [50]. Dysregulation of IRFs can lead to either suppression or hyperactivation, both of which may contribute to disease development. In particular, dysregulation of IRF4 is associated with certain types of lymphoid and myeloid malignancies [51]. IRF4 overexpression is a hallmark of the activated B cell-like (ABC) type of diffuse large B cell lymphoma (DLBCL) and multiple myeloma (MM) [52] [53], and is also overexpressed in almost 100% cases of classical Hodgkin lymphoma (cHL), plasma cell myeloma and primary effusion lymphoma (PEL) [54]. Moreover, high levels of IRF4 protein exist in Epstein-Barr virus (EBV)-transformed cells and associated B cell lymphomas with Type 3 latency (III) [55], as well as in human T cell leukemia virus-1 (HTLV1)-infected cell lines and associated adult T cell leukemia (ATL) [56] [57]. Chromosomal translocation and genetic mutation of IRF4 are present in MM, peripheral T cell lymphomas [58] [59], and chronic lymphocytic leukemia (CLL) [60] [61]. These observations make IRF4 an attractive target for the development of new therapies to treat these disease conditions.

1.4. Multiple Myeloma

IRF4 is central to the genesis of multiple myeloma (MM) [52]. MM is an aggressive and incurable plasma cell neoplasm that accounts for 1%-1.8% of all cancers and is the second most common haematological malignancy. Despite the significant improvements in new therapeutic drugs, only 10%-15% of patients achieve or exceed expected survival compared with the matched general population [62]. Deregulation of TFs contributes to MM pathogenesis through: (1) direct TF modifications (e.g., mutations); (2) intrinsic genetic alterations or extrinsic stimuli within the bone marrow microenvironment that trigger signaling pathway-mediated TF activation or inhibition; (3) epigenetic changes in DNA methylation, histone modifications and non-coding RNAs; and (4) TF dependency on prolonged oncogene activity (“oncogenic addiction”) [63] [64] [52] [65]. In particular, myelomas are addicted to an abnormal regulatory network controlled by the TF IRF4, regardless of their molecular subtype and underlying oncogenic abnormalities [52]. Interference with IRF4 expression is lethal to MM cells making IRF4 a fundamental therapeutic target to be exploited [52]. In the UK, there are around 5,800 new myeloma cases every year. Myeloma incidence rates are projected to rise by 11% in the UK between 2014 and 2035, to 12 cases per 100,000 people by 2035 [66]. For males, myeloma European age-standardised (AS) incidence rates in the UK are projected to rise by 13% between 2014 and 2035, to 16 cases per 100,000 by 2035 [66]. For females, rates are projected to rise by 7% between 2014 and 2035, to 10 cases per 100,000 by 2035 (Fig. 6) [66].

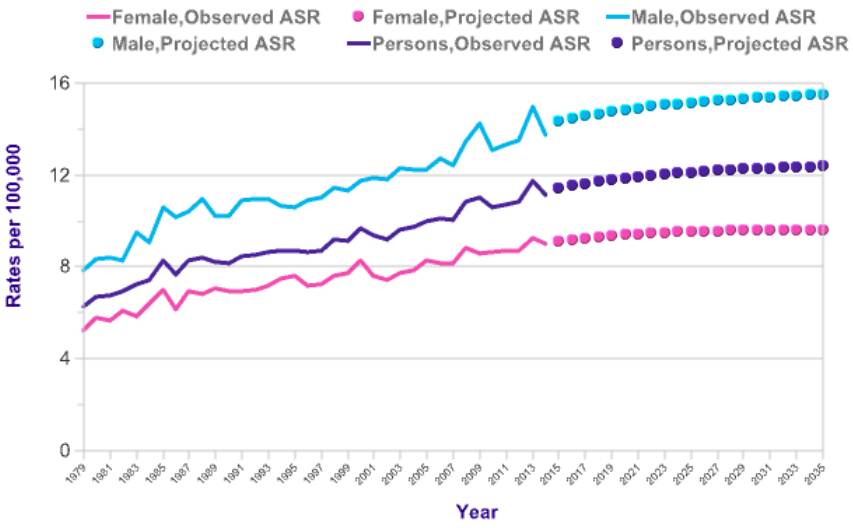


Fig. 6. Observed and Projected Age-Standardised Incidence Rates, by Sex, UK, 1979-2035. Adapted from [67].

MM is characterised by monoclonal plasma cells (PC) that accumulate in the bone marrow and produce monoclonal immunoglobulins (Fig. 7) [68].

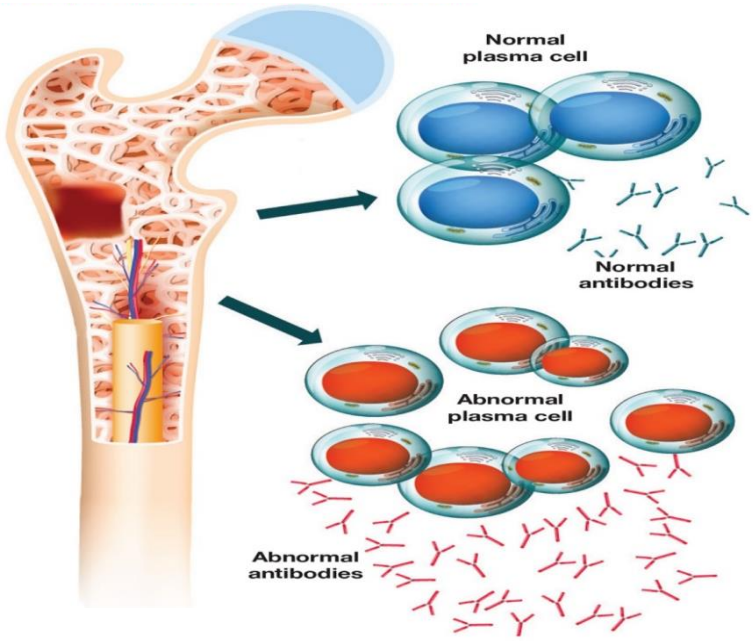


Fig. 7. MM in bone marrow. MM cells are abnormal PC that produce abnormal antibodies like monoclonal proteins. Adapted from [69].

Four major subtypes of MM characterised more than 80% of patients. They include hyper-diploid MM, t(11;14) MM, t(4;14) MM, and MM with translocations of t(14;16) or t(14;20) referred to as

MAF MM [70]. Secondary cytogenetic abnormalities such as deletion 17p, gain 1q, deletion 1p, deletion 13q, or monosomy 13 can occur in any of the primary cytogenetic types of myeloma, and can further modify disease course, response to therapy, and prognosis [71]. High-risk MM is defined by the presence of t(4;14), t(14;16), t(14;20), deletion 17p, gain 1q, or p53 mutation [72]. MM is an highly heterogenous diseases meaning that most of tumor PC share a common pool of mutations, but may differ by several subclonal mutations [73]. Several types of clonal evolution have been described in MM: stable evolution (identical genomic profile at diagnosis and relapse), linear evolution (appearance of novel mutations at relapse, but with the same mutational architecture), and branching evolution ("disappearance" of some mutations revealing evolution from a minor undetected subclone) (Fig. 8) [74].

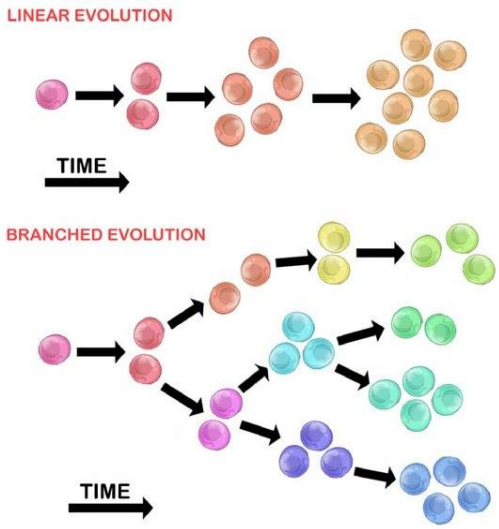


Fig. 8. Linear expansion and branched expansion of the clonal evolution tumor model. Adapted from [75].

There are different reasons why these subclones emerge and are selected. Nutriment accessibility or hypoxia can select the subclone(s) with the best fitness in the bone marrow niche. The selection can be caused also by a different proliferative capacity of subclones. Moreover, therapy targeting a mutation present only in a small fraction of tumor cells can result in a good benefit only for a specific subclone(s) (Fig. 9) [76] [77].

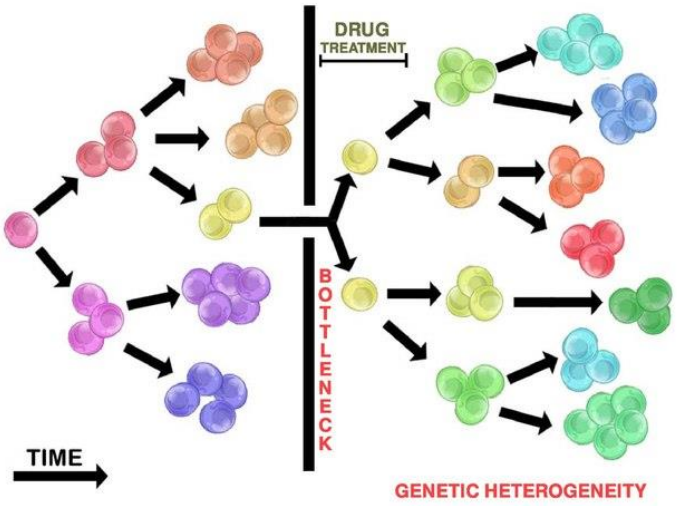


Fig. 9. Drug treatment induces a selection of resistant subclones that will survive and propagate to re-form a heterogeneous tumour. Adapted from [78].

Corre *et al.* showed that sequencing of MM patients treated with the same intensive treatment at diagnosis and at relapse result in no specific mutation or rearrangement, demonstrating that drug treatment has a nonspecific effect on clonal selection in MM [76]. On the contrary, chemoresistance and relapse can be caused by newly acquired mutations in myeloma drivers but also by (sub)clonal mutations preexisting to the treatment [76]. MM treatment typically includes drug therapy, such as targeted therapy and/or chemotherapy, with or without steroids [79]. Bone marrow/stem cell transplantation may be an option [79, 80]. Other types of treatments, such as radiation therapy and surgery, are used in specific circumstances [79]. The major classes of drugs include alkylating agents (melphalan,cyclophosphamide) corticosteroids (dexamethasone, prednisone), immunomodulatory drugs (thalidomide, lenalidomide, pomalidomide), and proteasome inhibitors (bortezomib, carfilzomib, ixazomib). The monoclonal antibodies daratumumab and isatuximab targeting CD38 play an important role in MM as well [71]. Other active approved compounds include elotuzumab, an IgG 1 monoclonal antibody targeting the SLAMF7 antigen; panobinostat, a histone deacetylase inhibitor; and selinexor, an inhibitor of exportin-1 (XPO1) [71]. A rapidly emerging and highly promising therapeutic option in MM is chimeric antigen receptor (CAR) T cell therapy, which has shown promising results in B cell malignancies [81]. B cell maturation antigen (BCMA), a tumor necrosis factor receptor superfamily

member 17 (TNFRSF17), is the most widely studied myeloma CAR target. It is an ideal antigen target because of its preferential expression on the PC but not on the HSCs [82]. BCMA is universally present in all MM cells and its overexpression carries important prognostic value [83] [84]. Clinical phase 2 studies using bb2121, a BCMA targeted CAR T cell therapy have shown an overall response rate (ORR) of 73% with significantly improved progression-free survival (PFS) improved (median PFS 8.8 months) in the relapsed/refractory multiple myeloma (RRMM) patients [85] [86]. Synthetic drugs are physiologically eliminated from the body over time, meaning that a repetitive administration is necessary to produce a durable response [87]. On the contrary, CAR T cells would be induced as long as targeted tumor antigen exists, thus making its therapeutic effect sustainable and rendering one day MM a highly manageable and curable disease [87]. Despite the promising results achieved by CAR T cell administration in MM in terms of response rates, the absence of a plateau corresponding with the absence of durable remissions is common to all studies [88]. Clinical experience with CAR T cell therapy has revealed several limitations of this technology such as lack of effectiveness, toxicities, antigen loss, interference with soluble proteins or manufacturing issues [89] [90]. Despite the introduction of new classes of cancer drugs, including immunotherapy drugs and monoclonal antibodies, MM remains an incurable disease. The majority of patients will relapse and will require additional treatment [91].

1.5. Targeting transcription factors in cancer

An interesting therapeutical approach in cancer is to target TFs playing important roles in the gene regulation of cancer cells (Fig. 10) [64].

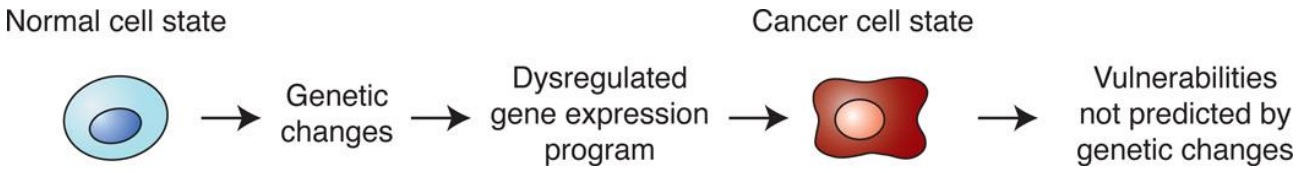


Fig. 10. Dysregulation program of a cancer cells create dependencies on transcriptional factors that make the tumor cells more sensitive to inhibition of these regulators than normal cells. Adapted from [64].

Extensive genome and transcriptome sequencing showed that cancer is characterised by deregulated expression of TFs [92]. Transcriptional activity is altered in numerous cancer types because of chromosomal translocations, gene amplification or deletion, point mutations and alteration of expression, as well as indirectly through non-coding DNA mutations that affect transcription factor binding [64]. TFs were previously viewed as “undruggable” [93]. This arose from the challenges associated with targeting either the protein-DNA or protein-protein interactions that mediate their function, as opposed to more tractable active sites of kinases or other enzymes [93]. Protein-protein interaction surfaces are typically flatter and are not characterised by the deep pockets present in enzyme active sites, making the development of small-molecule inhibitors more challenging [93]. However, some successes in targeting TFs suggest that this approach is achievable [94]. For example, inhibitors of the protein-protein interaction between p53 and its negative regulator MDM2 resulted in reduced proteasome degradation of p53. These inhibitors have shown in vivo activity against numerous cancers [95]. Another way to perform targeting of TFs would be by disrupting transcription factor-DNA binding with DNA binding compounds [93]. However, there is a profound shortage of small-molecule inhibitors that bind to a protein in order to inhibit protein-DNA binding. This is due to the fact that it is challenging to design drugs that can bind with specificity and potency to the highly positively charged and convex DNA binding interfaces found on TFs [93]. Another possible line of investigation when exploring inhibition of

protein is to look at allosteric inhibition [93]. Autoinhibition is a common property of many TFs where regions outside a functional domains (catalytic domain, DBD) bind to the functional domain to inhibit its activity [93] [96]. This process is regulated by post-translational modifications or protein-protein interactions [93]. The stabilization of the autoinhibited state has the potential to have broad utility to target TFs [93]. Indeed, the sequences of the elements mediating autoinhibition include a distribution of amino acids that more closely resembles that seen on the surfaces of other proteins, unlike the highly positively charged DNA binding interfaces [93]. That means that the likelihood of finding small-molecule inhibitors that can bind to these regions is much higher [93]. Another strategy to targeting TF could be the targeting of intrinsically disordered regions of TFs [93]. TFs frequently contain intrinsically disordered regions that do not form a stable 3D structure [98]. These regions become structured upon interaction with binding partners, a process referred to as coupled folding and binding [99] [93]. An analysis based on the structures of intrinsically disordered proteins when folded and bound to their partners suggests they actually have a higher proportion of potential cavities where a small molecule could bind than their folded counterparts [100]. This suggests that exploration of the druggability of these regions. These innovations in drug development hold great promise to perform the targeting of TFs in cancer.

1.6. Targeting transcription factors with bromodomain inhibitors

In eukaryotic cells TFs activate the expression of their target genes by recruiting multisubunit coactivator complexes, which use diverse biochemical mechanisms to activate RNA polymerase II [101]. One important class of coactivators possess lysine acetyltransferase (KAT) activity, which transfers acetyl groups from acetyl-coenzyme A to the epsilon amino group of lysine residues of various substrate proteins [101]. Lysine side-chain acetylation of many transcriptional regulators creates docking sites for proteins possessing acetyllysine binding/reader domains [101]. In this context, acetyllysine promotes the assembly of the transcriptional apparatus at enhancer and promoter elements [101]. The bromodomain is the most well known acetyllysine reader that is present in 46 different proteins encoded in the human genome [102]. A bromodomain is characterised by a left-handed bundle of four α helices, that is a small protein fold composed of 4 alpha helix oriented counter-clockwise [103]. The bromodomain includes also interhelical loops forming a hydrophobic binding pocket that engages in acetyllysine recognition [103]. The mammalian BET (bromodomain and extraterminal domain) protein family consists of four members, including the ubiquitously expressed BRD2, BRD3, BRD4, and the germ-cell-specific BRDT [104]. All four BET proteins possess two conserved bromodomains that preferentially bind to multiacetylated peptides (Fig. 11) [102]. In addition to two bromodomains, all four BET proteins have a conserved extraterminal (ET) domain that performs an effector role in transcriptional activation and in chromatin remodeling [101]. Moreover, BRD4 and BRDT possess a unique carboxyterminal motif (CTM), which binds to the serine/threonine kinase P-TEFb as an additional mechanism of gene activation [105].

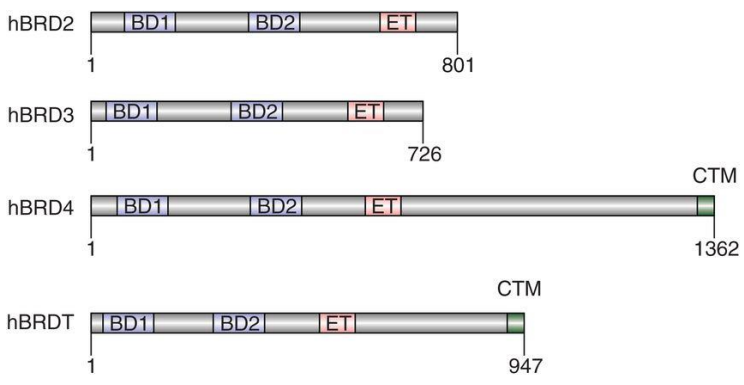


Fig. 11. Domain structure of the bromodomain and extra-terminal (BET) protein family. BD1 and BD2 indicate bromodomains while ET indicates the extra-terminal (ET) domain. Adapted from [101].

One of the most studied hallmarks of cancer is epigenetic dysregulation, which has been shown to induce tumorigenesis as well as resistance to standard-of-care treatments [106]. Bromodomains are druggable epigenetic targets, which has encouraged the discovery and development of several small-molecule inhibitors in recent years (Fig. 12) [107].

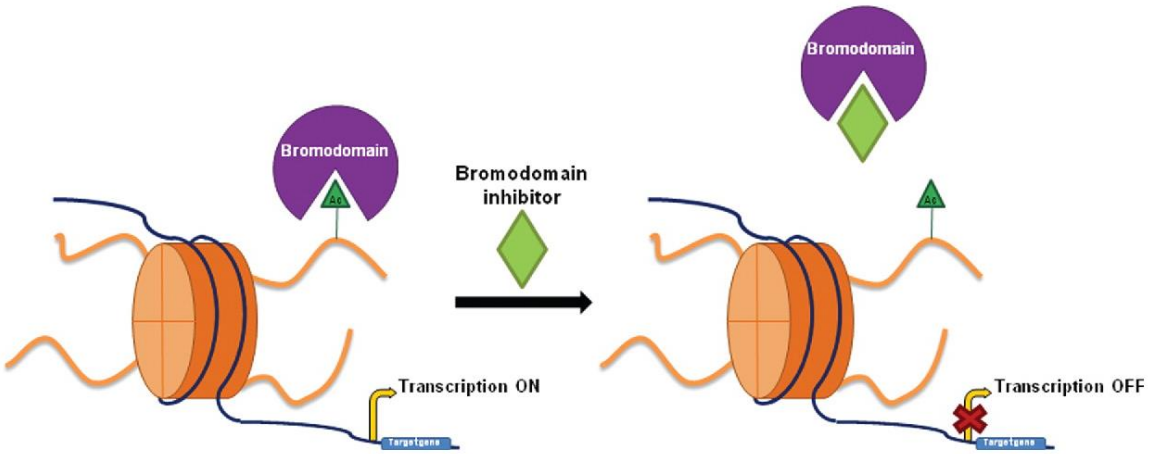


Fig. 12. Schematic picture of bromodomain inhibition. Bromodomains recognize acetylation marks in histone tails and recruit transcriptional machinery promoting target gene transcription. Bromodomain inhibitors prevent interaction between the bromodomain and the acetyl group, causing the downregulation of certain genes. Adapted from [107].

The first developed inhibitors of the BET bromodomain family were (+)-JQ1 (Fig. 13), reported by the Structural Genomics Consortium (SGC) and the Dana-Faber Cancer Institute [108], and I-BET762, reported by GlaxoSmithKline (GSK) [109] [110].

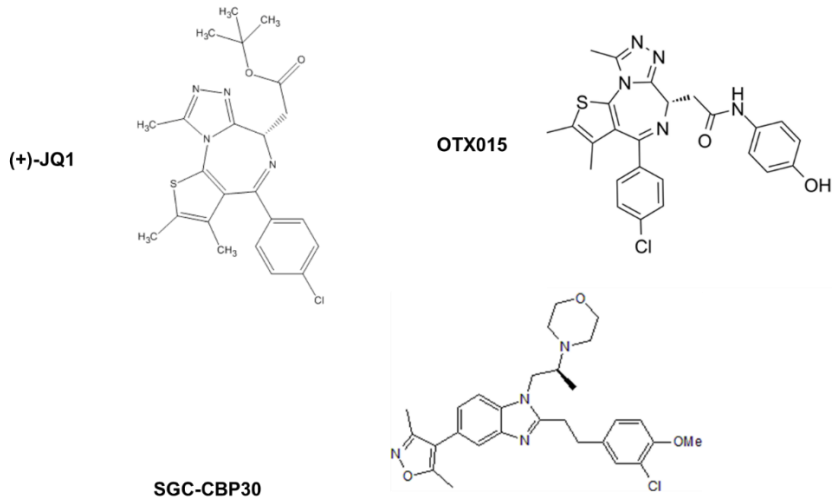


Fig. 13. Chemical structures of (+)-JQ1, OTX015 and SGC-CBP30.

A large number of studies were published showing the efficacy of JQ1 in hematological malignancies [111] [112] [113] [114] and solid tumors [61]. In particular, in these hematological cancers, BET inhibition by JQ1 downregulates *MYC* transcription and genome-wide MYC-dependent target genes, promoting cell cycle arrest and cellular senescence [107]. JQ1 does not discriminate between the two bromodomains within the same BET protein, nor among the four BET family members [108] [109]. Moreover, JQ1 is not being tested in clinical trials due to its short half-life [115]. OTX015 (MK-8628) is a selective BET inhibitor that competitively occupies the acetyl-binding pockets of BRD2/3/4, resulting in their release from active chromatin and suppression of downstream signaling to RNA polymerases [116]. OTX015 is the first BET inhibitor to have moved into the clinic, with phase three clinical trials initiated in hematologic malignancies (NCT01713582) (Fig. 13) [117] [118], selected solid tumors (NCT02259114) and glioblastoma multiforme (NCT02296476). BET family inhibitors have been extensively studied in recent years, but far less attention has been paid to the other bromodomains. Given that the latter are frequently in proteins with other epigenetic functions, such as histone acetyltransferases (HAT) activity, finding non-BET inhibitors will allow a better understanding of them [107]. This will also help identify new druggable targets for treating diseases like cancer [107]. The most promising non-BET BD targets are those of the transcriptional coactivators CREBBP and EP300 [119]. SGC-CBP30 has been developed to inhibit CREBBP and EP300 (Fig. 13). SGC-CBP30 was designed using nonselective bromodomain inhibitors as starting points and showed improvements in potency and some selectivity (12–140-fold) for the intended BDs over BET family BDs [120] [121]. Moreover, it has been shown that CREBBP and EP300 BDs, transcriptionally regulate the transcription factor IRF4 in MM, and treatment with SGC-CBP30 causes the downregulation of the IRF4-MYC axis and rapid cell death of MM cell lines [122]. The functional diversity and emerging “drugability” of bromodomain modules with small molecules has motivated a widespread interest in this class of proteins as therapeutic targets particularly in oncology, where BET proteins regulate the expression of key oncogenes and anti-apoptotic proteins [123]. The development of BET inhibitors has provided important insights into the key role of BET proteins in the transcriptional control of proto-oncogenes, and highlighted

the potential of these proteins as therapeutic targets [124]. A lack of biomarkers predicting sensitivity to BET inhibitors, coupled with the use of non-clinically relevant doses in preclinical studies, is limiting the application of these agents in clinical practice [124]. Further research and mechanistic studies will help to identify such biomarkers, and the development of novel, highly selective bromodomain inhibitors will help prevent toxicities [124].

1.7. Overview of the thesis

The work presented in this thesis focuses on the role of the protein IRF4 in the context of MM. IRF4 is a key transcription factor in the regulation of immune cells and is essential for PC differentiation [25]. IRF4 has also emerged to have a pivotal role in MM [52]. Knockdown experiments of this transcription factor caused apoptosis in MM cell lines showing an “IRF4 addiction in MM” [52]. Despite the essential role of IRF4 in MM, currently there are no therapeutic drugs targeting directly this transcription factor. This thesis analyses different strategies to target IRF4 in order to subvert MM. The second chapter of the thesis is the review entitled “IRF4 in multiple myeloma-Biology, disease and therapeutic target”. The first part of the review analyses the role of IRF4 in GC B cells and PC differentiation, describing the different IRF4 interacting proteins important in transcriptional circuitry of GC B cells and PC. In particular, it analyses how IRF4, depending on its protein levels, can interact with different DNA motifs and form heterodimers or homodimers in order to activate the expressions of specific genes related to GC or PC differentiation. The second part of the review outlines the mechanism of the IRF4-abnormal transcriptional network in MM. In particular, it describes the importance of the KDM3A-KLF2-IRF4 axis that is shown to be important for MM cell survival and homing to the bone marrow [125]. Moreover, it analyses the IRF4-MYC auto positive feedback loop that is essential for IRF4 overexpression and aberrant proliferation of MM cell lines [52]. In the last part of the review there is the discussion of the current and new therapies for MM and also an analysis of the direct target of IRF4. Interestingly, knockdown experiments and drug inhibition of IRF4 have shown significant decrease of MM cell viability [52] [122], indicating IRF4 as an attractive target for new therapeutic drugs. There are multiple ways in order to target IRF4. Previous studies have shown that

epigenetic alterations, including aberrant DNA methylation, histone modifications and chromatin remodeling, can be an attractive therapeutic targets in MM [126]. The third chapter of the thesis investigates the indirect target of IRF4 through its upstream epigenetic regulators. In this work we tested different epigenetic drugs JQ1, OTX015, SGC-CBP30 and ISOX-DUAL in different MM cell lines in order to disrupt the IRF4-MYC oncogenic loop. In particular, using different ranges of cell and molecular biology techniques like western blot and qPCR analysis together with network modelling, the results showed that the epigenetic regulator inhibitors tested were likely to kill MM cells by affecting MYC and its transcriptional pathway, not IRF4. In particular, both *IRF4* and *MYC* mRNA expression were significantly downregulated whereas only MYC protein expression was reduced after treatment with the epigenetic drugs. We noticed no significant reduction of IRF4 protein expression after treatment with different epigenetic inhibitors. This is due to the fact that IRF4 protein is characterised by a long half-life (more than 48 hours). These results clearly indicate that an indirect target of IRF4 cannot affect IRF4 protein expression suggesting that a direct target of this transcription factor would be more efficacious in order to subvert MM. IRF4 exerts its transcriptional activity by binding to specific DNA sequences called ISRE DNA sequences [43]. The fourth chapter of the thesis is the paper “Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4” published in the Acta Crystallographica Section F in 2021. This paper reports the use of native intrinsic phosphorus and sulfur single-wavelength anomalous dispersion (SAD) methods to solve the complex of IRF4 DBD bound to a specific ISRE sequence. When we started this study, only a nuclear magnetic resonance spectroscopy (NMR) structure was available and it did not lead to a clear phasing solution. The crystal structure showed three IRF4 DBD molecules bound to ISRE motif. The result was quite unexpected because previous studies suggested an IRF4 homodimerization to ISRE motifs [43]. This paper showed that native SAD phasing at long wavelengths can be used to solve the crystal structures of a large number of nucleic acid-protein complexes. To directly target IRF4 and to dissect the mechanism of ISRE binding in MM, the fifth chapter of the thesis presents the crystal structure of IRF4 DBD in complex with different ISRE DNA sequences. The different structures were solved by molecular replacement (MR) using the PHASER- Expert Mode

Molecular Replacement software in CCP4i2 [130]. From the different crystal structures, we can see that there is no evidence of IRF4 DBD protein dimerisation to the different ISRE sequences. In particular, as already shown by Sundararay *et al.* [47], IRF4 binds to ISRE sequences because of a specific DNA shape conformation achieved through protein-DNA contacts. Moreover, in order to determine the IRF4 DBD affinity to ISRE DNA sequences, we performed electrophoretic mobility shift assay (EMSA) and a more quantitative assay microscale thermophoresis (MST). Our data confirmed that IRF4 is characterised by low affinity to DNA [44] [131]. In particular we tested IRF4 affinity analysing its binding to different ISRE motifs characterised by different spacing between GAAA motifs, different upstream sequences and also testing single GAAA motifs. IRF4 affinity to single GAAA ISRE motifs was even lower than canonical ISRE sequences, suggesting that the sequences upstream and between the canonical ISRE motifs did not influence the interaction of IRF4 DBD to ISRE motifs but they affected the IRF4 DBD affinity to ISRE. These data are important in order to develop a small-molecules drug discovery programme to target IRF4.

2. Chapter 2. IRF4 in multiple myeloma—Biology, disease and therapeutic target

Leukemia Research 72 (2018) 52–58

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

IRF4 in multiple myeloma—Biology, disease and therapeutic target

Alessandro Agnarelli[a], Tim Chevassut[b], Erika J. Mancini[a,*]

[a] School of Life Sciences, Biochemistry Department, University of Sussex, Falmer, Brighton, BN1 9QG, United Kingdom
[b] Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom

ARTICLE INFO

Keywords:
Multiple myeloma
IRF4
Transcription regulation
Treatment
Drug discovery

ABSTRACT

Multiple Myeloma (MM) is an incurable hematologic malignancy characterized by abnormal proliferation of plasma cells. Interferon Regulatory Factor 4 (IRF4), a member of the interferon regulatory family of transcription factors, is central to the genesis of MM. IRF4 is highly expressed in B cells and plasma cells where it plays essential roles in controlling B cell to plasma cell differentiation and immunoglobulin class switching. Overexpression of IRF4 is found in MM patients' derived cells, often as a result of activating mutations or translocations, where it is required for their survival. In this review, we first describe the roles of IRF4 in B cells and plasma cells and then analyse the subversion of the IRF4 transcriptional network in MM. Moreover, we discuss current therapies for MM as well as direct targeting of IRF4 as a potential new therapeutic strategy.

1. Introduction

Multiple Myeloma (MM) is an aggressive and incurable cancer characterized by the clonal proliferation of bone marrow plasma cells. MM diagnosis follows the appearance of end-organ damage known as the CRAB criteria (increased calcium level, renal dysfunction, anaemia, and destructive bone lesions) but can also be diagnosed in presence of at least one myeloma defining event or MED (bone marrow plasma cells greater than or equal to 60%; serum free light chain ratio greater than or equal to 100 provided involved FLC level is at least 100 mg/L; more than one focal lesion on magnetic resonance imaging that is at least 5 mm or greater in size) [1]. MM represents approximately 2% of all cancers and about 10% of all hematologic malignancies [2] with a rising incidence estimated to be 6–10 cases per 100,000 persons per year. In the UK alone 5540 people were diagnosed and 3079 deaths were reported in 2016. The median age of patients at the time of diagnosis is about 65 years [2]. MM is considered a multistep disease since almost all patients with MM are characterized by an asymptomatic pre-malignant stage termed monoclonal gammopathy of undetermined significance (MGUS) and some patients by an intermediate asymptomatic but more advanced pre-malignant stage called smouldering multiple myeloma (SMM) [2,3].

Therapies used in the treatment of MM include initial therapy, autologous stem cell transplantation (when possible), consolidation/maintenance therapy, and treatment of relapse [4]. The most common regimes for MM initial therapy consist of a combination of drugs including immunomodulatory drugs (IMiD) (thalomide, lenalidomide), corticosteroids (dexamethasone) and proteasome inhibitors (PI) (bortezomib). Current treatments have dramatically improved the median overall survival of patients, however MM usually relapses with patients refractory to both IMiDs and PIs. In the last few years, treatment of relapsed refractory MM improved because of the introduction of pomalidomide, another immune-modifying drug, monoclonal antibodies daratumumab and elotuzumab, the histone deacetylase inhibitor panobinostat and new-generation proteasome inhibitors carfilzomib and ixazomib [4]. However with a median duration between MM diagnosis and relapse of 3.1 years and a median overall survival following relapse of 13 months [5], there is a clear need for new treatments to overcome the dismal survival rates of MM.

Interferon Regulatory Factor 4 (IRF4) is a transcription factor belonging to the interferon regulatory factor (IRF) family. IRFs are transcription factors playing a critical role in the regulation of immune responses, immune cell development, cell growth regulation and metabolism [6]. IRF4 is a critical regulator of the immune system and it is essential for PC differentiation [7,8]. IRF4 has also emerged as the master regulator of an aberrant and malignancy-specific gene expression programme in MM, where it is found to be overexpressed often as a result of activating mutations or translocations [9,10]. Knockdown experiments of IRF4 have shown a dramatic decrease in the viability of MM cells [10]. Enforced expression of miR-125b-5p, a miRNAs predicted to target the 3′ UTR of IRF4 mRNA, inhibits the growth and survival of MM cell lines [11]. Yet IRF4 has not been the direct target of therapeutic drug discovery programmes.

Here we describe the role of IRF4 during normal PC differentiation,

* Corresponding author.
E-mail address: erika.mancini@sussex.ac.uk (E.J. Mancini).

https://doi.org/10.1016/j.leukres.2018.07.025
Received 22 May 2018; Received in revised form 30 July 2018; Accepted 31 July 2018
Available online 03 August 2018

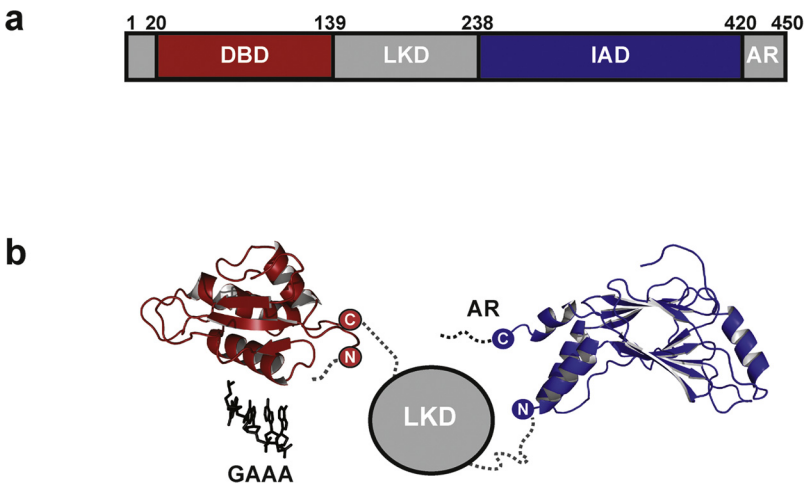


Fig. 1. Overall structure of IRF4. (a) Schematic representation showing the domain arrangement of IRF4: DNA binding domain (DBD, red), linker domain (LKD), Interferon Activating Domain (IAD, blue), auto inhibitory region (AR). **(b)** Cartoon representation of the crystal structure of the IRF4 DBD bound to GAAA consensus motif [12] and IAD (PDB: 5BVI). The LKD domain, which is thought to be folded into a domain structure, interacts with both DBD and IAD domains [6]. The AR domain is flexible and does not interact with either IAD or DBD domains [6]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

the mechanism of IRF4-driven deregulation of transcriptional activity in MM and we discuss the value of new therapeutic avenues to treat MM, including the direct targeting of IRF4.

2. IRF4 structure and transcriptional activity

IRF4 is characterized by an N-terminal tryptophan pentad repeat DNA-binding domain (DBD) connected to a C-terminal interferon activation domain (IAD), critical in mediating protein-protein interactions *via* a linker domain (LKD) (Fig. 1) [6]. The DBD domain resembles a winged helix-turn-helix motif with a 3-helix bundle (α1, α2, α3), a 4-stranded antiparallel beta-sheet (β1–β4) and two large loops (between β2 and α2 and α2 and α3) (Fig. 1b). The third helix slots into the major groove of the 5′-GAAA-3′ subsequence and is the major determinant of sequence-specific binding through contacts made by arginine residues on the hydrophilic face with the phosphate backbone (Fig.1b). Three of the five invariant tryptophan residues contact DNA [12,6].

Unlike other IRF protein, IRF4 binds DNA with low affinity and requires further protein-protein interactions to bind DNA [6]. IRF4 is essential for the expression of both GC B cell-specific and PC-specific genes and the low affinity for DNA is thought to be central to this role. Depending on its protein levels, IRF4 binds DNA as a heterodimer or a homodimer to different motifs, each motif uniquely activating the expression of genes related to GC B cell or PC differentiation. At low protein levels, IRF4 binds as a heterodimer either the Ets-IRF composite elements EICEs (GGAANN(N)GAAA) with PU.1 (Fig. 2a, d) or the AP-1-IRF composite elements AICEs (GAAATGAGTCA or GAAANNNNTGAG TCA) with AP-1 family such as Batf (Fig. 2b, e) [13,14]. During the differentiation of B cell into PCs protein levels increase and IRF4 binds as a homodimer to the interferon sequence response elements ISREs (GAAANNGAAA) (Fig. 2c, f) [15].

The low DNA binding affinity of IRF4 has been attributed to the inhibitory activity of the last 30 residues of the IAD domain [16,13] (Fig. 1a). It has been postulated that this auto-inhibitory region (AR) prevents the DBD from binding to DNA, whilst DBD interactions with transcription factor partners would release AR inhibition [13]. This hypothesis however does not explain how release of the inhibition would occur when IRF4 binds to ISRE sequences as a homodimer. Recent structural studies have shown that the AR region is a flexible unstructured peptide that does not dock into the IAD helical bundle, as seen in IRF3 [6,17,18]. Furthermore, the diversity in sequence homology and length of the IRF4 AR region, suggest that alternative mechanisms could induce IRF4 dimerization on DNA. Small-angle X-ray scattering (SAXS) studies of full length IRF4 suggests that the linker region (LKD) connecting the DBD and IAD domains most likely adopts a folded conformation able to interact with the domains located at either end of the molecule and that it may therefore play a role in the regulation of IRF4 activity [6] (Fig. 1b).

3. IRF4 role in transcriptional circuitry of GC B cells and plasma cells

IRF4 is the master regulator of two mutually antagonistic programmes of B and PC cells gene expression [15]. B cells play a fundamental role in the humoral immune response. During antigen-dependent activation, B cells can rearrange the constant region of the IgH region yielding antibodies with different effector functions by a process called class-switch recombination (CSR). Moreover, after antigen-dependent activation, mature B cells undergo somatic hypermutation (SHM), a process that alters the variable regions of the immunoglobulin in order to select B cells producing high affinity antibodies. SHM leads to the affinity maturation of B cells in germinal centres (GCs) that are transient structures within secondary lymphoid organs where B cells are selected based on their ability to produce high-affinity antibodies [19]. GCs are characterized by two compartments: the dark zone (DZ) where B cells proliferate extensively undergoing SHM and the light zone (LZ) in which B cells are selected on the basis of their affinity for the antigen. The GCs ultimately produce memory B cells and high-affinity, long-lived PCs characterized by high level of antibody secretion [20]. Molecular alterations occurring during early and late phases of B cell development can lead to the generation of lymphoid tumours.

According to the ‘‘kinetic model’’ proposed by Ochiai et al. [15]. IRF4 regulates CSR, SHM, the generation of GC B cells and PC differentiation in a temporal and dose-dependent manner [8,7,15]. Specifically, IRF4 levels appear to define cell fate decisions by coordinating binding partner- and DNA-binding activity.

In the early stages of the GC reaction IRF4 is present at low levels and its binding to AICE and EICE motifs up-regulates activation-induced cytidine deaminase (AID) expression. AID (encoded by the *AICDA* gene), an enzyme that creates mutations in DNA by deamination of cytosine base, is absolutely necessary for CSR and SHM [21]. IRF4 also activates B-cell lymphoma 6 protein (BCL6), a transcriptional repressor mainly required for GC formation and antibody affinity maturation (Fig. 3a) [15,22]. On the other hand, elevated levels of IRF4 during PC differentiation favour binding of IRF4 to the ISREs of direct target genes such as *PRDM1*, which encodes protein PRDM1 (also known as BLIMP1) a key component of the PC differentiation transcription programme [23,24]. The shift to ISREs binding therefore mediates activation of *PRDM1* and repression of *BCL6*, bringing the GC programme to an end and promoting the differentiation into PCs (Fig. 3b, c).

IRF4 is absolutely required for GC formation. Studies looking at the effect in mice of B cell specific knockdown of IRF4, show a failure in GC formation caused by insufficient induction of BCL6 *and AID* [15,8,7]. BCL6, which is highly expressed in GC B cells, facilitates their rapid

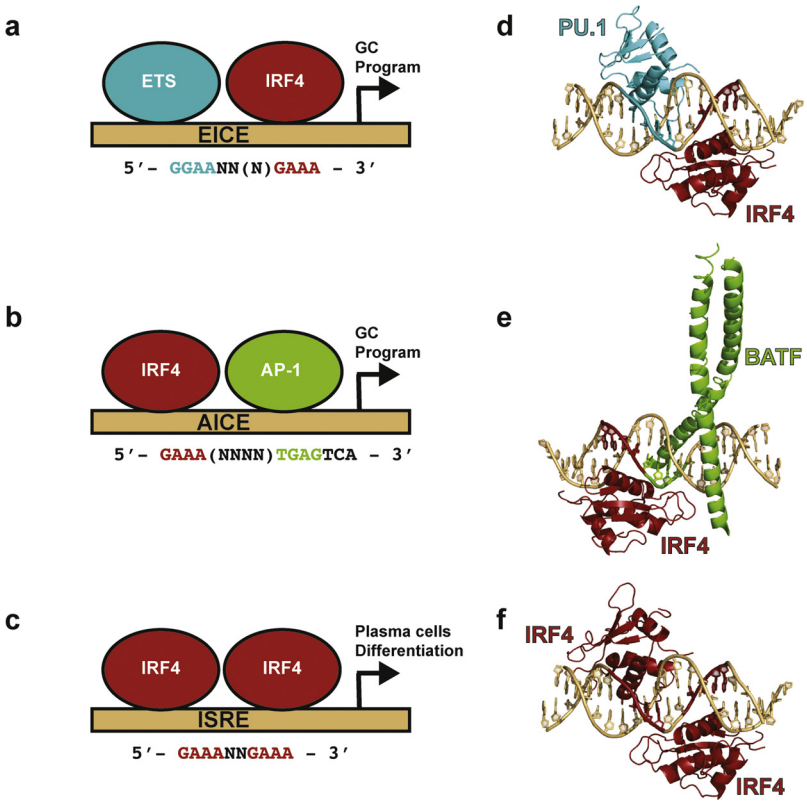


Fig. 2. IRF4 cooperative DNA binding and transcription outcome. (a–c) Schematic representation of the different IRF4 DNA binding modes. IRF4 binds the affinity high affinity composite DNA binding motif ETS–IRF (EICE) with members of the Ets family **(a)** or AP-1-IRF (AICE) with members of the AP-1 family **(b)**. At high concentrations IRF4 binds the DNA interferon response elements (ISRE) as a homodimer **(c)**. The different outcomes of cooperation between IRF4 and other transcription factors or itself are listed. The crystal structure of the PU.1 (teal) – IRF4 (red) DNA binding domains bound on a high affinity EICE composite motif is shown in a cartoon representation **(d)**. A model of the BATF-Jun (green) – IRF4 (red) DNA binding domains complex bound on an AICE motif **(e)**, based on the known structures of the Jun-ATF2-IRF3B complex in the interferon-β enhanceosome (PDB: 1T2K) and the PU.1-IRF4 complex on the IgL λ gene enhancer [12], is shown in a cartoon representation. The model was built using AICE motifs from Bcl11b (0 bp spacing) loci. A model for the IRF4 DNA binding domain homodimer (red) bound on an ISRE motif **(f)** based on the structure of the PU.1-IRF4 complex on the IgL λ gene enhancer [12] is shown in a cartoon representation. The model was built using an ISRE motif with 2bp spacing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

proliferation in the dark zone through repression of cell cycle controlling genes, such as p53 and p21, and through inhibition of the DNA damage response that facilitates tolerance for high rates of SHM [25,26]. *In vivo* studies in mice have shown that transient expression of *IRF4* directly activates but does not maintain *BCL6* expression in GC B cells [15], suggesting that *IRF4* might play an essential role for the establishment but not for the maintenance of GC state

Once the germinal centre is formed, IRF4 expression needs to be inhibited to avoid premature differentiation into PCs [15,22]. Evidence suggests that BCL6 itself might down-regulate *IRF4* in GC B cells by sustaining the expression of Microphthalmia-associated transcription factor MITF, a known suppressor of IRF4 [27]. *In vivo* studies showed that during germinal centre B cell differentiation, BCL6 represses BLIMP1 levels in cooperation with BACH2, a transcriptional repressor expressed during B cell differentiation (Fig. 3a, b) [28]. Specifically, BCL6 and BACH2 cooperate in regulating B cells GC transcriptional programme by forming a complex and recruiting each other to their respective *PDRM1* DNA binding sites [28]. Moreover, *in vivo* studies showed that BCL6 directly inhibits *c-Myc* (*MYC*) expression and confirmed the absence of *MYC* expression in the centroblasts located in DZ of the GC (Fig. 3b) [29,30]. *MYC* promotes cell growth and proliferation of most cell types, therefore *MYC* repression in active proliferating DZ B cells may explain the reduced number of cell divisions in DZ which allows for the affinity maturation process in the DZ of the GC [29]. On the other hand *MYC* is expressed in LZ GC B cells to enable their re-entry into the DZ and the continuation of the GC reaction (Fig. 3b) [29].

During PC differentiation, high levels of IRF4 induce *PRDM1* expression and together they repress *BCL6* expression to terminate the GC programme (Fig. 3b) [23,24,31]. Once PC differentiation has taken place, BLIMP1 enhances *IRF4* expression and represses *MYC* transcription causing an arrest in the PC cell cycle, as MYC is required for cell proliferation and growth (Fig. 3c) [8,31,32]. During PC differentiation, BLIMP1 also inhibits the expression of master regulator of B-cell identity protein PAX5, which during B cell development regulates IRF4, BACH2 and AID expression (Fig. 3a) [33–36]. The BLIMP1-mediated inhibition of PAX5 causes in turn expression of X-box binding protein 1 (XBP1), a transcription factor required for PC development, which induces Unfolded Protein Response (UPR) target genes [36] (Fig. 3c). In PCs, the UPR functions as a physiological pathway and it is activated during the early differentiation that precedes the high immunoglobulin expression [37]. IRF4 but not BLIMP1 is essential for PC survival. BLIMP1 instead sustains immunoglobulin secretion as Blimp-1-deficient PCs retain their transcriptional identity but loose their ability to secrete antibody [38].

Recently, a novel mechanism of IRF4-dependent gene repression during PC differentiation involving newly identified DNA binding motifs called ZICE, was described [39]. Zinc finger–IRF composite elements ZICE (GGGAANNNGAAA), composed of the zinc finger motif (GGGAA) and the IRF motif (GAAA), embed an ISRE motif which allows IRF4 to bind the ZICE sequence as a heterodimer with Ikaros or the ISRE sequence as a homodimer. Surprisingly, and despite the high levels of IRF4 during PC differentiation, IRF4 is more efficiently recruited to ZICE motifs in the presence of Ikaros. Crucially, the ZICEs were identified among a subset of IRF4 target genes whose expression is reduced upon PC differentiation, including *Ebf1* a positive regulator of B cell activation and GC reactions [39]. This report expands the number of transcription factors that partner with IRF4 to orchestrate GC B cell and PC differentiation and raises the question of how the delicate balance of transcription factors is accurately sustained.

4. IRF4 and multiple myeloma

4.1. IRF4 transcriptional network in multiple myeloma

IRF4 plays a central role in the pathogenesis of MM [9,10,40]. Chromosomal translocation t(6;14)x(p25;q32), which juxtaposes the immunoglobulin heavy-chain (IgH) locus to IRF4, is recurrently found in about ~21% of MM cases [41,42]. An identical mutation in the DNA binding domain of IRF4 (K123R) was initially found in two MM patients in a parallel sequencing of 38 tumour genomes [43]. Subsequently a

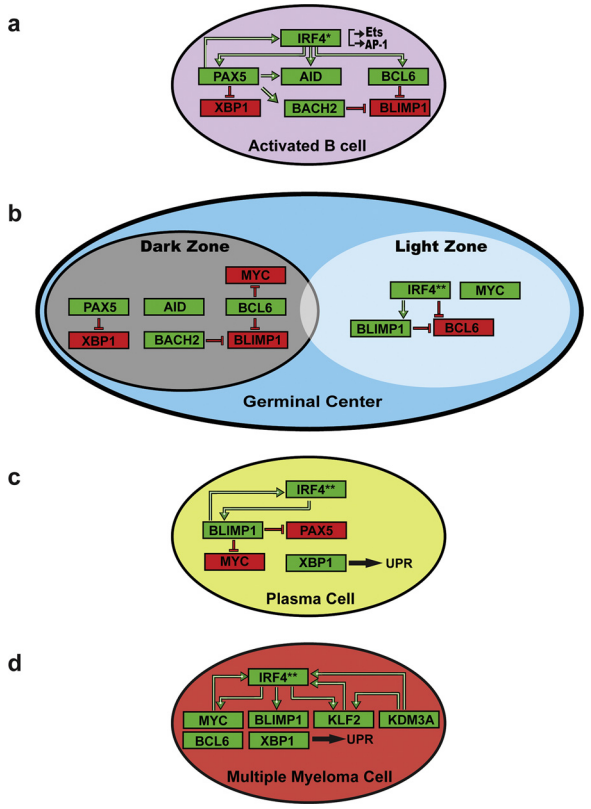


Fig. 3. IRF4-transcriptional networks in activated B cells, Germinal Centers, Plasma Cells and Multiple Myeloma Cells. Schematic representation of IRF4-transcriptional networks where green and boxes denote actively expressed and repressed protein respectively **(a)** Schematic representation of the IRF4-transcriptional network in activated B cells where IRF4 is expressed at low levels (*). **(b)** Schematic representation of the IRF4-transcriptional network in GCs. IRF4 is not expressed in the Dark Zone whilst it is expressed at high levels (**) in the Light Zone. **(c)** Schematic representation of the IRF4-transcriptional network in a plasma cells where IRF4 is expressed at high level (**). **(d)** Schematic representation of the aberrant IRF4 transcriptional network in myeloma cells where IRF4 is overexpressed (**). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

massively parallel sequencing of paired tumour/normal samples from 203 MM patients identified four more mutations in the DNA binding domain of IRF4, one being L116R and the other three being K123R, establishing K123R as a recurrent "hot spot" mutation in IRF4 [44]. IRF4 is highly expressed in MM patients when compared to healthy PCs and an important prognostic marker for MM with longer survival in patients with low IRF4 expression [40]. Overexpression is consistent with an oncogene role for IRF4 in MM and its mutations and translocations are thought to be activating. MM tumours are however known to be highly heterogeneous and in many MM cases IRF4 is overexpressed without genetic alterations [10,9]. In the context of MM, IRF4 is known to up-regulate over 100 genes that are quiescent in healthy PCs. Most of these genes, among them *Stag2*, *CDK6*, and *MYC* are associated with cellular growth and survival [10]. A study utilizing small hairpin RNAs (shRNAs) showed that *IRF4* silencing results in loss of cell viability in 10 different MM cell line models (representing different MM subtypes most of them lacking IRF4 genetic abnormalities) suggesting that MM cells depend on the ability of IRF4 to sustain an aberrant gene transcription programme [10]. Additionally, IRF4 downregulation and impairment of downstream effectors (such as BLIMP1) by enforced expression of miR-125b-5p, an miRNA that is predicted to target the 3′ UTR of IRF4 mRNA, affects growth and survival in patient-derived MM cells and MM cell lines [11].

This dependency has led to proposal of the "non-oncogenic addiction" of MM cells to IRF4 where the aberrant functions of normal genes, which themselves are not classical oncogenes, is required for cancer cells survival [10]. ChIP-chip analysis showed that the regulatory network that IRF4 controls in MM includes genes involved in many cellular process like cell cycle regulation, membrane biogenesis, cell death regulation, PC function [10]. This regulatory network does not reflect the genetic programme of normal PCs and instead more closely resembles that of antigen stimulated mature B cells [10].

Emerging evidence suggests that MM is characterized by a plethora of epigenetic alterations [45] and that this altered landscape can allow IRF4 access to loci that are usually not accessible in normal PCs [10]. Recent studies showed that MM is characterized by global DNA hypomethylation that correlates with disease progression and poor prognosis [46–50]. Promoter hypomethylation of specific genes such as NOTCH ligand JAG2 has been reported in MGUS, MM patients and MM cell lines suggesting DNA hypomethylation could be a an early event in MM pathogenesis [51]. Other studies reported an aberrant DNA hypermethylation of promoter regions of different tumour suppressor genes in MM [52]. Specifically, DNA hypermethylated regions were associated with intronic enhancer regions harbouring binding sites of B cell-differentiation transcription factors (TFs) like PAX5 [53]. As these enhancers are found to be hypermethylated in embryonic stem cells (ESCs) and to become gradually demethylated during differentiation into plasma cells, this suggests that MM cells are characterized by epigenetic features of undifferentiated cells [53].

A direct IRF4 target of particular interest in MM is Krϋppel zinc-finger family transcription factor KLF2, a negative regulator of pre-B cell clonal expansion and B cell activation [54] (Fig. 3d). Ohguchi et al. showed that knockdown of KLF2 caused apoptosis of MM cell lines indicating that KLF2, like IRF4, is essential for MM cells [55,10]. Interestingly, later studies demonstrated that KLF2 and IRF4 are activated by, and in turn activate KDM3A. KDM3A is a member of the Jumonji C-domain-containing histone demethylases, which catalyses the removal of H3K9 mono- and di-methylation (H3K9me1 and H3K9me2) [56]. The KDM3A–KLF2–IRF4 auto positive feedback loop was shown to be important for MM cell survival and homing to the bone marrow [55]. As KDM3A regulates KLF2 and IRF4 expression through its H3K9 demethylation activity, targeting the enzymatic activity of KDM3A could therefore open an interesting therapeutic window [55].

Another direct target of IRF4 in MM is MYC (Fig. 3d) [10,57,11]. Unexpectedly, IRF4 itself is also a direct target of MYC transactivation, generating an auto regulatory circuit in MM [57,10,58]. ChiP assay experiments showed that IRF4 and MYC regulate each other in MM cell lines, creating a positive regulatory loop resulting in an aberrant proliferation of MM cells [10] (Fig. 3d). MYC expression in MM cells is unusual since normal PCs do not express MYC due to repression by BLIMP1 [32,59] (Fig. 3c). BLIMP1 ensures the survival of MM cells by the interaction with Aiolos [60]. Particularly, Aiolos-BLIMP1 interaction plays an important role in MM cells survival, probably through the collaborative down-regulation of pro-apoptotic genes [60]. Moreover, Gyory et al. and Ocana et al. reported that MM cell lines overexpressed an alternative isoform of BLIMP1, called BLIMP1β, when compared to normal PCs [61,62]. BLIMP1β lacks the first 101 amino-terminal residues [61] and has a disrupted PR domain, a domain with similarities to SET domains found in Histone methyltransferases [61,63]. In addition, BLIMP1β is characterized by a diminished capacity to repress target genes, like *MYC* [61]. The expression of the truncated protein BLIMP1β could explain the inability of BLIMP1 to silence MYC in MM cells.

Similarly, *BCL6* that does not usually express in normal PCs because of inhibition by BLIMP1, is instead up regulated in MM cells in the bone marrow microenvironment [23,64] (Fig. 3d). *BCL6* over expression in MM cells, which is modulated at least in part *via* Janus kinase/STAT3 and canonical nuclear factor-κB pathways, can attenuate the DNA Damage Response, conferring a selective advantage to MM cell growth

[64,65]. More studies are required to understand how over expression of BCL6, a transcriptional repressor of MYC, does not lead to MYC inhibition in MM.

4.2. Targeting the IRF4 transcriptional network in multiple myeloma

Multiple myeloma (MM) is a clonal PC malignancy characterized by the growth of tumour cells in the bone marrow and an aggressive clinical course [66]. During the past decade, the advent of new generation proteasome inhibitors and immunomodulatory agents has improved the treatment of MM [67], however it remains an incurable disease. Almost all patients with MM who survive initial treatment will eventually relapse and require further therapy [67].

Targeting IRF4 is an attractive and broadly applicable therapeutic option for MM as various studies have shown that blocking its expression or interfering with its transcriptional network has profound consequences on the survival of MM cells [10,11,55]. This hypothesis is arguably validated by the observation that mice with only one *IRF4* allele are phenotypically normal, whilst 50% decrease of IRF4 at both the mRNA and protein levels causes MM cell death [10,68]. An IRF4-directed therapy could therefore kill MM cells with little or negligible effect effects on normal cells making it potentially an exciting therapeutic avenue for MM patients. However, efficient *in vivo* strategies aimed at directly blocking IRF4 are still lacking.

IRF4 has been targeted indirectly *via* targeting of its up-stream transcriptional regulators IKZF1 (Ikaros) and IKZF3 (Aiolos). Immunomodulatory drugs (IMiDs) like lenalidomide and thalidomide have been shown to have potent anti-tumour activities in MM resulting from IMiDs/Cereblon-mediated selective degradation of Ikaros and Aiolos, leading to a decrease of IRF4 and MYC expression [69–71].

Another way to indirectly target IRF4 is through its upstream epigenetic regulators. Panobinostat, a pan-Histone DeACetylase Inhibitor (HDACi) has been approved to treat MM patients by the FDA's accelerated approval programme. By activation of caspase-3 and cell apoptotic programmes, Panobinostat leads to downregulation of HO-1/IRF4/MYC mRNA and protein expressions in MM cells, where Heme oxygenase (HO-1) itself regulates IRF4/MYC expression [72]. Panobinostat was also found to increase the acetylation of histone H3K9 and induce the apoptosis of MM cells by IRF4 inhibition [72]. Finally, combining panobinostat with lenalidomide synergistically inhibited MM growth by facilitating caspase-3-mediated HO-1/IRF4/MYC degradation [72]. Treatment with panobinostat in combination with proteasome inhibitors bortezomib or carfilzomib was also shown to cause MM cell death by caspase-8 activation and IRF4 and MYC decreased expression [73].

Inhibition of transcriptional co-activator CBP/EP300 *via* its bromodomain selectively abrogates the viability of multiple myeloma cell lines as a result of transcriptional suppression of IRF4 and of its target genes [74]. In particular, CBP/EP300 bromodomain inhibition caused down-regulation of *MYC*, suggesting that CBP/EP300 plays an important role in the regulation of the IRF4/MYC axis in MM [74]. More recently, a selective inhibitor of CBP/EP300 catalytic activity called A-485 was shown to selectively inhibit cell proliferation, confirming the relevance of histone acetyltransferase (HAT) inhibition in MM [75].

The importance of histone methyltransferase activity in MM has been highlighted by studies showing that the inhibition of the catalytic subunit EZH2 of the polycomb repressive complex 2 (PRC2) causes a reduction of MM cells viability and a down regulation of IRF4, MYC and BLIMP1 expression *via* up-regulation of potent tumour suppressor microRNAs miR-125a-3p and miR-320c [76].

Growing experimental and clinical evidence underscore the importance of natural killer (NK) cells in the immune response against MM. Combination therapies that also enhance the activity of NK cells against MM are showing promise in treating this hematologic cancer. For example, inhibition of BET through its bromodomain causes an increase of NK cell-activating MICA ligand in MM cells resulting in enhanced NK cell-mediated cytotoxicity and an up-regulation of the tumour suppressor microRNA-125b-5p (miR-125b), involved in the downregulation of IRF4 expression and of its downstream signalling [77,11]. Incidentally, IMiD drug lenalidomide has both tumouricidal and immunomodulatory activity in MM as the Cereblon-dependent degradation of IKZF1/3 proteins also causes an increase of NK cell-activating ligands MICA and PVR [78].

Given the importance of the MM- specific auto regulatory loop between IRF4 and MYC [10,57], IRF4 could be down regulated by direct targeting of MYC. Previous studies show that knockdown of MYC results in a decreased viability of MM cells [10], whilst inhibition of MYC-MAX heterodimerization, by the small-molecule compound 10058-F4, causes MM cell death [79]. Since BET protein BRD4 directly regulates *MYC* expression in MM cells [80,81], treatment of MM cells with BET inhibitor JQ1 causes release of BRD4 from the *MYC* promoter resulting in the reduction of proliferation and viability of MM cells [80,81]. JQ1 also suppresses the secretion of the key survival factor IL-6 in MM cells [82]. JQ1 is not being tested in clinical trials due to its short half life [83], but new BET inhibitors in development for MM therapeutic use include I-BET762 (GSK525762) (preclinical) [84], CPI203 for multiple myeloma resistant to bortezomib and melphalan (preclinical) [85] and CPI-0610 (clinical trials) (NCT02157636) [86].

Another strategy to target IRF4 could be to disrupt MM-specific IRF4 direct protein-protein interactions, although it is not clear if IRF4 directly interacts with other protein in the MM context. Co-occupancy of EICE composite motifs sites by PU.1 and IRF4 is important for gene regulation during B cell activation [15]. During PC differentiation however high concentrations of IRF4 promote binding to ISRE motifs, up-regulating PC specific genes, like *PDRM1*, and inhibiting PU.1 (*Spib1*) [15,23]. In the majority of MM cells studied, PU.1 has been shown to be down regulated [87], however induced overexpression of PU.1 in MM cells causes down regulation of IRF4 expression and cell death by activation of the IRF7-INFβ pathway [88]. Up regulation of PU.1 could therefore represent a promising therapeutic strategy for MM.

The MM specific IRF4 aberrant downstream transcriptional network could also represent a valid target for inhibition. Previous studies reported that knockdown of BLIMP1 by short hairpin RNA causes apoptosis in MM cells [89] and that the interaction between Aiolos and BLIMP1 plays an important role in MM cells survival, probably through the collaborative down-regulation of pro-apoptotic genes [60]. Moreover, treatment with IMiD lenalidomide caused proteasomal degradation of BLIMP1 and reduced Aiolos levels leading to apoptosis of MM cells [60].

MM cells secrete an excess of monoclonal proteins and a stringent endoplasmatic reticulum (ER) quality control is essential for these high levels of protein synthesis [90]. During ER stress, activated IRE1α protein mediates splicing of the XBP1 mRNA and fully initiates the UPR [91]. Because of its fundamental role in ER quality control, Xbp1 is highly expressed in MM cells and is required for their growth and survival [90]. Mimura et al. showed that targeting the IRE1α-Xbp1 pathway by small-molecule inhibition results in a decrease of MM cell viability [90]. Moreover, IRE1α-Xbp1 pathway inhibition in MM cells causes an enhanced activity of the proteasome inhibitor (PI) drug bortezomib (BTZ) [90]. PIs, including BTZ, are the backbone of MM therapies but they fail to cure because of resistance whose mechanisms remain controversial. Leung-Hagesteijn et al. showed that primary MM tumours are characterized by the presence of $Xbp1^+$ plasma cells but also by subpopulations of $Xbp1^-$ pre-plasmablasts and earlier $CD20^+$ B cell progenitors that are intrinsically PI-insensitive [92]. As $Xbp1^-$ pre-plasmablasts are deficient in full secretory status and produce less Ig [93,94], they are less vulnerable to lethal ER stress when Endoplasmic-reticulum-associated protein degradation (ERAD) is inhibited by PIs treatment [92]. Moreover, *XBP1* is characterized by two inactivating mutations in MM [43,92] both promoting BTZ resistance in MM cell lines [92]. Conspicuously, *PRDM1*, which like *XBP1* is essential for

plasma cell maturation and for Ig production [38], is also mutated in MM tumors [43] suggesting that mutation of genes different from *XBP1* and *IRE1* can drive maturation arrest in MM causing PI-resistance [92].

Finally, a therapeutic option for MM could be direct targeting of IRF4. Transcription factors, like IRF4, have been traditionally considered a very attractive class of biologically potent but un-druggable targets. Transcription factors usually lack pockets amenable to small molecule inhibition and instead operate largely through protein-DNA and protein-protein interactions, mediated by relatively featureless surfaces or helices [95]. However, few transcription factors have been successfully targeted and their inhibitors are in development as drugs for various cancers (p53 agonists (Nutlin-3), SREBP inhibitors (Fatostatin), LXR agonists (T0901317), MYC inhibitors (10058-F4) and NFkB inhibitors [96] suggesting that IRF4 might also be a druggable target.

5. Summary and future directions

Despite the introduction of novel immunomodulatory drugs, such as thalidomide lenalidomide and pomalidomide, proteasome inhibitors such as bortezomib, carfilzomib and ixazomib, monoclonal antibodies such as daratumumab and elotuzumab and histone deacetylase inhibitors such as panobinostat, MM remains an incurable cancer. In this review, we firstly examined the biology of IRF4 in healthy B cells and PCs and we then focused on the aberrant IRF4-transcriptional network, characteristic of MM cells. Knockdown experiments and pharmacological suppression have both shown significant decrease in the viability of MM cells [10,74] confirming IRF4 as an attractive target for novel therapeutic strategies. Various pharmacological approaches have been discussed based on the inhibition of IRF4 MM-specific upstream and downstream pathways.

Future work might focus on the direct inhibition of IRF4, as currently no IRF4-specific small molecule inhibitor is available. Concentration induced homodimerization and consequent binding to ISRE composite motifs has been shown to be a requirement for IRF4 to mediate its transcriptional activity in the context of PCs [15]. Overexpression in MM suggests that homodimerization is the prevalent IRF4 mode of action in MM cells and that targeting such dimerization could constitute a valid approach to MM subversion.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] S.V. Rajkumar, et al., International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma, Lancet Oncol. 15 (12) (2014) e538–e548.
[2] S.V. Rajkumar, Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management, Am. J. Hematol. 91 (7) (2016) 719–734.
[3] C. Botta, et al., A gene expression inflammatory signature specifically predicts multiple myeloma evolution and patients survival, Blood Cancer J. 6 (12) (2016) e511.
[4] S.V. Rajkumar, S. Kumar, Multiple myeloma: diagnosis and treatment, Mayo Clin. Proc. 91 (1) (2016) 101–119.
[5] S.K. Kumar, et al., Natural history of relapsed myeloma, refractory to immunomodulatory drugs and proteasome inhibitors: a multicenter IMWG study, Leukemia 31 (11) (2017) 2443–2448.
[6] S.G. Remesh, V. Santosh, C.R. Escalante, Structural studies of IRF4 reveal a flexible autoinhibitory region and a compact linker domain, J. Biol. Chem. 290 (46) (2015) 27779–27790.
[7] U. Klein, et al., Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination, Nat. Immunol. 7 (7) (2006) 773–782.
[8] R. Sciammas, et al., Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation, Immunity 25 (2) (2006) 225–236.
[9] L. Wang, et al., Gene expression profiling identifies IRF4-associated molecular signatures in hematological malignancies, PLoS One 9 (9) (2014) e106788.
[10] A.L. Shaffer, et al., IRF4 addiction in multiple myeloma, Nature 454 (7201) (2008) 226–231.
[11] E. Morelli, et al., Selective targeting of IRF4 by synthetic microRNA-125b-5p mimics induces anti-multiple myeloma activity in vitro and in vivo, Leukemia 29 (11) (2015) 2173–2183.
[12] C.R. Escalante, et al., Crystal structure of PU.1/IRF-4/DNA ternary complex, Mol. Cell 10 (5) (2002) 1097–1105.
[13] A.L. Brass, A.Q. Zhu, H. Singh, Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers, EMBO J. 18 (4) (1999) 977–991.
[14] R. Tussiwand, et al., Compensatory dendritic cell development mediated by BATF-IRF interactions, Nature 490 (7421) (2012) 502–507.
[15] K. Ochiai, et al., Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4, Immunity 38 (5) (2013) 918–929.
[16] A.L. Brass, et al., Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1, Genes Dev. 10 (18) (1996) 2335–2347.
[17] B.Y. Qin, et al., Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation, Nat. Struct. Biol. 10 (11) (2003) 913–921.
[18] K. Takahasi, et al., X-ray crystal structure of IRF-3 and its functional implications, Nat. Struct. Biol. 10 (11) (2003) 922–927.
[19] S.L. Nutt, et al., The generation of antibody-secreting plasma cells, Nat. Rev. Immunol. 15 (3) (2015) 160–171.
[20] M.J. Shlomchik, F. Weisel, Germinal center selection and the development of memory B and plasma cells, Immunol. Rev. 247 (1) (2012) 52–63.
[21] M. Muramatsu, et al., Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme, Cell 102 (5) (2000) 553–563.
[22] S.N. Willis, et al., Transcription factor IRF4 regulates germinal center cell formation through a B cell-intrinsic mechanism, J. Immunol. 192 (7) (2014) 3200–3206.
[23] A.L. Shaffer, et al., Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program, Immunity 17 (1) (2002) 51–62.
[24] M. Saito, et al., A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma, Cancer Cell 12 (3) (2007) 280–292.
[25] K. Basso, R. Dalla-Favera, Roles of BCL6 in normal and transformed germinal center B cells, Immunol. Rev. 247 (1) (2012) 172–183.
[26] K. Basso, et al., Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells, Blood 115 (5) (2010) 975–984.
[27] J. Alinikula, et al., Alternate pathways for Bcl6-mediated regulation of B cell to plasma cell differentiation, Eur. J. Immunol. 41 (8) (2011) 2404–2413.
[28] C. Huang, et al., Cooperative transcriptional repression by BCL6 and BACH2 in germinal center B-cell differentiation, Blood 123 (7) (2014) 1012–1020.
[29] D. Dominguez-Sola, et al., c-MYC is required for germinal center selection and cyclic re-entry, Nat. Immunol. 13 (11) (2012) 1083–1091.
[30] D.P. Calado, et al., MYC is essential for the formation and maintenance of germinal centers, Nat. Immunol. 13 (11) (2012) 1092–1100.
[31] R. Sciammas, M.M. Davis, Modular nature of Blimp-1 in the regulation of gene expression during B cell maturation, J. Immunol. 172 (9) (2004) 5427–5440.
[32] Y. Lin, K. Wong, K. Calame, Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation, Science 276 (5312) (1997) 596–599.
[33] T. Decker, et al., Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis, Immunity 30 (4) (2009) 508–520.
[34] A. Schebesta, et al., Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function, Immunity 27 (1) (2007) 49–63.
[35] H. Gonda, et al., The balance between Pax5 and Id2 activities is the key to AID gene expression, J. Exp. Med. 198 (9) (2003) 1427–1437.
[36] K.I. Lin, et al., Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells, Mol. Cell. Biol. 22 (13) (2002) 4771–4780.
[37] J.N. Gass, N.M. Gifford, J.W. Brewer, Activation of an unfolded protein response during differentiation of antibody-secreting B cells, J. Biol. Chem. 277 (50) (2002) 49047–49054.
[38] J. Tellier, et al., Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response, Nat. Immunol. 17 (3) (2016) 323–330.
[39] K. Ochiai, et al., Zinc finger-IRF composite elements bound by Ikaros/IRF4 complexes function as gene repression in plasma cell, Blood Adv. 2 (8) (2018) 883–894.
[40] H. Bai, et al., Bone marrow IRF4 level in multiple myeloma: an indicator of peripheral blood Th17 and disease, Oncotarget 8 (49) (2017) 85392–85400.
[41] S. Iida, et al., Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma, Nat. Genet. 17 (2) (1997) 226–230.
[42] S. Yoshida, et al., Detection of MUM1/IRF4-IgH fusion in multiple myeloma, Leukemia 13 (11) (1999) 1812–1816.
[43] M.A. Chapman, et al., Initial genome sequencing and analysis of multiple myeloma, Nature 471 (7339) (2011) 467–472.
[44] J.G. Lohr, et al., Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy, Cancer Cell 25 (1) (2014) 91–101.
[45] M. Alzrigat, A.A. Parraga, H. Jernberg-Wiklund, Epigenetics in multiple myeloma: from mechanisms to therapy, Semin. Cancer Biol. (2017), https://doi.org/10.1016/j.semcancer.2017.09.007 pii: S1044-579X(17)30104-9. PMID: 28962927.
[46] B. Salhia, et al., DNA methylation analysis determines the high frequency of genic

hypomethylation and low frequency of hypermethylation events in plasma cell tumors, Cancer Res. 70 (17) (2010) 6934–6944.
[47] B.A. Walker, et al., Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma, Blood 117 (2) (2011) 553–562.
[48] C.J. Heuck, et al., Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis, J. Immunol. 190 (6) (2013) 2966–2975.
[49] V. Bollati, et al., Differential repetitive DNA methylation in multiple myeloma molecular subgroups, Carcinogenesis 30 (8) (2009) 1330–1335.
[50] J.I. Sive, et al., Global hypomethylation in myeloma is associated with poor prognosis, Br. J. Haematol. 172 (3) (2016) 473–475.
[51] C. Houde, et al., Overexpression of the NOTCH ligand JAG2 in malignant plasma cells from multiple myeloma patients and cell lines, Blood 104 (12) (2004) 3697–3704.
[52] P. Martin, et al., Aberrant gene promoter methylation in plasma cell dyscrasias, Exp. Mol. Pathol. 84 (3) (2008) 256–261.
[53] X. Agirre, et al., Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers, Genome Res. 25 (4) (2015) 478–487.
[54] R. Winkelmann, et al., KLF2–a negative regulator of pre-B cell clonal expansion and B cell activation, PLoS One 9 (5) (2014) e97953.
[55] H. Ohguchi, et al., The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival, Nat. Commun. 7 (2016) 10258.
[56] K. Yamane, et al., JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor, Cell 125 (3) (2006) 483–495.
[57] N. Li, et al., Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression, Nat. Commun. 7 (2016) 13656.
[58] A. Lopez-Girona, et al., Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response, Br. J. Haematol. 154 (3) (2011) 325–336.
[59] K.I. Lin, Y. Lin, K. Calame, Repression of c-myc is necessary but not sufficient for terminal differentiation of B lymphocytes in vitro, Mol. Cell. Biol. 20 (23) (2000) 8684–8695.
[60] K.H. Hung, et al., Aiolos collaborates with Blimp-1 to regulate the survival of multiple myeloma cells, Cell Death Differ. 23 (7) (2016) 1175–1184.
[61] I. Gyory, et al., Identification of a functionally impaired positive regulatory domain I binding factor 1 transcription repressor in myeloma cell lines, J. Immunol. 170 (6) (2003) 3125–3133.
[62] E. Ocana, et al., The expression of PRDI-BF1 beta isoform in multiple myeloma plasma cells, Haematologica 91 (11) (2006) 1579–1580.
[63] S. Huang, G. Shao, L. Liu, The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression, J. Biol. Chem. 273 (26) (1998) 15933–15939.
[64] T. Hideshima, et al., A proto-oncogene BCL6 is up-regulated in the bone marrow microenvironment in multiple myeloma cells, Blood 115 (18) (2010) 3772–3775.
[65] K. Tahara, et al., Overexpression of B-cell lymphoma 6 alters gene expression profile in a myeloma cell line and is associated with decreased DNA damage response, Cancer Sci. 108 (8) (2017) 1556–1564.
[66] A. Hanbali, et al., The evolution of prognostic factors in multiple myeloma, Adv. Hematol. 2017 (2017) 4812637.
[67] L. Naymagon, M. Abdul-Hay, Novel agents in the treatment of multiple myeloma: a review about the future, J. Hematol. Oncol. 9 (1) (2016) 52.
[68] H.W. Mittrucker, et al., Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function, Science 275 (5299) (1997) 540–543.
[69] G. Lu, et al., The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins, Science 343 (6168) (2014) 305–309.
[70] J. Kronke, et al., Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells, Science 343 (6168) (2014) 301–305.
[71] Y.X. Zhu, et al., Identification of cereblon-binding proteins and relationship with response and survival after IMiDs in multiple myeloma, Blood 124 (4) (2014) 536–545.
[72] S. Tang, et al., Crucial role of HO-1/IRF4-dependent apoptosis induced by panobinostat and lenalidomide in multiple myeloma, Exp. Cell Res. 363 (2) (2018) 196–207.
[73] A. Bat-Erdene, et al., Synergistic targeting of Sp1, a critical transcription factor for myeloma cell growth and survival, by panobinostat and proteasome inhibitors, Oncotarget 7 (48) (2016) 79064–79075.
[74] A.R. Conery, et al., Bromodomain inhibition of the transcriptional coactivators CBP/EP300 as a therapeutic strategy to target the IRF4 network in multiple myeloma, Elife 5 (2016).
[75] L.M. Lasko, et al., Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours, Nature 550 (7674) (2017) 128–132.
[76] M. Alzrigat, et al., EZH2 inhibition in multiple myeloma downregulates myeloma associated oncogenes and upregulates microRNAs with potential tumor suppressor functions, Oncotarget 8 (6) (2017) 10213–10224.
[77] M.P. Abruzzese, et al., Inhibition of bromodomain and extra-terminal (BET) proteins increases NKG2D ligand MICA expression and sensitivity to NK cell-mediated cytotoxicity in multiple myeloma cells: role of cMYC-IRF4-miR-125b interplay, J. Hematol. Oncol. 9 (1) (2016) 134.
[78] C. Fionda, et al., The IMiDs targets IKZF-1/3 and IRF4 as novel negative regulators of NK cell-activating ligands expression in multiple myeloma, Oncotarget 6 (27) (2015) 23609–23630.
[79] T. Holien, et al., Addiction to c-MYC in multiple myeloma, Blood 120 (12) (2012) 2450–2453.
[80] J.A. Mertz, et al., Targeting MYC dependence in cancer by inhibiting BET bromodomains, Proc. Natl. Acad. Sci. U. S. A. 108 (40) (2011) 16669–16674.
[81] J.E. Delmore, et al., BET bromodomain inhibition as a therapeutic strategy to target c-Myc, Cell 146 (6) (2011) 904–917.
[82] R.R. Ghurye, H.J. Stewart, T.J. Chevassut, Bromodomain inhibition by JQ1 suppresses lipopolysaccharide-stimulated interleukin-6 secretion in multiple myeloma cells, Cytokine 71 (2) (2015) 415–417.
[83] E. Wadhwa, T. Nicolaides, Bromodomain inhibitor review: bromodomain and extra-terminal family protein inhibitors as a potential new therapy in central nervous system tumors, Cureus 8 (5) (2016) e620.
[84] A. Chaidos, et al., Potent antimyeloma activity of the novel bromodomain inhibitors I-BET151 and I-BET762, Blood 123 (5) (2014) 697–705.
[85] M.B. Siegel, et al., Small molecule inhibitor screen identifies synergistic activity of the bromodomain inhibitor CPI203 and bortezomib in drug resistant myeloma, Oncotarget 6 (22) (2015) 18921–18932.
[86] M. Perez-Salvia, M. Esteller, Bromodomain inhibitors and cancer therapy: From structures to applications, Epigenetics 12 (5) (2017) 323–339.
[87] H. Tatetsu, et al., Down-regulation of PU.1 by methylation of distal regulatory elements and the promoter is required for myeloma cell growth, Cancer Res. 67 (11) (2007) 5328–5336.
[88] N. Ueno, et al., PU.1 acts as tumor suppressor for myeloma cells through direct transcriptional repression of IRF4, Oncogene 36 (31) (2017) 4481–4497.
[89] F.R. Lin, et al., Induction of apoptosis in plasma cells by B lymphocyte-induced maturation protein-1 knockdown, Cancer Res. 67 (24) (2007) 11914–11923.
[90] N. Mimura, et al., Blockade of XBP1 splicing by inhibition of IRE1alpha is a promising therapeutic option in multiple myeloma, Blood 119 (24) (2012) 5772–5781.
[91] R. Bravo, et al., Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration, Int. Rev. Cell Mol. Biol. 301 (2013) 215–290.
[92] C. Leung-Hagesteijn, et al., Xbp1s-negative tumor B cells and pre-plasmablasts mediate therapeutic proteasome inhibitor resistance in multiple myeloma, Cancer Cell 24 (3) (2013) 289–304.
[93] M. Jourdan, et al., Characterization of a transitional preplasmablast population in the process of human B cell to plasma cell differentiation, J. Immunol. 187 (8) (2011) 3931–3941.
[94] A.M. Reimold, et al., Plasma cell differentiation requires the transcription factor XBP-1, Nature 412 (6844) (2001) 300–307.
[95] K.A. Papavassiliou, A.G. Papavassiliou, Transcription factor drug targets, J. Cell. Biochem. 117 (12) (2016) 2693–2696.
[96] A.S. Bhagwat, C.R. Vakoc, Targeting transcription factors in cancer, Trends Cancer 1 (1) (2015) 53–65.

3. Chapter 3. Disrupting the IRF4-MYC oncogenic loop in Multiple Myeloma

Disrupting the IRF4-MYC oncogenic loop in Multiple Myeloma

Alessandro Agnarelli[1], Simon Mitchell[2], David Wood[1], Leanne Milton-Harris[1], Timothy Chevassut[2], Michelle J West[1] and Erika J Mancini[1,3]

[1]School of Life Sciences, Biochemistry and Biomedicine Department, University of Sussex, Falmer, Brighton, BN1 9QG, United Kingdom

[2]Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom.

[3]Erika J. Mancini, JMS Building, University of Sussex, +44 1273 678613 erika.mancini@sussex.ac.uk

Abstract

B cell progenitor fate determinant interferon regulatory factor 4 (IRF4) exerts key roles in the pathogenesis and progression of multiple myeloma (MM). Aberrant expression of IRF4 and the establishment of a positive auto-regulatory loop with oncogene MYC, drives an MM specific gene expression programme leading to the abnormal expansion of malignant immature plasma cells. Targeting the IRF4-MYC oncogenic loop might provide a more selective and effective therapy for MM. Here we evaluate the use of bromodomain inhibitors to target the IRF-MYC axis through combined disruption of their known epigenetic regulators BRD4 and CBP/EP300. We found that within the time frames chosen for the study there is no synergistic effect on the viability of MM cell lines. Within that time frame and for all inhibitors tested we experimentally measure largely unaffected levels of IRF4 protein and downstream target proteins mRNA levels. We find that these results can be partially explained by the high stability of the IRF4 protein over a prolonged time frame. Gene network modelling of MM suggests that the cell viability effects of bromodomain inhibition are not exerted through IRF4 but indirectly through MYC.

Keywords: Multiple Myeloma, IRF4, MYC, BET/BRD4, CBP/EP300, Dual inhibition

Introduction

Transcription factor IRF4 (interferon regulatory factor 4) is a key activator of lymphocyte development, affinity maturation and terminal differentiation into immunoglobulin-secreting plasma cells [1] [2]. Faulty regulation of IRF4 expression is associated with numerous lymphoid malignancies, including notably multiple myeloma (MM), an aggressive and incurable hematologic cancer characterized by the abnormal proliferation of bone marrow plasma cells [2] [3]. At the molecular level MM is an heterogenous disease with several subgroups defined by specific gene expression profiles and recurrent chromosomal rearrangements. Moreover, MM develops via a multistep process where the malignant clone accumulates further genetic damage as the disease progresses. In a minority of MM cases, chromosomal translocation t(6;14)(p25;q32) brings the *IRF4* gene under the control of immunoglobulin heavy-chain regulatory regions [4] [5]. Interestingly while IRF4 is not always genetically altered in MM [6], its expression levels are higher than in plasma cells [7]. Over-expression of IRF4 leads to an aberrant gene expression programme and to the mis-regulated transcription of a wide network of target genes. Remarkably, gene expression profiling and genome-wide chromatin immunoprecipitation analysis in MM cells identified MYC as a direct target of IRF4 and IRF4 itself as a direct target of MYC transactivation, giving rise to an autoregulatory circuit [6]. IRF4 loss-of-function, RNA-interference-based experiments have shown that MM cells are “addicted” to this abnormal gene expression programme since IRF4 inhibition causes rapid and extended non-apoptotic cell death, irrespective of genetic etiology [7]. Similarly, targeting the 3’ UTR of IRF4 for degradation by overexpression of miR-125-b, also leads to MM cell death [8].

MM accounts for 2% of all cancers and about 10% of all hematologic malignancies [9]. In the UK around 5800 new MM cases are diagnosed every year (2015-2017) and the incidence rates are projected to rise by 11% by 2035. The past decade has seen a revolution in in the management of MM with the availability of novel therapies which are both more effective and less toxic, including proteasome inhibitors, immunomodulatory agents, and monoclonal antibodies. Despite the ensuing improvement of clinical outcomes, nearly every patient

becomes refractory to current therapies and the overall 5-year survival rate is 52% [10]. Considering that existing treatments are not curative, new therapeutic approaches are needed, ideally targeting a molecular pathway shared by multiple subtypes of MM. Targeting IRF4 might provide a powerful therapeutic strategy aimed at MM. Firstly, IRF4 inhibition likely presents manageable side effects as phenotypes in IRF4-deficient mice are restricted to lymphoid and myeloid lineages and mice-lacking one allele of IRF4 are phenotypically normal [6]. Additionally, because of the MM cells' "addiction" to IRF4 even fairly a small decrease in IRF4 levels could lead to cell death. The aberrant IRF4-driven gene-expression program in MM cells directly controls genes critical for cell-cycle control, transcriptional regulation, plasma-cell differentiation, and membrane biogenesis [6]. Finally, IRF4 inhibition is lethal to all MM cells regardless of their underlying transforming oncogenic mechanism [6] .

A particularly attractive approach to inhibit IRF4 might be to target MYC, a known regulator of IRF4 expression in MM. Constitutive activation of MYC signalling is detected in more than 60% of patient-derived myeloma cells and one of the most common somatic genomic aberrations in early- and late-stage MM is rearrangement or translocation of MYC [11]. MYC transactivates *IRF4* by binding to an evolutionarily conserved intronic region whilst IRF4 binds directly to the *MYC* promoter region in MM cells and transactivates its expression, creating a positive autoregulatory feedback loop [6]. The expression of MYC in MM cells is unquestionably abnormal since normal plasma cells do not express MYC as a result of repression by PR domain zinc finger protein 1 (*PRDM1* gene that encodes for BLIMP1 protein) [12]. Moreover, IRF4 binds to its own promoter region, creating a second positive autoregulatory loop which would potentiate any therapeutic effect of targeting the MYC-IRF4 loop [6]. The IRF4-MYC axis is thus considered to be a very promising therapeutic target in MM.

One possible way to target the IRF4-MYC axis is through upstream epigenetic regulators. For this purpose, bromodomain and extra-terminal (BET) proteins inhibitors have emerged as promising molecules for the treatment of haematologic malignancies [13]. BET proteins are chromatin adaptors that bind acetylated lysine residues on histone tails [14] and play

prominent roles in cellular proliferation and cancer signalling. BET protein BRD4 is specifically enriched at immunoglobulin heavy chain (IgH) enhancers in MM cells bearing IgH rearrangement at the *MYC* locus, causing aberrant proliferation of MM cells [15]. BET inhibitors such as JQ1 or OTX015, which inhibit BRD4 by competitively binding to its bromodomain acetyl-lysine recognition pocket, have been described as MM therapeutics. By selectively displacing BRD4 from chromatin they trigger inhibition of *MYC* transcription in a dose and time-dependent manner.

CREB (cyclic-AMP response element binding protein) binding protein (CBP) and its highly homologous E1A binding protein of 300 kDa (EP300) are bromodomain-containing histone acetyltransferases [16]. Previous data suggest that CBP/EP300 are transcriptional coactivators with important role in the preservation of super-enhancers [17] [18]. Inhibitors of the bromodomains of CBP/EP300, such as SGC-CBP30, induce cell cycle arrest and apoptosis in MM cell lines [19] . Whilst however the effects of BET bromodomain inhibition are most likely due to direct suppression of MYC, inhibition of CBP/EP300 bromodomain has been proposed to affect MM cells through suppression of IRF4 [19].

Given the positive auto regulation loop between MYC and IRF4 in MM, we hypothesised that the combination of the two classes of inhibitors with distinct transcriptional effects would have a synergistic impact on MM cells. To confirm this hypothesis, we explored the effect of various combination of BET and CBP/Ep300 inhibitors on the cell viability of a panel of MM cell lines. To assess whether the protein and mRNA levels for MYC, IRF4 and their downstream targets following drug exposure were consistent with those expected from the IRF4-MYC auto-regulatory loop model, we compared their experimentally measured expression with their simulated expression in a network model of MM molecular interactions. We found that within the time frames chosen for the study there is no synergistic effect on the viability of MM cell lines. Within that time frame and for all inhibitors tested we experimentally measure largely unaffected levels of IRF4 protein and downstream target proteins mRNA levels. We find that these results can be partially explained by the high stability of the IRF4 protein over a prolonged time frame. However, discrepancies between the measured and modelled levels of

molecular species point at additional and yet uncovered regulatory interactions within the IRF4 network in MM cells. Finally, our network modelling of MM suggests that cell viability effects of CBP/EP300 bromodomain inhibition are not exerted directly through IRF4 but indirectly through MYC.

Results

Combination of bromodomain inhibitors does not have a synergistic effect on multiple myeloma cells viability

To explore the effect of the combination of BET and CBP/EP300 inhibitors on MM cell viability, we employed BET inhibitors JQ1 and OTX015, CPB/EP300 inhibitor SGC-CBP30 and ISOX-DUAL, a dual inhibitor of BET and CPB/EP300. Three MM (KMS-12-BM, NCI-H929, SKMM-1) and one acute leukaemia (OCI-AML3) cells lines were treated for 48hrs with different concentrations of these compounds. As shown in Fig 1a-e, JQ1 was the most effective inhibitor of cell viability with an IC_{50} between 0.27 and 0.32μM. Very similar values were obtained by treatment with OTX015 and a combination of JQ1 and SGC-CBP30. However, treatment with SGC-CBP30 alone and ISOX-DUAL showed reduced efficacy with ISOX-DUAL showing the least effect on cell viability. Following treatment with SGC-CBP30 or ISOX-DUAL, we could observe a cell-line specific response, with the SKMM-1 cells being the least and KMS-12-BM the most affected respectively. In contrast, treatment with JQ1, OTX015 or combination of JQ1 and SGC-CBP30 had similar effect on cell viability, with the exception of SKMM-1 with consistently higher IC_{50}s. The dual inhibitor ISOX-DUAL was found to be less effective than either JQ1 or SGC-CBP30, for which we measured IC_{50}s in line with those reported in other studies. This result could be explained by the reduced affinity of ISOX-DUAL for BRD4 (IC_{50} 1.5μM) and CPB/EP300 (IC_{50} 0.65μM), when compared to JQ1 and SGC-CBP30 respectively [20]. To test this hypothesis, we compared the effect on cell viability of ISOX-DUAL treatment with that of a combination of JQ1 and SGC-CBP30. Fig 1e suggests this hypothesis is correct and that the treatment with a combination of JQ1 and SGC-CBP30 has a stronger inhibitory effect on cell viability compared to the dual inhibitor alone, with an

IC_{50} that is comparable with that of JQ1 and not of SGC-CBP30. Taken together, these results suggest that ISOX-DUAL has the least effect on cell viability and that combining JQ1 and SGC-CBP30 does not lead to synergistic or antagonistic cytotoxic effects.

Bromodomain inhibitors impact mRNA but not protein levels of IRF4 in MM cell lines

We next investigated the effects of different bromodomain inhibitors on the protein and gene expression levels of IRF4 and MYC in MM cell lines. We treated KMS-12-BM, NCI-H929, SKMM-1 and OCI-AML3 cells with a concentration of drugs corresponding to their IC_{50} value as determined in Fig.1. As shown by western blotting analysis, we observed a dramatic decrease in the level of MYC protein, following treatment for 4hrs, 8hrs, 24hrs (supplementary Fig.S1) and 48hrs (Fig. 2a) with JQ1, OTX015, SGC-CBP30 and ISOX-DUAL or a combination of JQ1 and SGC-CBP30. OCI-AML3 cell lines do not express IRF4 (Fig. 2a). In the control experiment, the different cell lines were exposed to DMSO (supplementary Fig.S1, Fig. 2a). The observed reduction in MYC levels in the control experiments over time is due to the effect of exposure of cells to DMSO. However, drug treatments did not strongly affect IRF4 protein levels at the tested time points for the chosen MM cell lines (Supplementary Fig.S1 and Fig2a). Interestingly there was a slight reduction (up to 30%) in IRF4 protein levels consistently across all MM cell lines when a combination of JQ1 and SGC-CBP30 was used. In this experiment the different cell lines were treated with the different drugs up to 48hrs. The fully drug inhibitory effect is probably not reached after 48hrs. A longer exposure of cells to the drugs (i.e. 72hrs) could have started to cause an IRF4 decrease. We then examined the effect of drug treatment on the mRNA levels of *IRF4* and *MYC*. Treatment with all drugs significantly decreased both *IRF4* and *MYC* mRNA expression in all cell lines after 4hrs, 8hrs, 24hrs (supplementary Fig.S2) and 48hrs (Fig. 3). In summary, our data finds that bromodomains inhibitors strongly impact *MYC* and *IRF4* mRNA levels and MYC protein levels but do not significantly reduce IRF4 protein levels in MM cell lines over the tested time points. Interestingly, whilst no significant reduction could be seen with either JQ1 or SGC-CBP30 alone, the combination of the two bromodomain inhibitors has a small but consistent effect

across all MM cell lines studied on IRF4 protein levels. This suggests a small but consistent additive effect of JQ1 and SGC-CBP30 treatment on IRF4 protein levels but not mRNA levels in MM cell lines.

Bromodomain inhibitors affect the gene expression levels of MYC but not IRF4 downstream targets

As protein levels of MYC and IRF4 were unequally affected by drug treatment, we hypothesize that their downstream gene target would be differentially disturbed. To test this hypothesis, we evaluated the impact of drug treatment on IRF4 downstream targets, *KLF2* and *PRDM1* and MYC downstream targets, *CDK4* and *hTERT*. KMS-12-BM, NCI-H929, SKMM-1 and OCI-AML3 cells were treated with a concentration of drugs corresponding to their IC_{50} value for 4hrs, 8hrs, 24hrs and 48hrs. At the early time points of 4hrs, 8hrs and 24hrs, no significant reduction of mRNA levels could be seen in the MM cell lines tested for IRF4 downstream target *KLF2* (supplementary Fig.S3), whilst minor effects could be seen after 48hrs (Fig.4). A similar trend was observed for *PRDM1* mRNA levels, with small decreases after different drug treatments for 4hrs, 8hrs, 24hrs (Fig. S4) and more significant decreases only after 48hrs (Fig.4).

In contrast to IRF4 downstream targets *PRDM1* and *KLF2*, the mRNA levels of MYC downstream targets *hTERT* and *CDK4* were rapidly and strongly affected by drug treatment in all cell lines tested (Supplementary Fig. S4 and Fig.5).

In summary, these results confirmed our hypothesis that mRNA levels of MYC and IRF4 downstream targets are not equally affected by bromodomain inhibition, with MYC-dependent genes *hTERT* and *CDK4* mRNA levels being affected more strongly than those of IRF4 downstream targets *KLF2* and *PRDM1.* This behaviour is largely independent of the drug and cell line combination.

Gene and protein network modelling suggest a long half-life for IRF4 protein

The data on protein and mRNA levels following treatment with bromodomain inhibitors suggests that, independently from the MM cell line and drug tested, MYC protein and mRNA

levels as well as the mRNA levels of MYC downstream proteins are strongly and rapidly affected. In contrast, IRF4 protein levels and IRF4 downstream target genes mRNA levels remain largely unperturbed for at least 48 hours. Given the known feedback loop between MYC and IRF4 in MM cells we asked whether the unexpected results can be explained by a long half-life of IRF4 protein.

To test this hypothesis and to assess whether the protein and mRNA levels for MYC, IRF4 and their downstream targets following drug exposure were consistent with those expected from the IRF4-MYC auto-regulatory loop model we used computational techniques to model the MYC and IRF4 gene and protein network in MM cells. We then used this model to compare the experimentally measured gene expression and protein levels for MYC, IRF and BLIMP1 following drug treatment with their simulated levels. As the results are independent from the drug and cell line used, we initially modelled our response based on a drug inhibiting MYC expression (Fig. 6a) and assuming a half-life for MYC of 30min [20] and of 7 hours for IRF4 (no data on IRF4 half-life could be found in literature). The squared distance between the mean experimental result and modelled response for each timepoint shows a discrepancy between the model and the data (Fig. 6b), specifically for IRF4 protein and BLIMP1 mRNA levels, suggesting that IRF4 has a half-life significantly longer than 7 hours. To measure IRF4 protein half-life, we treated KMS-12-BM, NCI-H929 and SKMM-1 MM cell lines with 10μg/mL cycloheximide for up to 72 hours and measured the corresponding protein levels by western blotting analysis. As shown in Fig.7a, IRF4 is characterized by a long half-life in all three MM cell lines tested, specifically more than 48hrs in KMS-12-BM and NCI-H929 and more than 24hrs in SKMM-1 cell lines. In contrast to the stability of IRF4, levels of MYC decreased after cycloheximide exposure within 30min in all cell MM cell lines tested, with half-lives of 1hr, 22min and 30min respectively for KMS-12-BM, SKMM-1 and NCI-H929, in line with published reports [21]. To test whether a half-life of 48 hours for IRF4 can explain the observed response to the drug we modelled again MYC and IRF4 gene and protein network using a longer half-life of 48 hours. The squared distance between the mean experimental result and modelled response for each timepoint now shows a good agreement between the model and the data

(Fig. 7b). Despite the overall improvement of the fit, a discrepancy persists for IRF4 protein levels between 24 and 36 hours suggesting that the model does not completely recapitulate the data, especially at the later time points.

Gene and protein network modelling suggest that bromodomain inhibitors effects on MM cell lines are mainly exerted through MYC transcription repression and not IRF4

The initial computational modelling of the predicted drug response on MM cell lines was formulated on the assumption of bromodomain inhibition affecting mainly MYC transcription. This was a reasonable assumption based on the observation that unperturbed IRF4 protein levels in MM cell lines could be measured following most drug treatment. However, because of a small (30%) but consistent reduction of IRF4 protein levels in response to treatment with the JQ1+ SGC-CBP30 combination we then asked the question as to whether bromodomain inhibitors work through repression of MYC, IRF4 or both. To answer this question, we used gene and protein network modelling to simulate the effect of a drug acting on the transcription of MYC, IRF4 or both (Fig.8a) using the measured half-lives of IRF4 and MYC. When comparing the predicted to the experimentally measured expression of MYC, IRF4 and BLIMP1 in KMS-12-BM, NCI-H929 and SKMM-1 cell lines H929, SKMM-1, KMS cell lines exposed to SGC-CBP30, JQ1, OTX015, ISOX-DUAL, and JQ1+SGC-CBP30 combination we could conclude that the main effect of the drugs is predicted to be through disruption of MYC transcription. The modelled response of the effects of a drug acting only on IRF4 transcription predicts poorly the observed protein and mRNA levels, especially those of MYC. On the other hand, simulating the effects of a drug treatment targeting both MYC and IRF4 transcription improves the match between the predicted and observed data, however not as well when using a single-hit to MYC model. In summary, our gene and protein network modelling suggest that bromodomain inhibitors effects on MM cell lines are mainly exerted through MYC transcription repression and not through IRF4 or a combination of both. However, for all models a discrepancy remains between the measured and modelled levels of IRF4 protein

after 24hrs, pointing at additional and yet uncovered regulatory interactions within the IRF4 network in MM cells.

Discussion

Despite the crucial role of IRF4 and of the IRF4-MYC oncogenic loop in the pathogenesis of MM [6], no drug that disrupts this target is currently available to patients. Given the known effects of bromodomain inhibitors on the regulation of transcription of both MYC and IRF4 [22], their use in the clinical treatment of MM is an attractive therapeutical option. In this paper we studied the effects on MM cell lines of two classes of bromodomain (BET and CBP/Ep300) inhibitors, with putatively distinct transcriptional effects, with the aim of disrupting the oncogenic feedback loop between MYC and IRF4. Specifically, we wanted to evaluate the possibility that the combination of these bromodomain inhibitors would have synergistic impact on the viability of MM cells and on the transcription and protein levels of IRF4 and MYC.

To test this hypothesis, we explored the effect of various combination of inhibitors on a panel of MM cell lines. Our data showed that while the two BET inhibitors JQ1 and OTX015 had the most effective inhibition on KMS-12-BM, SKMM-1 and NCI-H929 cell viability, the CBP/Ep300 inhibitor SGC-CBP30 and the dual BET-CBP/Ep300 inhibitor ISOX-DUAL caused the least effect. Since the combination of JQ1 and SGC-CBP30 has a stronger inhibitory effect on cell viability compared to the dual inhibitor alone this suggests that the limited effect of ISOX-DUAL is caused by its reduced affinity for BRD4 and CPB/EP300. Also, our data suggests that combining JQ1 and SGC-CBP30 does not lead to synergistic or antagonistic cytotoxic effects on MM cell lines. In line with previous studies looking at the effects of JQ1, OTX015 and SGC-CBP30 treatment on MM cells lines [15] [23] [24] [19], we found that these drugs cause MYC downregulation at protein and mRNA levels. Interestingly, within the time frame and for all inhibitors tested we have observed largely unaffected levels of IRF4 protein and downstream target proteins mRNA levels. Using computational modelling of a network of MM molecular interactions, we could show that these results can be partially explained by the high stability of the IRF4 protein over a prolonged time frame (>48hrs). Finally, the modelling data also

implies that any effect observed on MM cell lines for both BRD4 and CBP/Ep300 inhibitors is not exerted through IRF4 but mainly through MYC. These results are in contrast with previous data by Conery et al., [19] supporting the idea that SGC-CBP30 treatment on MM cell line causes cell cytotoxicity via targeting of IRF4. However, more recent data show that inhibition of CBP/EP300 bromodomains can interfere with GATA1 and MYC-driven transcription by displacing CBP/EP300 from GATA1 and MYC binding sites at enhancers leading to a decrease in the level of acetylation of these regulatory regions. This in turn reduces gene expression of both GATA1 and MYC [25].

Our data shows that IRF4 is characterized by a long half-life (>24hrs) in a panel of MM cell lines. Previous studies have shown a variability in th half-lives values for IRFs proteins (IRF1 ~ 30min, IRF2 ~ 8hrs, IRF7~ 5hrs, IRF3 ~ 60hrs) [26] [27]. Ubiquitination is a post-translational protein modification that causes protein degradation by proteosomes or lysosomes [28]. Ubiquitin-specific proteases (USPs) are deubiquitinating enzymes that interfere with ubiquitination by removing ubiquitin from the targeted proteins [29]. Alterations of USP enzymes are implicated in the pathogenesis of various cancer [30]. USP15 for example has been reported to be overexpressed in MM cells and to inhibit MM apoptosis by activating a feedback loop with the transcription factor NF-κBp65 [29]. Interestingly, USP4 interacts, stabilizes and deubiquitinates IRF4 facilitating IL-4 expression in Th2 cells [31], which could be provide an explanation for the long IRF4 half-life we have observed. Future works might focus on the role of USP4 and its interaction with IRF4 in MM.

In conclusion, our data suggests that despite the autofeedback positive regulatory loop between IRF4 and MYC, bromodomain inhibition via BRD4 and CBP-EP300 might not be the most effective way to interfere with IRF4 expression, which is lethal for MM cells irrespectively of their genetic makeup. In addition, we suggest that additional and yet uncovered regulatory interactions exist within the IRF4 network in MM cells.

Future work aimed at targeting the IRF4-addiction in MM should focus on the direct inhibition or degradation of IRF4, taking also into account its elevated stability.

Materials and Methods

Cell lines

Multiple Myeloma cell lines KMS-12-BM, NCI-H929, SKMM-1 and acute myeloid leukemia OCI-AML3 are grown at 37°C, 5% CO_2 as follows. KMS-12-BM, SKMM-1 and OCI-AML3: RPMI medium with 20% FBS, 1% Penicillin-Streptomycin-L-glutamine, 1% Sodium-Pyruvate and 0.05% Thioglycerol. NCI-H929 RPMI medium with 20% FBS, 1% Penicillin Streptomycin-L-glutamine, 1% Sodium Pyruvate, 0.05% β-mercaptoethanol.

Cell Titer-Blue viability assay

Cells were plated at 20000 cells per well in 96 well plates and treated with 2mM DMSO or indicated compounds for 48 hours at 37°C (full concentration range used 0.01-10μM). Cell viability was analysed by performing CellTiter-Blue® Cell Viability Assay following manufactures instructions (Promega). After 48 hours treatment, 20μL CellTiter-Blue® Reagent (Promega G8080) was added to each 96 wells plate. The plates were shaked for 10 seconds and then incubated for 2 hours at 37°C. After 2 hours the plates were shaked 10 seconds and fluorescence was recorded at 560/590nm. Cells were plate in triplicate/per condition and at least n=3 assays were performed. IC_{50} was determined using nonlinear regression (curve fit) with log(inhibitor) vs response—Variable slope (four parametres) by GraphPad Prism Version 6.01. Mean IC_{50} was analyzed and statistical significance was performed by one-way ANOVA followed by Tukey's test to compare statistical differences among IC_{50s} of different cell lines treated with a specific drug.

Western Blotting

Cell lysates samples were prepared adding 100μL 1x Gel Sample Buffer/ 10^6 cells. Gel sample buffer: 50mM Tris pH 6.8, 4% SDS, 5% β-Mercaptoethanol, 0.01% Bromophenol blue, 10% Glycerol, 1mM EDTA. The lysates were then sonicated at 25% 5x 10 seconds with 10 seconds gaps on ice. Samples were separated by SDS-PAGE with Tris-Glycine 1X and SDS 1%

running buffer for 90 minutes at 120 Volt. Proteins were transferred on to Protran nitrocellulose membranes (Schleicher and Schuell) for 90 minutes at 85 Volt. Membrane were blocked with 5% milk in PBS-Tween for 1 hour. Membranes were incubated with primary antibody overnight at 4 °C. Membranes were then washed 3x10 minutes in PBS-Tween. After that, membranes were incubated with HRP-conjugated secondary antibody for 1 hour at room temperature. Membranes were then washed 3x10 minutes in PBS-Tween. Pierce™ ECL western blotting substrate (Thermo Fisher Scientific 32209) was added to the membrane for visualization at LI-COR machine. Stripping was performed by adding Tween® 20 [BP337-100, Thermo Fisher] to the membranes for 10 minutes. Membranes were then washed 3x10 minutes in PBS-Tween and block with 5% milk in PBS-Tween for 1 hour. Primary and secondary antibodies were added as described above. Primary antibodies specific for IRF4 (1:10000, Anti-MUM1 antibody [EP5699] (ab133590), Abcam), c-Myc (1:300, Antibody (9E10): sc-40, Santa Cruz Biotechnology), β-actin (1:5000, Anti-Actin antibody A2066, Sigma-Aldrich) and HRP-conjugated secondary antibodies anti-rabbit (1:3000, abcam ab205718), anti-mouse (1:5000, Cell signalling 7076S) were used.

RNA extraction, cDNA synthesis, and qRT-PCR

Total RNA was extracted using Monarch total RNA miniprep kit (T2010S). RNA concentrations were determined using a NanoDrop 2000 instrument (Thermo Scientific). cDNA was synthesized by using ImProm-II™ Reverse Transcription System kit with random primers (Promega A3800). RNA/primers mixes were prepared on ice with 1µg RNA, 1µl random primers and Nuclease-Free Water up to 5µL. Reverse transcription reaction contained 4.0µL ImProm-II™ 5X Reaction Buffer, 4µL $MgCl_2$ (6.6mM), 1.0µL dNTP Mix, 0.5µL Recombinant RNasin® Ribonuclease Inhibitor, 1.0µL ImProm-II™ Reverse Transcriptase and Nuclease-Free Water up to 15µL. 15µl of Reverse transcription reaction were then mixed to 5µL RNA/primers mixes in PCR tubes (Axygen® PCR-02-C). cDNA was synthesized by placing the PCR tubes first in a controlled-temperature heat block at 25°C for 5 minutes for the annealing reaction, then in a controlled-temperature heat block at 42°C for 1 hour for the extension

reaction. The reverse transcriptase was then inactivated incubating the reaction tubes in a controlled-temperature heat block at 72°C for 15 minutes. cDNA was then used for PCR amplification.

Real time PCR was performed using an Applied Biosystems StepOnePlus PCR machine. In the Real time PCR reactions cDNAs represents 20% of the reaction volume. For each sample we used a reaction volume of 15µL that was composed by 1X of GoTaq® qPCR Master mix, 2X (Promega A6002), 0.15µM of each primer, Nuclease-Free Water and 3µL cDNA. cDNA was amplified by heating samples to 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute prior to dissociation curve analysis. Serial dilutions of cDNA were used to generate standard curves for each primer set.

Primer sequences used for Real time PCR: *IRF4* Fw 5'-AACAAACTGGAGAGAGACCAGACC-3' Rv 5'-CCTCTCCAAAGCATAGAGTCACC-3'; *MYC* Fw 5'-CCTGGTGCTCCATGAGGAGAC-3' Rv 5'-CAGACTCTGACCTTTTGCCAGG-3'; *PRDM1* Fw 5'-TACATACCAAAGGGCACACG-3' Rv 5'- TGAAGCTCCCCTCTGGAATA-3'; *KLF2* Fw 5'-AGACCTACACCAAGAGTTCGCATC-3' Rv 5'-CATGTGCCGTTTCATGTGCAGC-3'; *CDK4* Fw 5'-CTTCTGCAGTCCACATATGCAACA-3' Rv 5'- CAACTGGTCGGCTTCAGAGTTTC-3'; *hTERT* Fw 5'-GGAGCAAGTTGCAAAGCATTG-3' Rv 5'-TCCCACGACGTAGTCCATGTT-3'; ***β-actin*** Fw 5'-TTCTACAATGAGCTGCGTGTG-3' Rv 5'- GGGGTGTTGAAGGTCTCAAA-3'.

Protein half-life

To analyse protein stability KMS-12-BM, NCI-H929, SKMM-1 cells were treated with 10µg/mL cycloheximide in T25 flasks. Cells were incubated with cycloheximide for up to 72 hours. After 72 hours cells were pipetted into 50mL tubes and span at 1300rpm for 10 minutes at 4°C. The media was aspirated off and a same or half volume of PBS was added. An aliquot of cell was removed to count the total cell number. Cells were span again (1300rpm for 10 minutes at 4°C) and 1mL/5x10^6 cells of PBS was added to each tube, followed by a final spin at 13000rpm for 10-20 seconds (pulse). PBS was aspirated off and the pellet was frozen on dry ice. Western

blotting analysis was then performed. The half-life protein was quantified by using nonlinear regression (curve fit) with one phase decay GraphPad Prism Version 6.01.

Statistical analysis

Statistical analysis was carried out as described previously [32] by using Statgraphics (version XVI) and GraphPad Prism (version 6.01) software. Data tabulation and descriptive statistics were performed by using Excel program (Office 2016). Data are expressed as a mean of three independent experiments with three replicates. Error bars represent standard error of the mean (SEM). Normality was tested by Shapiro-Wilk and Kolmogorov-Smirnov tests. Homoscedasticity was tested by Levene's test. For multiple comparisons of normally distributed data, one-way ANOVA analysis of variance with the Tukey's HSD post-hoc test was performed. P-values < 0.05 were considered to be statistically significant.

Gene and protein network modelling

Computational models were constructed using Ordinary Differential Equations and solved using MATLAB 2020a and ode15s. All code, equations and parameters used in modelling are available for on Github (https://github.com/SiFTW/MMModel/). Briefly, each modelled molecular species is modelled as having an expression and a degradation term, such that:

$$\frac{d[X]}{dt} = Expression - Degradation$$

Expression and degradation terms were assumed constant unless influenced by regulation as indicated by promotion or inhibition lines in the diagram (Figure X). Regulated reactions were modelled as described previously (Mitchell et al. Immunity 2019). Drugs are modelled by dividing expression by the drug's activity such that:

$$Expression(t) = ExpressionWithoutDrug(t)/drug(t)$$

The drug's activity was assumed to rapidly increase within 15 minutes and then slowly decay over 48 hours as shown in supplemental figure X.

Conflicts of interest: The authors declare no competing interests.

Acknowledgments

This research was supported by a Wellcome Trust Institutional Strategic Support Fund

References

1. Klein, U., et al., *Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination.* Nat Immunol, 2006. **7**(7): p. 773-82.
2. Agnarelli, A., T. Chevassut, and E.J. Mancini, *IRF4 in multiple myeloma-Biology, disease and therapeutic target.* Leuk Res, 2018. **72**: p. 52-58.
3. Wang, L., et al., *Gene expression profiling identifies IRF4-associated molecular signatures in hematological malignancies.* PLoS One, 2014. **9**(9): p. e106788.
4. Iida, S., et al., *Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma.* Nat Genet, 1997. **17**(2): p. 226-30.
5. Yoshida, S., et al., *Detection of MUM1/IRF4-IgH fusion in multiple myeloma.* Leukemia, 1999. **13**(11): p. 1812-6.
6. Shaffer, A.L., et al., *IRF4 addiction in multiple myeloma.* Nature, 2008. **454**(7201): p. 226-31.
7. Bai, H., et al., *Bone marrow IRF4 level in multiple myeloma: an indicator of peripheral blood Th17 and disease.* Oncotarget, 2017. **8**(49): p. 85392-85400.
8. Morelli, E., et al., *Selective targeting of IRF4 by synthetic microRNA-125b-5p mimics induces anti-multiple myeloma activity in vitro and in vivo.* Leukemia, 2015. **29**(11): p. 2173-83.
9. Rajkumar, S.V., *Multiple myeloma: 2016 update on diagnosis, risk-stratification, and management.* Am J Hematol, 2016. **91**(7): p. 719-34.
10. Dingli, D., et al., *Therapy for Relapsed Multiple Myeloma: Guidelines From the Mayo Stratification for Myeloma and Risk-Adapted Therapy.* Mayo Clin Proc, 2017. **92**(4): p. 578-598.
11. Neri, P., *Enhancer Deregulation in Myeloma.* Blood, 2018. **132**: p. SCI-38.
12. Lin, Y., K. Wong, and K. Calame, *Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation.* Science, 1997. **276**(5312): p. 596-9.
13. Chaidos, A., V. Caputo, and A. Karadimitris, *Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: emerging preclinical and clinical evidence.* Therapeutic advances in hematology, 2015. **6**(3): p. 128-141.
14. Shi, J. and C.R. Vakoc, *The mechanisms behind the therapeutic activity of BET bromodomain inhibition.* Mol Cell, 2014. **54**(5): p. 728-36.
15. Delmore, J.E., et al., *BET bromodomain inhibition as a therapeutic strategy to target c-Myc.* Cell, 2011. **146**(6): p. 904-17.
16. Ogryzko, V.V., et al., *The transcriptional coactivators p300 and CBP are histone acetyltransferases.* Cell, 1996. **87**(5): p. 953-9.
17. Ebrahimi, A., et al., *Bromodomain inhibition of the coactivators CBP/EP300 facilitate cellular reprogramming.* Nature Chemical Biology, 2019. **15**(5): p. 519-528.
18. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.

19. Conery, A.R., et al., *Bromodomain inhibition of the transcriptional coactivators CBP/EP300 as a therapeutic strategy to target the IRF4 network in multiple myeloma.* Elife, 2016. **5**.
20. Chekler, E.L., et al., *Transcriptional Profiling of a Selective CREB Binding Protein Bromodomain Inhibitor Highlights Therapeutic Opportunities.* Chem Biol, 2015. **22**(12): p. 1588-96.
21. Hann, S.R. and R.N. Eisenman, *Proteins encoded by the human c-myc oncogene: differential expression in neoplastic cells.* Mol Cell Biol, 1984. **4**(11): p. 2486-97.
22. Badeaux, A.I. and Y. Shi, *Emerging roles for chromatin as a signal integration and storage platform.* Nat Rev Mol Cell Biol, 2013. **14**(4): p. 211-24.
23. Mertz, J.A., et al., *Targeting MYC dependence in cancer by inhibiting BET bromodomains.* Proc Natl Acad Sci U S A, 2011. **108**(40): p. 16669-74.
24. Shi, J., et al., *Potent Activity of the Bromodomain Inhibitor OTX015 in Multiple Myeloma.* Mol Pharm, 2018. **15**(9): p. 4139-4147.
25. Garcia-Carpizo, V., et al., *CREBBP/EP300 bromodomains are critical to sustain the GATA1/MYC regulatory axis in proliferation.* Epigenetics Chromatin, 2018. **11**(1): p. 30.
26. Watanabe, N., et al., *Activation of IFN-beta element by IRF-1 requires a posttranslational event in addition to IRF-1 synthesis.* Nucleic Acids Res, 1991. **19**(16): p. 4421-8.
27. Prakash, A. and D.E. Levy, *Regulation of IRF7 through cell type-specific protein stability.* Biochemical and biophysical research communications, 2006. **342**(1): p. 50-56.
28. Stringer, D.K. and R.C. Piper, *Terminating protein ubiquitination: Hasta la vista, ubiquitin.* Cell cycle (Georgetown, Tex.), 2011. **10**(18): p. 3067-3071.
29. Zhou, L., et al., *USP15 inhibits multiple myeloma cell apoptosis through activating a feedback loop with the transcription factor NF-κBp65.* Experimental & molecular medicine, 2018. **50**(11): p. 151-151.
30. Wei, R., et al., *Deubiquitinases in cancer.* Oncotarget, 2015. **6**(15): p. 12872-12889.
31. Guo, Z., et al., *Ubiquitin specific peptidase 4 stabilizes interferon regulatory factor protein and promotes its function to facilitate interleukin-4 expression in T helper type 2 cells.* Int J Mol Med, 2017. **40**(4): p. 979-986.
32. Agnarelli, A., et al., *Cell-specific pattern of berberine pleiotropic effects on different human cell lines.* Scientific Reports, 2018. **8**(1): p. 10599.

Figure legends

Figure 1. Characterization of the effect of JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30 treatments on MM cell lines viability. Reduction of KMS-12-BM **(a)**, NCI-H929 **(b)**, SKMM-1 **(c)** and OCI-AML3 **(d)** cell viability after treatment with different concentrations of bromodomain inhbitors for 48hrs. Cell survival is plotted against the logarithm of treatment concentrations. JQ1 (red curves), JQ1+SGC-CBPEP30 (purple curves), OTX015 (pink curves), SGC-CBP30 (brown curves) and ISOX-DUAL (light blue curves). Results are represented as mean±Standard Error of Mean (SEM) of triplicate assays. **(e)** Graph shows the IC_{50} values of JQ1, JQ1+SGC-CBP30, OTX015, SGC-CBP/EP30, ISOX-DUAL after 48 hours treatments of KMS-12-BM (green bars), NCI-H929 (black bars), OCI-AML3 (blue bars) and SKMM-1 (orange bars) cells.

Figure 2. IRF4 and MYC protein levels in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30. Changes in MYC and IRF4 protein levels were analysed by Western Blot following IC_{50} drug treatments for 48 hours in KMS-12-BM, SKMM-1, NCI-H929 and OCI-AML3. The control (CTRL) is 2mM DMSO treatment. β-actin was used as loading control. Quantification was performed by using LI-COR machine and protein levels were expressed relative to the control treatment.

Figure 3. *IRF4* and *MYC* mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30.

IRF4 and *MYC* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 48hrs in KMS-12-BM (green bars), SKMM-1 (orange bars), NCI-H929 (black bars) and OCI-AML3 (blue bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

Figure 4. IRF4 downstream gene mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30.

KLF2 and *PRDM1* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 48hrs in KMS-12-BM (green bars), SKMM-1 (orange bars), and NCI-H929 (black bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

Figure 5. MYC downstream gene mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30.

CDK4 and *hTERT* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 48hrs in KMS-12-BM (green bars), SKMM-1 (orange bars), and NCI-H929 (black bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

Figure 6. Computational model of the molecular regulatory network in MM cells.

(a) Systems Biology Graphical Notation (SBGBN) diagram of the model of IRF4, cMyc and Blimp1 regulation. Positive regulation is indicated by lines capped with circles. Negative regulation is indicated by lines capped with bars. (b) Experimentally measured expression of the indicated molecular species in H929, SKMM-1, KMS cell lines exposed to SGC-CBP30, JQ1, OTX015, ISOX-DUAL, and JQ1+SGC-CBP30 combination. Each shaded region represents the standard deviation of 3 experimental replicates. The modelled response is shown with a solid line. The model assumes a half-life for IRF4 of 7 hour. The squared distance between the mean experimental result and modelled response for each timepoint is shown in the bottom right with colours consistent with other panels.

Figure 7. Analysis of IRF4 stability in MM cell lines and updated computational model of the molecular regulatory network in MM cell . (a) KMS-12-BM, SKMM-1, NCI-H929 were incubated with 10µg/mL cycloheximide for the indicated time points and cell lysates analysed by Western blotting for protein levels of IRF4 and MYC. β-actin was used as a loading control. **(b)** Experimentally measured expression of the indicated molecular species in H929, SKMM-1, KMS cell lines exposed to SGC-CBP30, JQ1, OTX015, ISOX-DUAL, and JQ1+SGC-CBP30 combination. Each shaded region represents the standard deviation of 3 experimental replicates. The modelled response is shown with a solid line. The model uses the experimentally determined IRF4 half-life. The squared distance between the mean experimental result and modelled response for each timepoint is shown in the bottom right with colours consistent with other panels.

Figure 8. Computational model simulating the effect of a drug acting on *MYC* transcription, *IRF4* transcription or both.

(a) Systems Biology Graphical Notation (SBGBN) diagram of the model of IRF4, MYC and Blimp1 regulation. Positive regulation is indicated by lines capped with circles. Negative regulation is indicated by lines capped with bars**.** Drugs are shown impacting IRF4 transcription (A) and MYC transcription (B) **(b)** Experimentally measured expression of the indicated molecular species in H929, SKMM-1, KMS cell lines exposed to SGC-CBP30, JQ1, OTX015, ISOX-DUAL, and JQ1+SGC-CBP30 combination. The impact of single targeting IRF4 (A, left) and Myc (B, middle) is shown, along with the combination (A+B, right). Each shaded region represents the standard deviation of 3 experimental replicates. The modelled response is shown with a solid line. The model uses the experimentally determined IRF4 half-life. The squared distance between the mean experimental result and modelled response for each timepoint is shown in the bottom right with colours consistent with other panels.

Supplementary Figure 1. IRF4 and MYC protein levels in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30 for 4, 8 and 24 hours. Changes in MYC and IRF4 protein levels were analysed by Western Blot following IC_{50} drug treatments for 4, 8 and 24 hours in KMS-12-BM, SKMM-1, NCI-H929 and OCI-AML3. The control (CTRL) is 2mM DMSO treatment. β-actin was used as loading control. Quantification was performed by using LI-COR machine and protein levels were expressed relative to the control treatment.

Supplementary Figure 2. *IRF4* and *MYC* mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30 for 4, 8 and 24 hours.

IRF4 and *MYC* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 4, 8 and 24 hours in KMS-12-BM (green bars), SKMM-1 (orange bars), NCI-H929 (black bars) and OCI-AML3 (blue bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

Supplementary Figure 3. IRF4 downstream gene mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30 for 4, 8 and 24 hours.

KLF2 and *PRDM1* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 4, 8 and 24 hours in KMS-12-BM (green bars), SKMM-1 (orange bars), and NCI-H929 (black bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

Supplementary Figure 4. MYC downstream gene mRNA expression in MM cell lines following treatment with JQ1, OTX015, SGC-CBP30, ISOX-DUAL and JQ1+ SGC-CBP30 for 4, 8 and 24 hours.

CDK4 and *hTERT* mRNA expression was analysed by qPCR following IC_{50} drug treatments for 4, 8 and 24 hours in KMS-12-BM (green bars), SKMM-1 (orange bars), and NCI-H929 (black bars) cells. The control (CTRL) is 2mM DMSO treatment. Transcript levels were normalized against β-actin expression and expressed relative to the control treatment. Data are shown as mean±SEM. A t-test was performed with reference to the control. *P<0.05, **P<0.01, ***P < 0.001.

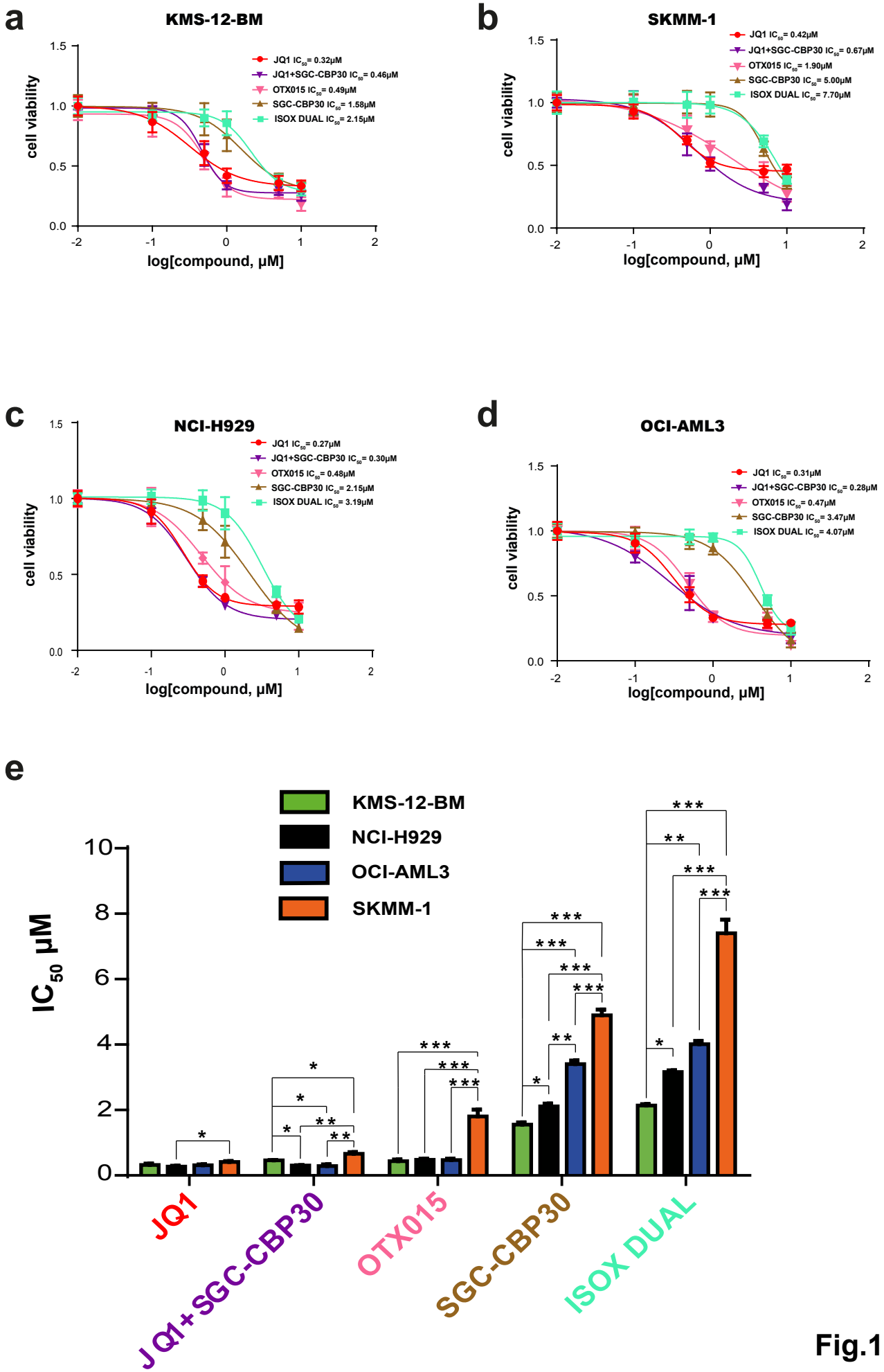

a
KMS-12-BM
JQ1 IC50= 0.32μM
JQ1+SGC-CBP30 IC50= 0.46μM
OTX015 IC50= 0.49μM
SGC-CBP30 IC50= 1.58μM
ISOX DUAL IC50= 2.15μM
cell viability
log[compound, μM]
b
SKMM-1
JQ1 IC50= 0.42μM
JQ1+SGC-CBP30 IC50= 0.67μM
OTX015 IC50= 1.90μM
SGC-CBP30 IC50= 5.00μM
ISOX DUAL IC50= 7.70μM
cell viability
log[compound, μM]
c
NCI-H929
JQ1 IC50= 0.27μM
JQ1+SGC-CBP30 IC50= 0.30μM
OTX015 IC50= 0.48μM
SGC-CBP30 IC50= 2.15μM
ISOX DUAL IC50= 3.19μM
cell viability
log[compound, μM]
d
OCI-AML3
JQ1 IC50= 0.31μM
JQ1+SGC-CBP30 IC50= 0.28μM
OTX015 IC50= 0.47μM
SGC-CBP30 IC50= 3.47μM
ISOX DUAL IC50= 4.07μM
cell viability
log[compound, μM]
e
KMS-12-BM
NCI-H929
OCI-AML3
SKMM-1
IC50 μM
JQ1
JQ1+SGC-CBP30
OTX015
SGC-CBP30
ISOX DUAL

Fig.1

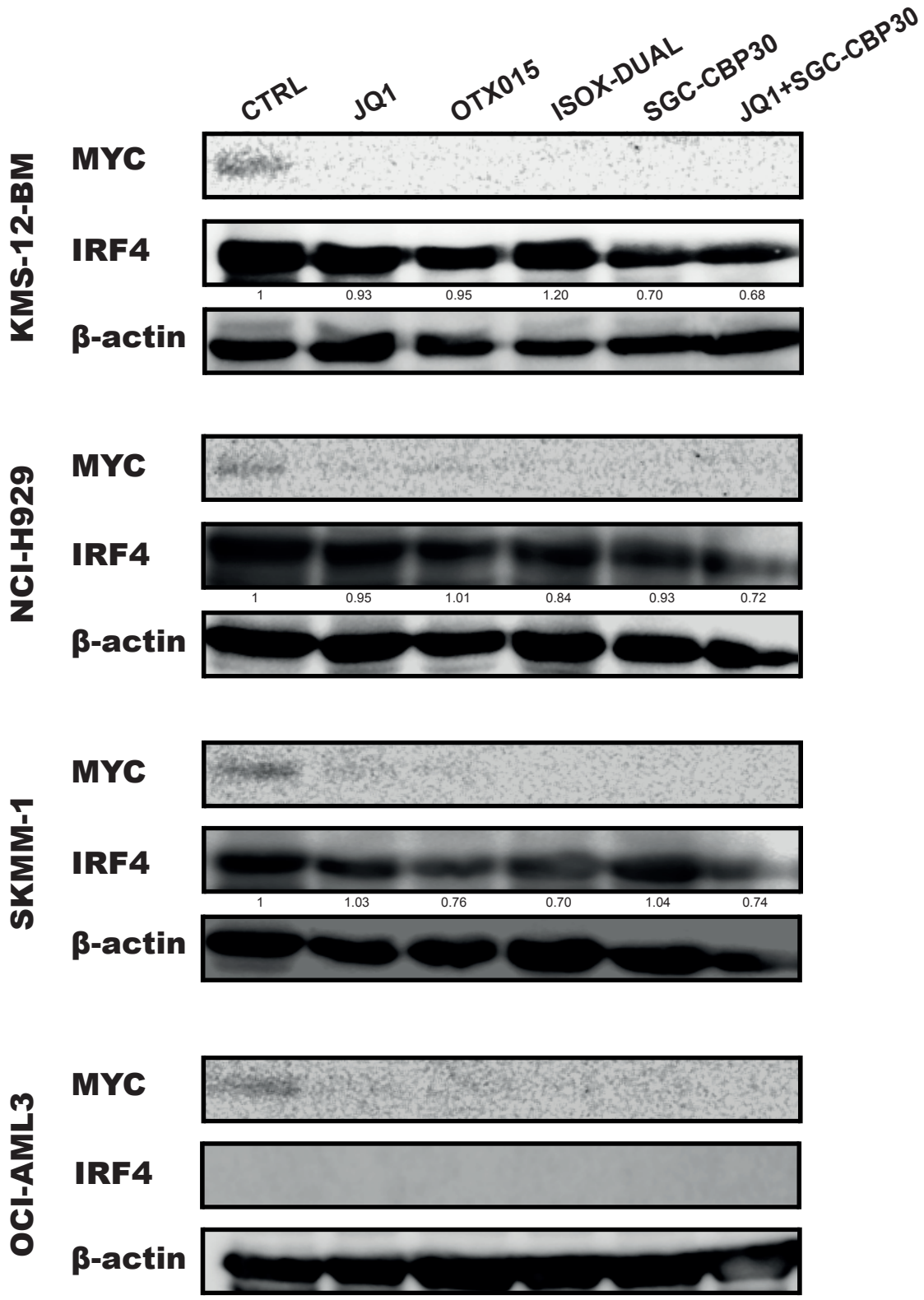
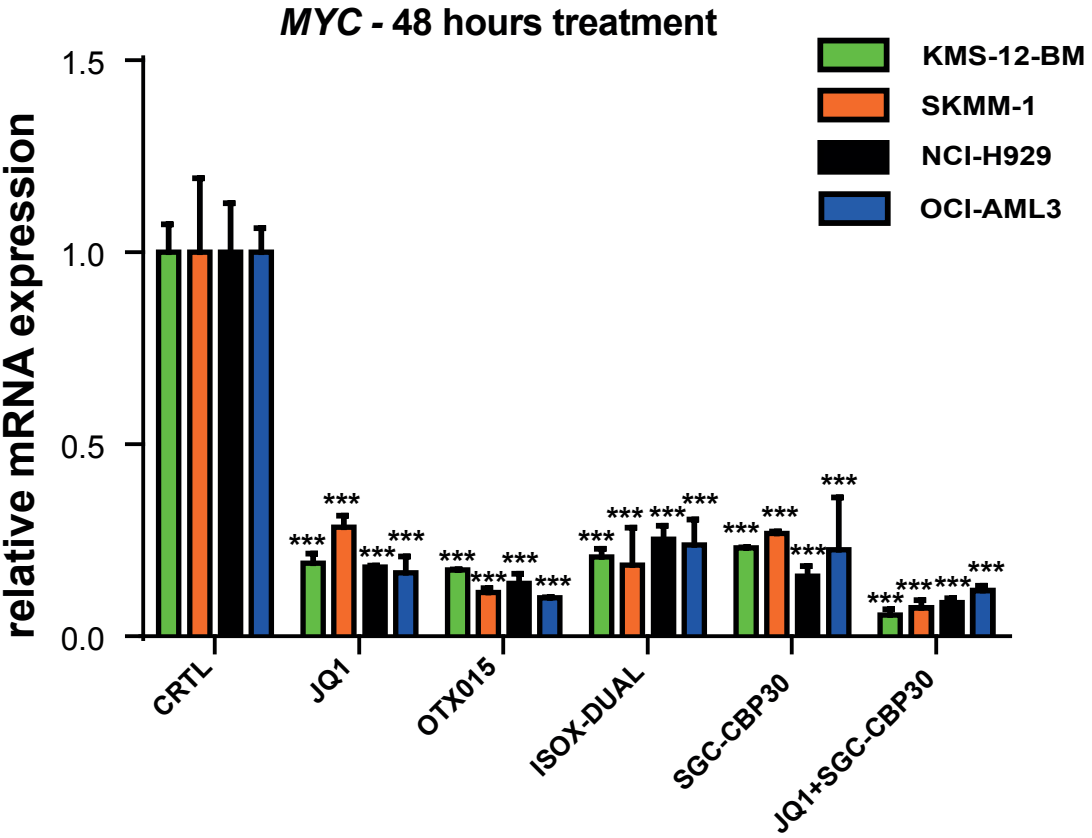

CTRL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30
KMS-12-BM
MYC
IRF4
1
0.93
0.95
1.20
0.70
0.68
β-actin
NCI-H929
MYC
IRF4
1
0.95
1.01
0.84
0.93
0.72
β-actin
SKMM-1
MYC
IRF4
1
1.03
0.76
0.70
1.04
0.74
β-actin
OCI-AML3
MYC
IRF4
β-actin

Fig.2


MYC - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
OCI-AML3
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

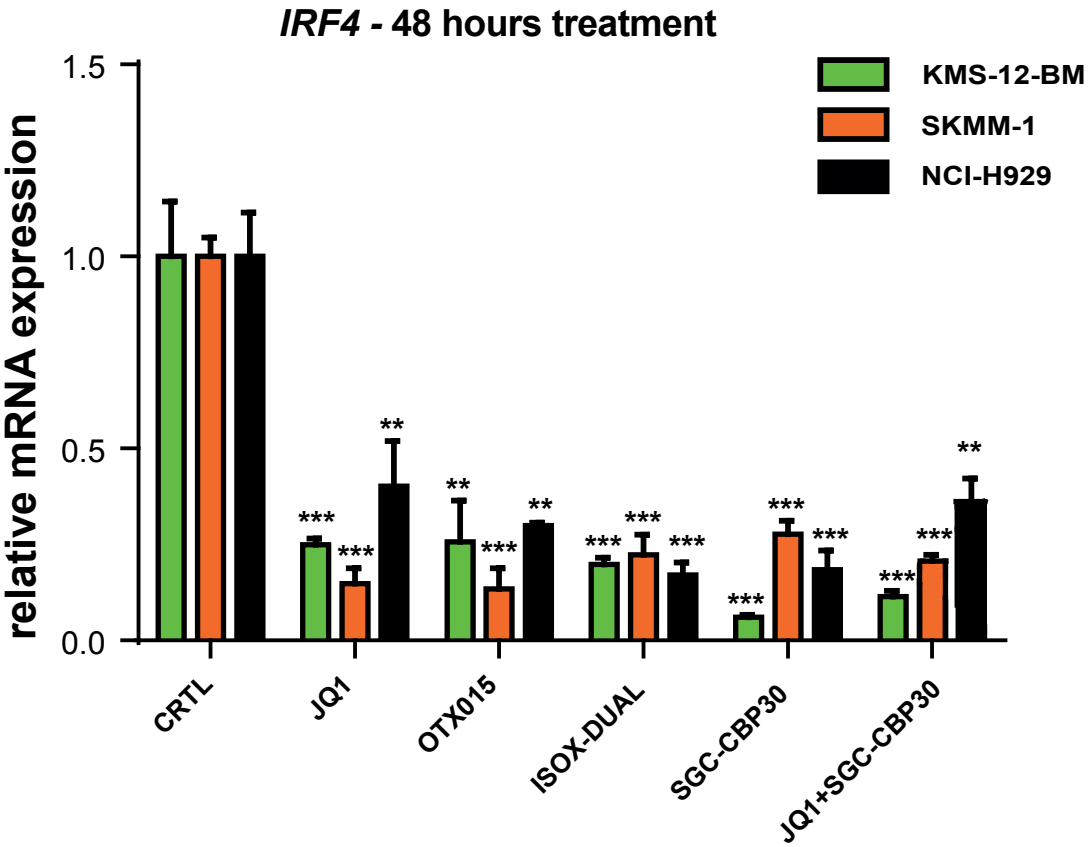
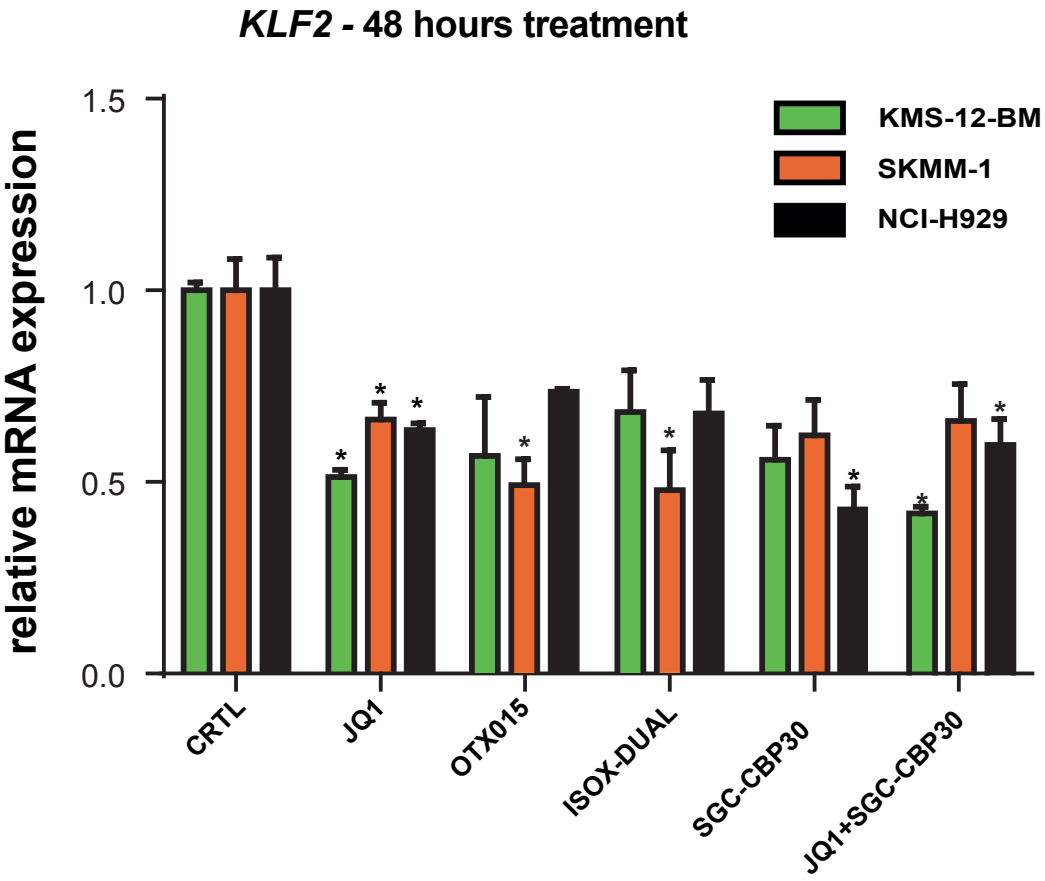

IRF4 - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

Fig.3


KLF2 - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

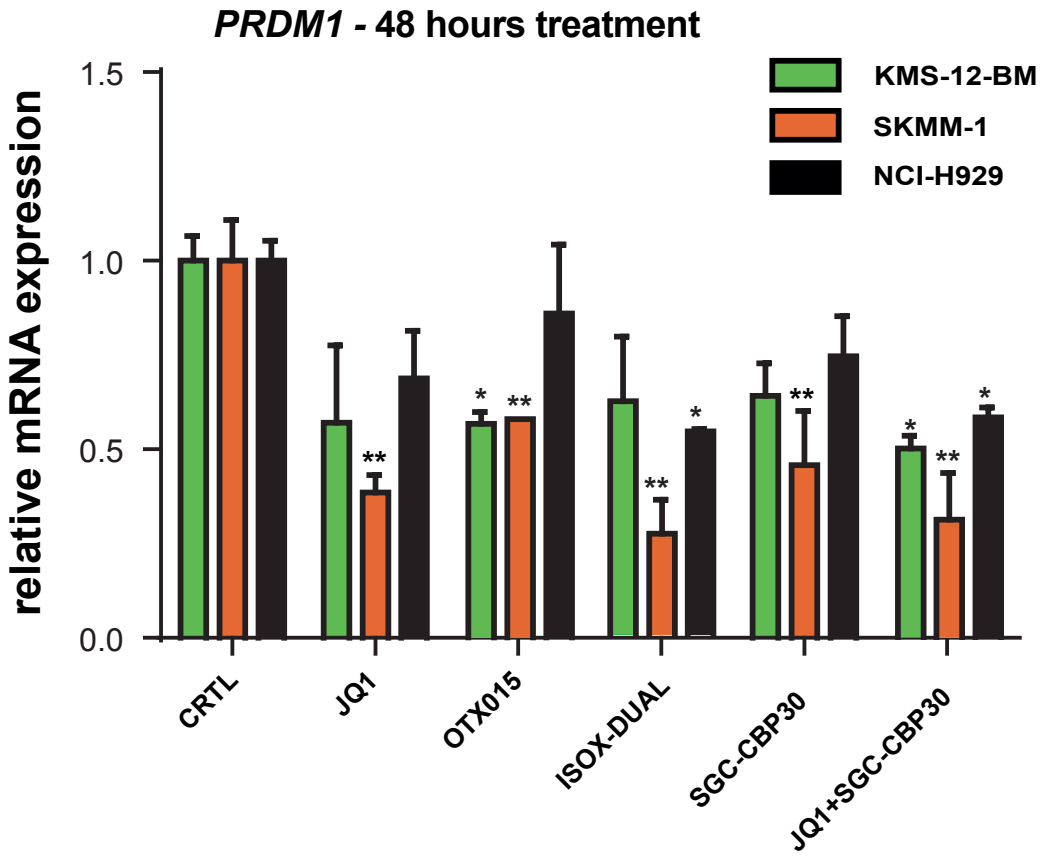
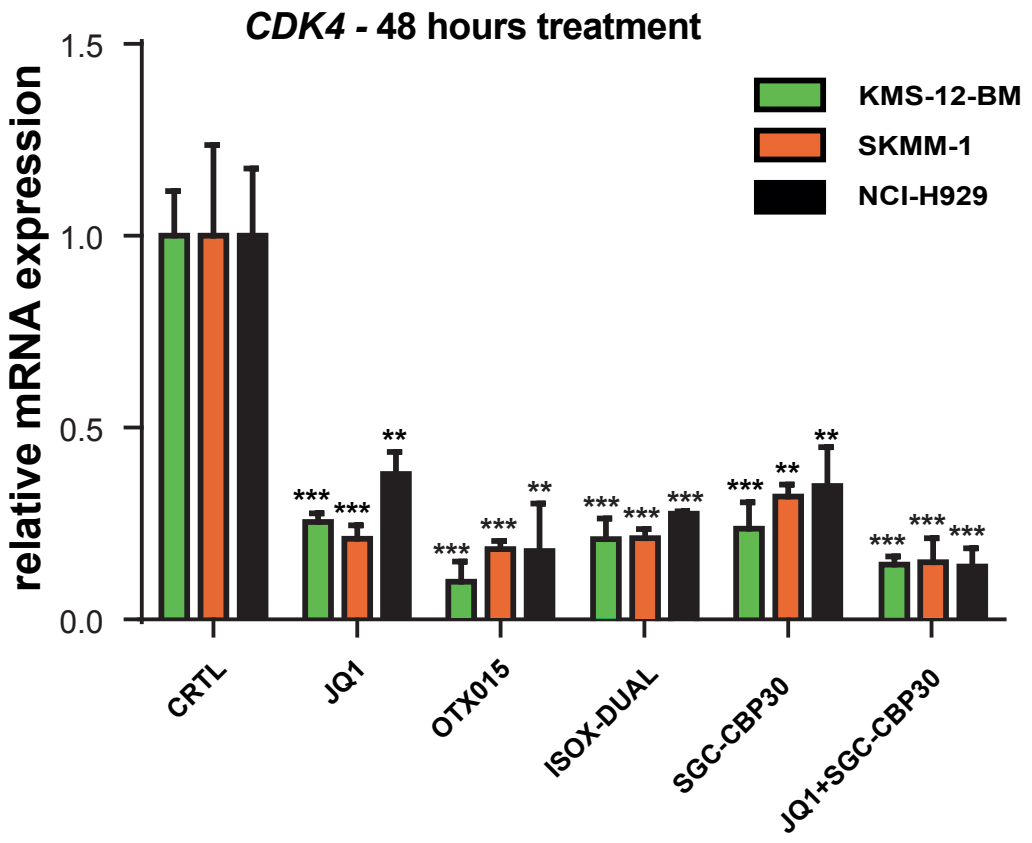

PRDM1 - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

Fig.4


CDK4 - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

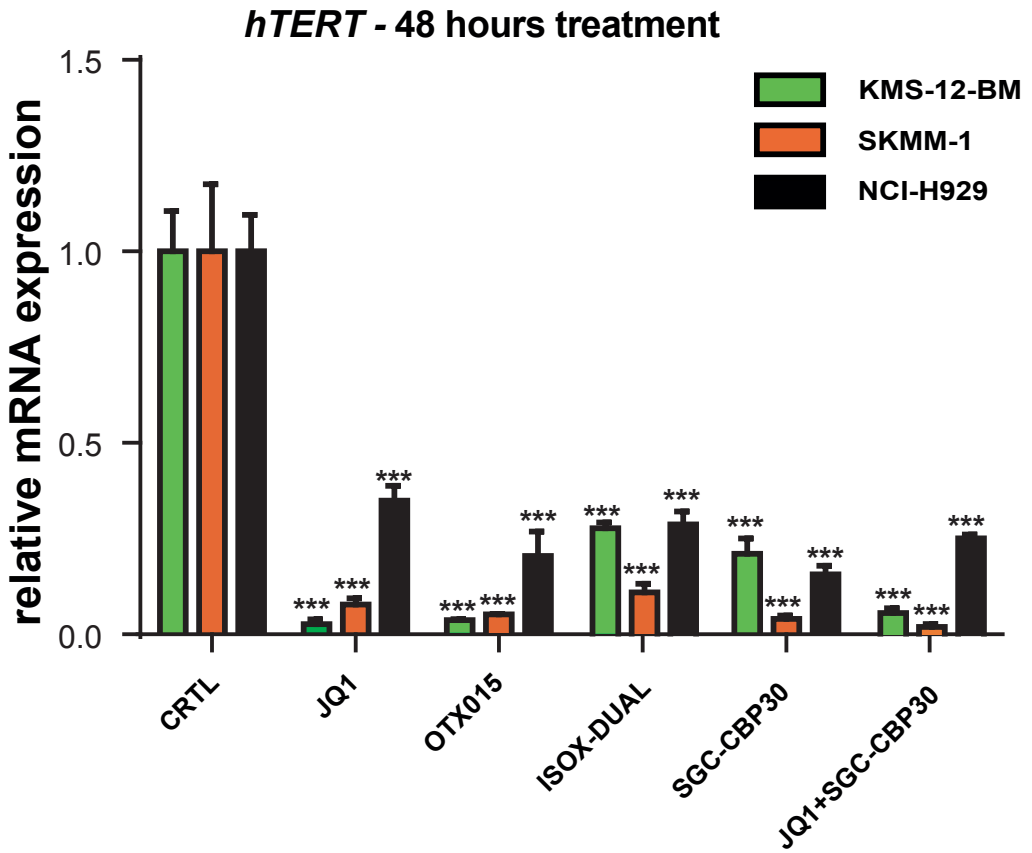

hTERT - 48 hours treatment
relative mRNA expression
1.5
1.0
0.5
0.0
KMS-12-BM
SKMM-1
NCI-H929
CRTL
JQ1
OTX015
ISOX-DUAL
SGC-CBP30
JQ1+SGC-CBP30

Fig.5

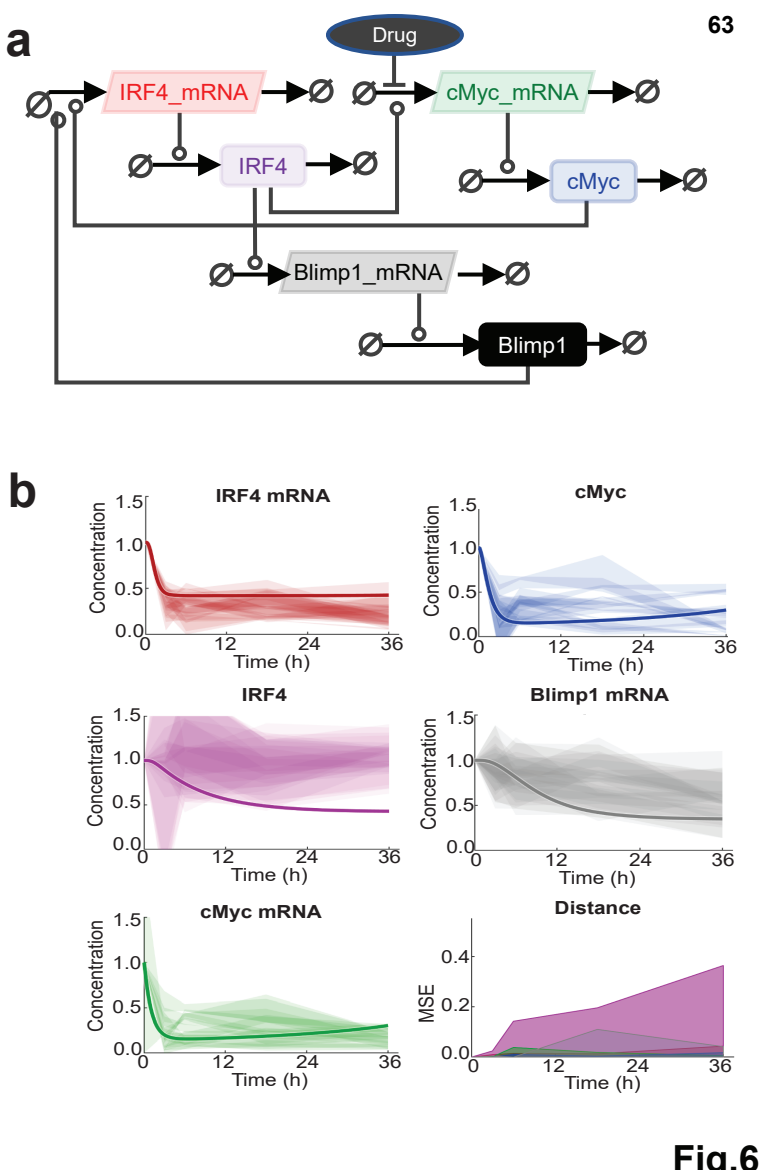

a
Drug
IRF4_mRNA
cMyc_mRNA
IRF4
cMyc
Blimp1_mRNA
Blimp1
b
IRF4 mRNA
cMyc
IRF4
Blimp1 mRNA
cMyc mRNA
Distance
Concentration
MSE
Time (h)
0
12
24
36
0.0
0.5
1.0
1.5
0.2
0.4

Fig.6

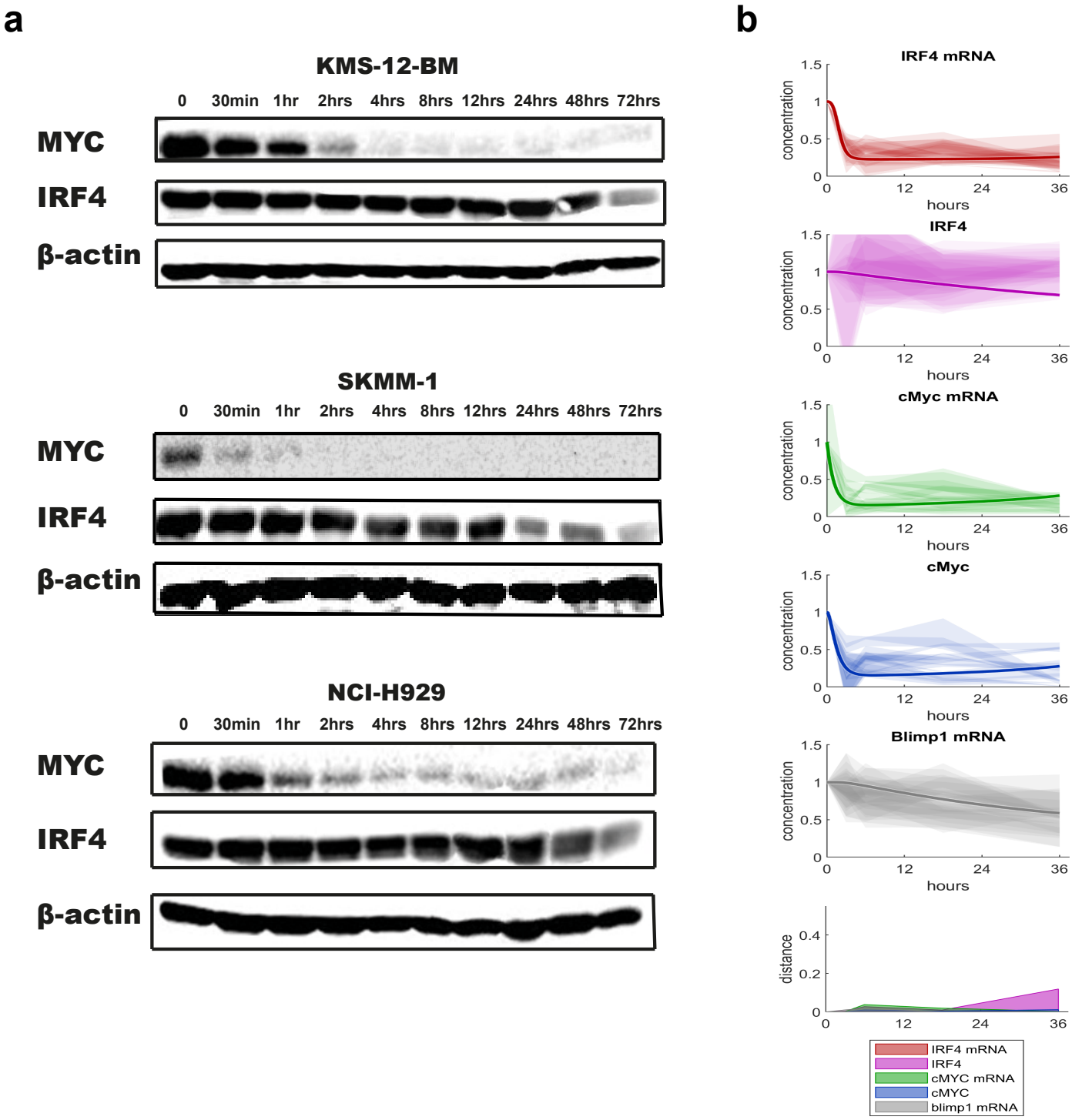

a
KMS-12-BM
0 30min 1hr 2hrs 4hrs 8hrs 12hrs 24hrs 48hrs 72hrs
MYC
IRF4
β-actin
SKMM-1
0 30min 1hr 2hrs 4hrs 8hrs 12hrs 24hrs 48hrs 72hrs
MYC
IRF4
β-actin
NCI-H929
0 30min 1hr 2hrs 4hrs 8hrs 12hrs 24hrs 48hrs 72hrs
MYC
IRF4
β-actin
b
IRF4 mRNA
IRF4
cMyc mRNA
cMyc
Blimp1 mRNA
concentration
hours
distance
IRF4 mRNA
IRF4
cMYC mRNA
cMYC
blimp1 mRNA

Fig.7

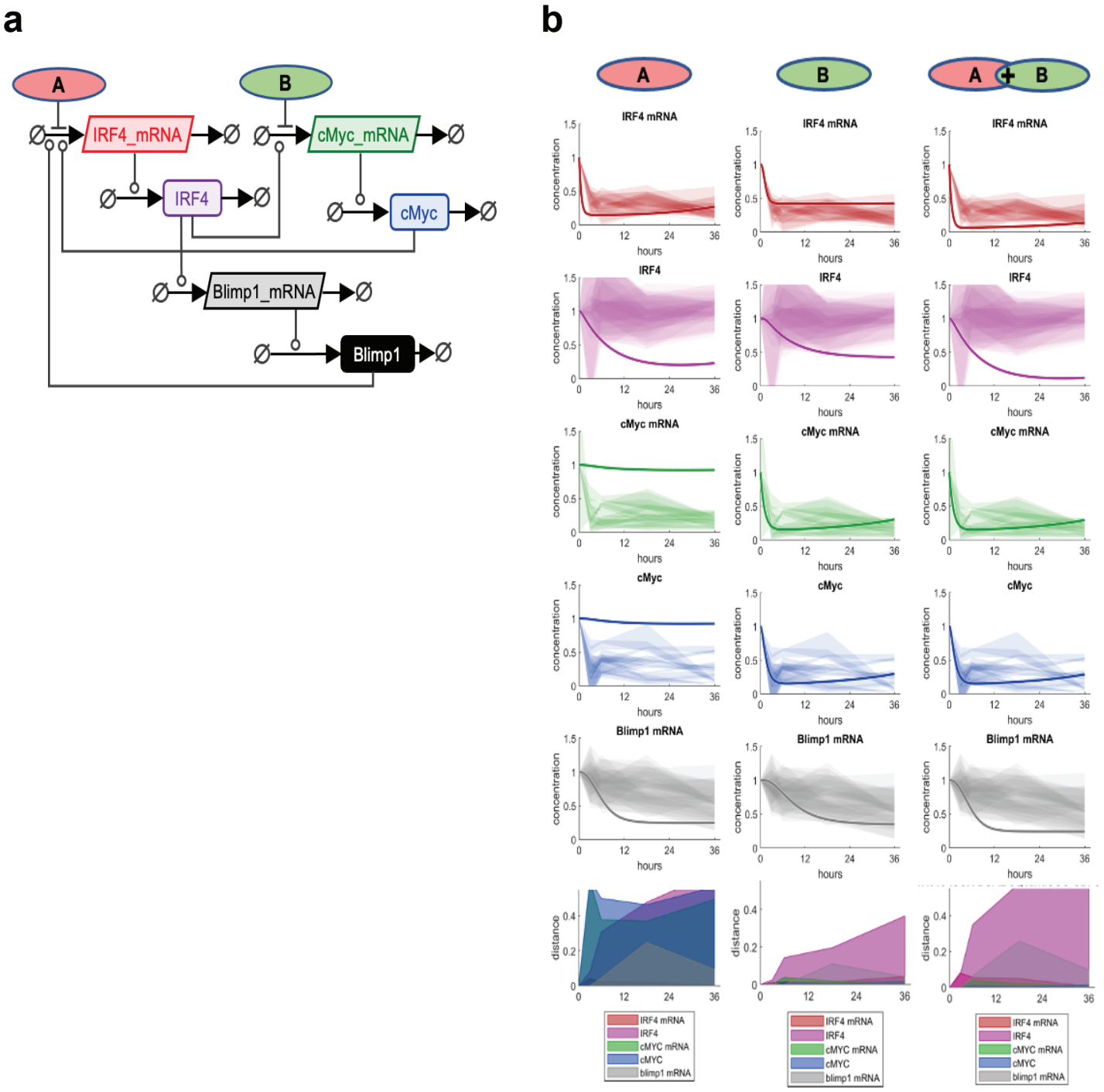
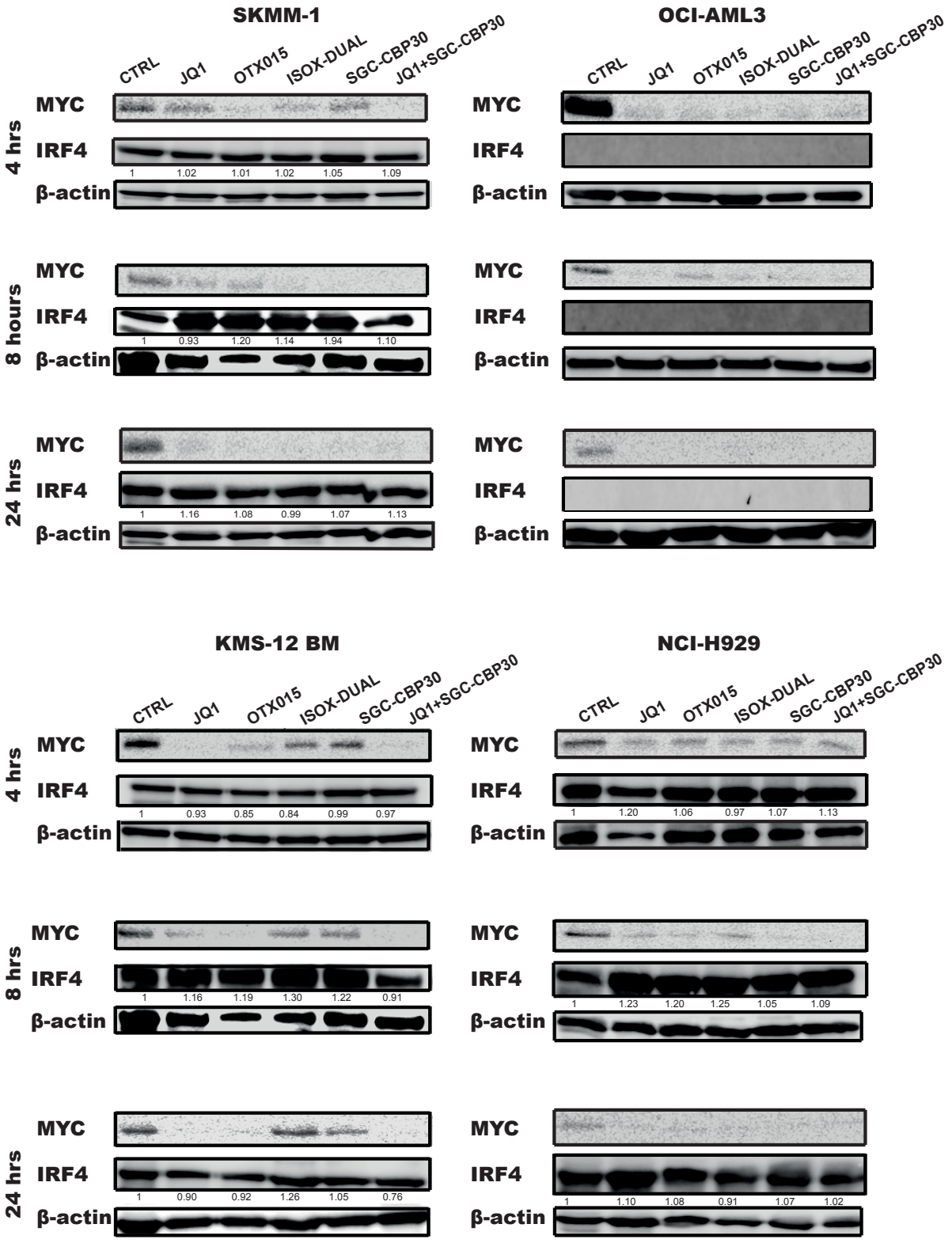
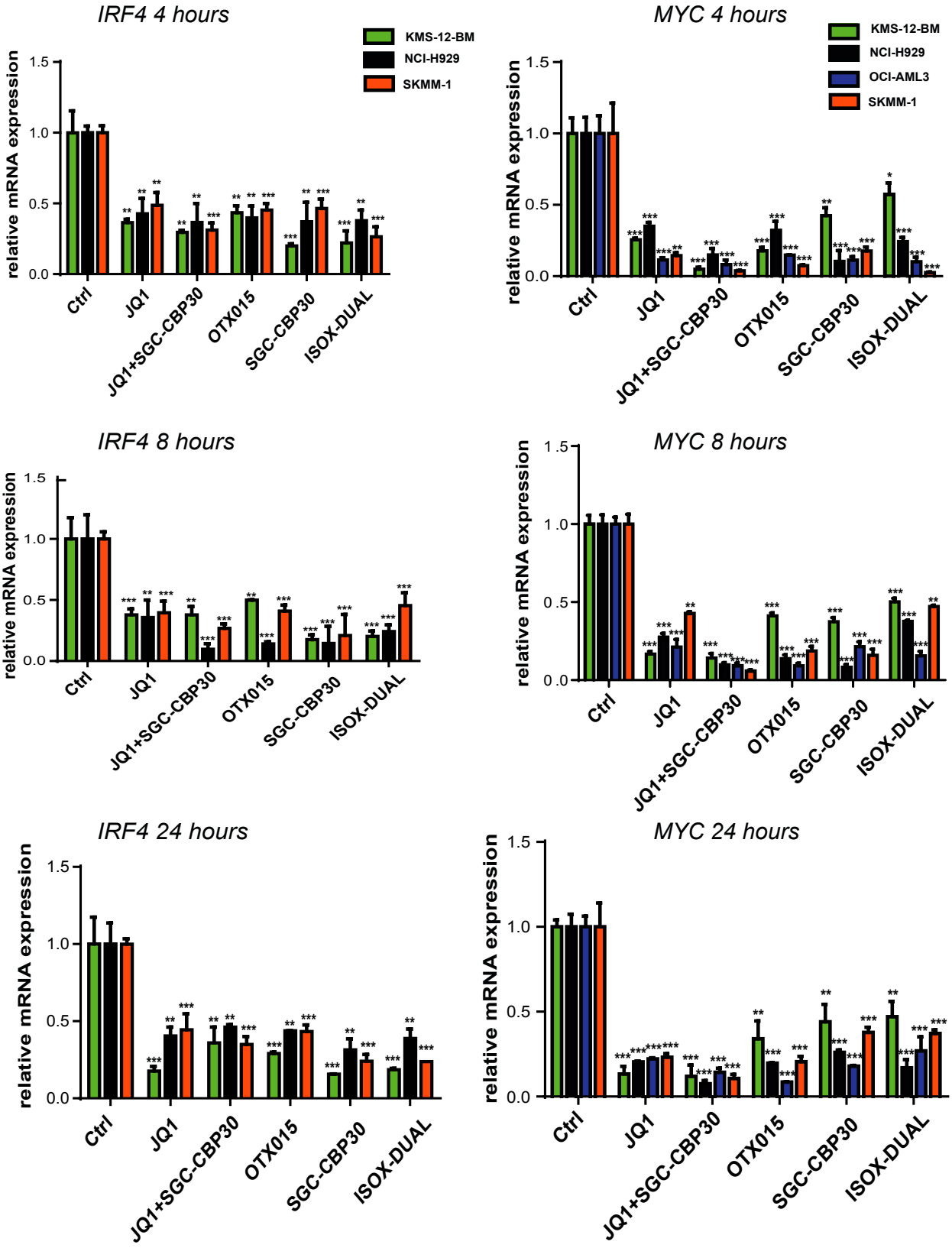

a
A
B
IRF4_mRNA
cMyc_mRNA
IRF4
cMyc
Blimp1_mRNA
Blimp1
b
A
B
A + B
IRF4 mRNA
IRF4
cMyc mRNA
cMyc
Blimp1 mRNA
concentration
hours
distance
IRF4 mRNA
IRF4
cMYC mRNA
cMYC
blimp1 mRNA

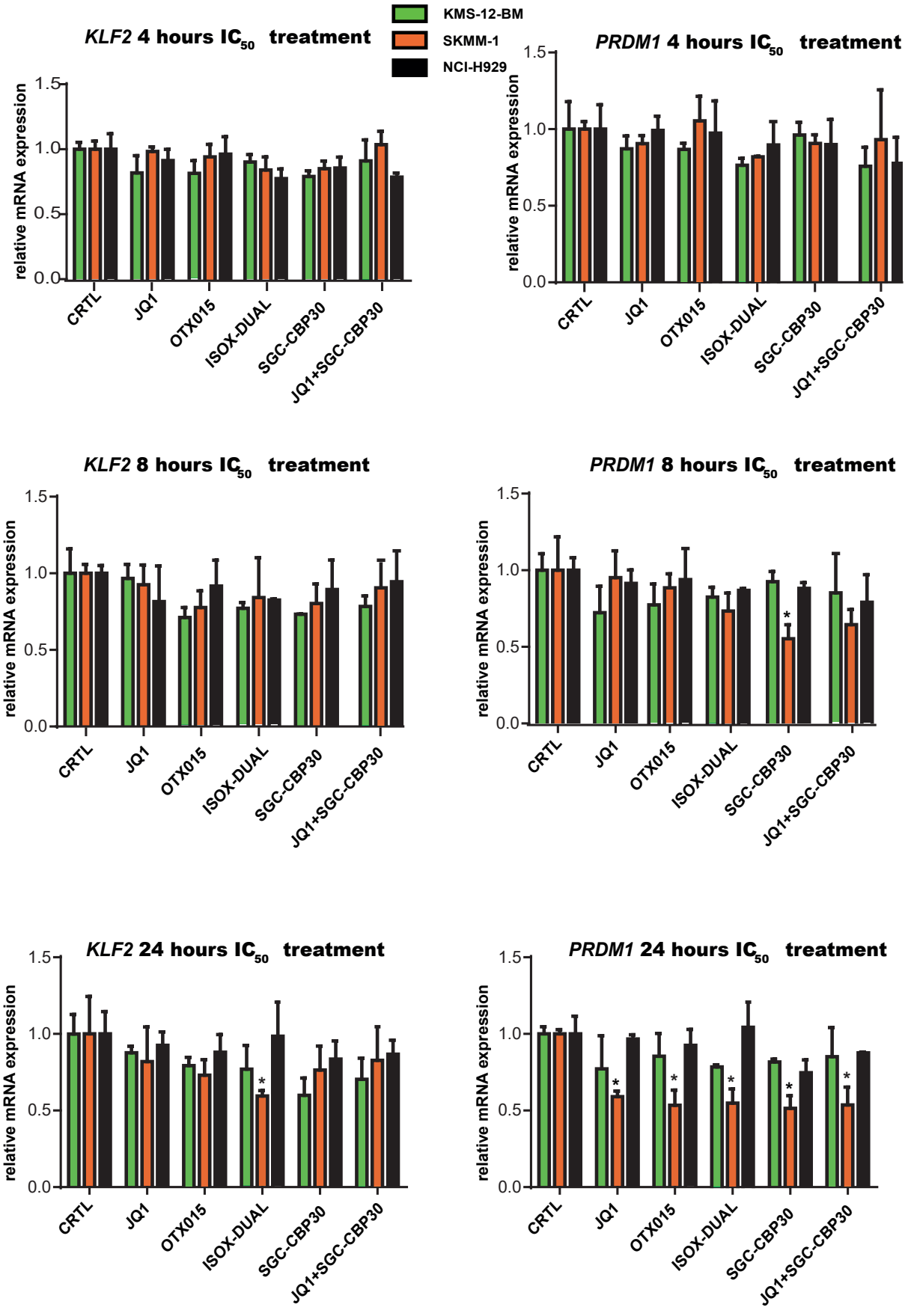
Fig.8



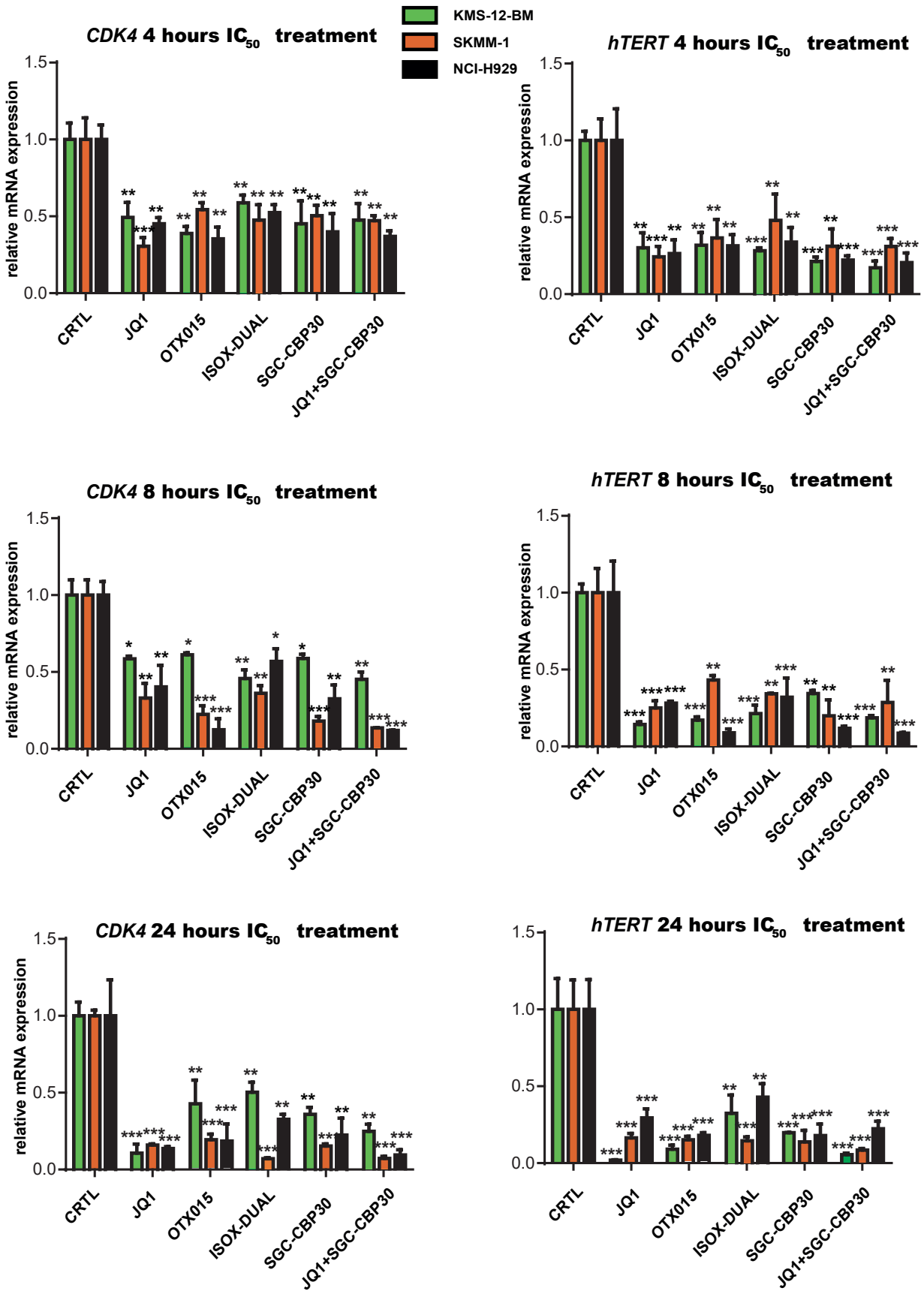
Supplementary Fig.1



Supplementary Fig.2



Supplementary Fig.3



Supplementary Fig.4

4. Chapter 4. Phosphorus and sulphur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4

Alessandro Agnarelli,[a] Kamel El Omari,[b,c] Ramona Duman,[b,c] Armin Wagner[b,c] and Erika J. Mancini[a]*

[a]School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom, [b]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, United Kingdom, and [c]Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, United Kingdom. *Correspondence e-mail: erika.mancini@sussex.ac.uk

Received 1 May 2021
Accepted 21 June 2021

Edited by M. J. van Raaij, Centro Nacional de Biotecnología – CSIC, Spain

Keywords: experimental phasing; native SAD; phosphorus; DNA; DNA-binding proteins; IRF4; interferon regulatory factor 4.

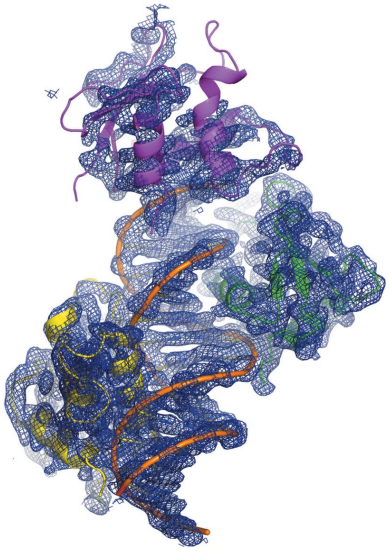
PDB reference: DNA-binding domain of interferon regulatory factor 4, 7o56

Supporting information: this article has supporting information at journals.iucr.org/f

Pivotal to the regulation of key cellular processes such as the transcription, replication and repair of DNA, DNA-binding proteins play vital roles in all aspects of genetic activity. The determination of high-quality structures of DNA-binding proteins, particularly those in complexes with DNA, provides crucial insights into the understanding of these processes. The presence in such complexes of phosphate-rich oligonucleotides offers the choice of a rapid method for the routine solution of DNA-binding proteins through the use of long-wavelength beamlines such as I23 at Diamond Light Source. This article reports the use of native intrinsic phosphorus and sulfur single-wavelength anomalous dispersion methods to solve the complex of the DNA-binding domain (DBD) of interferon regulatory factor 4 (IRF4) bound to its interferon-stimulated response element (ISRE). The structure unexpectedly shows three molecules of the IRF4 DBD bound to one ISRE. The sole reliance on native intrinsic anomalous scattering elements that belong to DNA–protein complexes renders the method of general applicability to a large number of such protein complexes that cannot be solved by molecular replacement or by other phasing methods.

1. Introduction

DNA-binding proteins are essential components of all biological systems, where they perform crucial roles. Deregulation or mutation of DNA-binding proteins, such as transcription factors, is closely associated with the pathogenesis of several human diseases, including cancer, making them attractive therapeutic targets (Lee & Young, 2013; Hudson & Ortlund, 2014). Structure solution of protein–DNA complexes provides the basis of our understanding of normal and pathogenic DNA metabolism and underpins attempts to develop novel drugs targeting disease-associated DNA-binding proteins (Bushweller, 2019). The last ten years have witnessed a step-change increase in the number of experimentally determined protein–nucleic acid complexes. More than two thirds of all structures of complexes deposited in the Protein Data Bank (PDB) as of April 2021 (6145 out of 9204) were solved in the last ten years. However, the number of protein–nucleic acid complex structures solved remains only a small part of the deposited structures as their experimental determination often remains challenging. The lack of suitable homologous structures can be an obstacle to solving the crystallographic phase problem. Even when molecular replacement (MR) can be employed, DNA-binding proteins can be flexible and/or disordered



(Munshi *et al.*, 2018; Dyson & Komives, 2012; Varadi *et al.*, 2015), whilst nucleic acids can depart from canonical structures (Tateishi-Karimata & Sugimoto, 2020). Occasionally, multiple conformational folds are displayed, producing highly dynamic structural ensembles (Fuxreiter *et al.*, 2011). As a result, MR alone can often deliver electron-density maps that are of poor quality and are unsuitable for model building and structure solution.

Experimental phasing sidesteps the lack of homologous structures and helps in obtaining interpretable electron-density maps; however, the artificial addition of anomalous scatters by heavy-atom derivatization or selenomethionine substitution can be a time-consuming and often arduous task. On the other hand, native single-wavelength anomalous dispersion (SAD) phasing using the weak anomalous scattering signal of light atoms that are intrinsically present in proteins and nucleic acids, such as phosphorus, sulfur, chlorine, potassium and calcium, obviates the need for covalent or noncovalent heavy-atom modifications. In comparison with metals, however, the anomalous scattering signal from these light atoms is relatively small, and native SAD phasing is critically dependent on accurate recording (Rose & Wang, 2016). The challenges associated with native SAD phasing are illustrated by the observation that whilst the first native SAD structure was reported in 1981 (Hendrickson & Teeter, 1981), it took almost 20 years for more structures to be solved (Dauter *et al.*, 1999; Liu *et al.*, 2000) by using solvent-flattening approaches (Wang, 1985). Over the past 20 years, advances in hardware, software, data-collection methods and strategies have allowed the collection of highly accurate data with an increase in the anomalous signal-to-noise ratio, which in turn has enabled the 'routine' use of native SAD phasing for *de novo* structure solution (Rose *et al.*, 2015). Furthermore, the use of native SAD phasing, for example from S atoms, has been successfully used in combination with MR (MRSAD) to overcome model bias and assist with phasing, model building and refinement (Schuermann & Tanner, 2003).

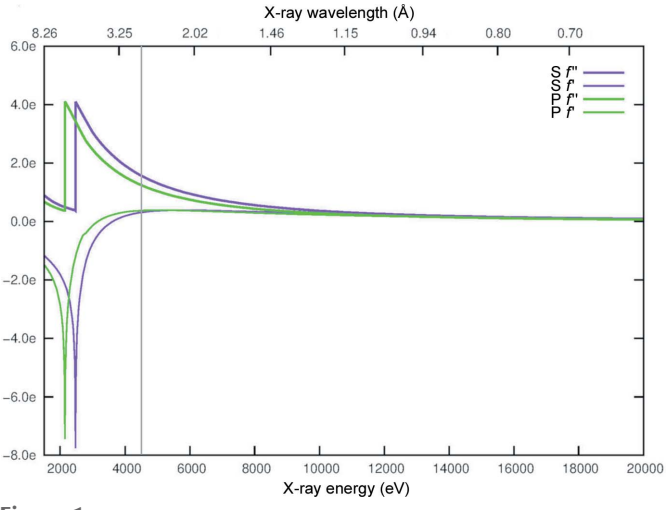


Figure 1
Theoretical values of f' and f'' for the elements sulfur (purple) and phosphorus (green) over energies from 1.5 to 20 keV. The grey bar indicates the wavelength/energy (2.7552 Å/4.5 keV) at which the IRF4 DBD–ISRE DNA data sets were collected. The plot was generated using the http://www.bmsc.washington.edu/scatter website.

The use of the intrinsic anomalous signal of phosphorus to phase oligonucleotide crystal diffraction data using SAD (P-SAD) was first theoretically and practically demonstrated in 2001 (Dauter & Adamiak, 2001). However, when not in complex with proteins, phasing of nucleic acid structures using P-SAD is very challenging and has in effect been limited to a very small number of cases where crystals diffracted to high resolution (Raiber *et al.*, 2015; Luo *et al.*, 2014). There are two possible explanations for the lack of success of P-SAD on nucleic acid structures: the high mobility and consequent high *B* factors of P atoms in the nucleic acid backbone (Harp *et al.*, 2016) and the reduced number of reflections available for phasing compared with the large number of P atoms (typically small unit cells and often high-symmetry space groups). Lower *B* factors and a higher ratio of reflections to sites in the substructure have been shown to be crucial for SAD phasing in general (Terwilliger *et al.*, 2016). On the other hand, because interactions with proteins usually stabilize nucleic acid backbones and the number of reflections is greater in larger unit cells, P-SAD can be routinely used for phasing protein–nucleic acid complexes as long as the anomalous signal can be precisely retrieved. The level of difficulty of extracting the intrinsic anomalous signal at in-house or synchrotron beamline wavelengths can be appreciated from a graph of f' and f'' of phosphorus, as seen in Fig. 1. In practice, the signal-to-noise ratios necessary to adequately and routinely retrieve the anomalous signal of phosphorus are achievable only with very high multiplicity data or at wavelengths that are only obtainable at state-of-the-art long-wavelength beamlines such as I23 at Diamond Light Source. This beamline operates under vacuum with a large semi-cylindrical detector (PILATUS 12M, Dectris) to minimize absorption effects and allow measurements of larger diffraction angles at longer wavelengths (Wagner *et al.*, 2016).

Here, the structure of the DNA-binding domain (DBD) of interferon regulatory factor 4 (IRF4) bound to its interferon-stimulated response element (ISRE), solved by the use of native intrinsic phosphorus and sulfur single-wavelength anomalous dispersion methods at I23, is presented. The structure shows the presence of three molecules of the IRF4 DBD bound to one molecule of DNA, which is unexpected in the light of previous studies suggesting the homodimerization of IRF4 on ISRE elements (Ochiai *et al.*, 2013). This study suggests that native intrinsic SAD methods can be used successfully and routinely on long-wavelength beamlines such as I23 to solve protein–nucleic acid structures *de novo*, eliminating the need for molecular replacement.

2. Materials and methods

2.1. Protein expression and purification

The IRF4 DBD (amino acids 20–139) was cloned into a pCDFDuet-1 bacterial expression plasmid containing an N-terminal 6×His tag, transformed into the *Escherichia coli*

BL21 strain (Novagen) and grown at 310 K by shaking at 180 rev min^{-1} in Luria–Bertani (LB) broth until the absorbance at 600 nm reached a value of 0.6. Overexpression of the fusion protein was induced by the addition of 0.4 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and growth was continued for 16 h at 291 K. The cells were harvested by centrifugation, resuspended in lysis buffer [25 m*M* HEPES pH 7.4, 150 m*M* NaCl, 5 m*M* imidazole, 0.1 m*M* $MgCl_2$, 0.01% Triton X-100, 0.5 m*M* tris(2-carboxyethyl)phosphine (TCEP), protease-inhibitor cocktail] and lysed by sonication on ice. The lysate was clarified by centrifugation at 26 700*g* for 45 min at 277 K. The supernatant was applied onto a HisPur Cobalt Resin column (Thermo Fisher) previously equilibrated with wash buffer (25 m*M* HEPES pH 7.4, 150 m*M* NaCl, 5 m*M* imidazole, 0.5 m*M* TCEP). Following a 10 min incubation at 227 K and the application of five column volumes of wash buffer, the protein was eluted by the addition of elution buffer (25 m*M* HEPES pH 7.4, 150 m*M* NaCl, 150 m*M* imidazole, 0.5 m*M* TCEP). The collected eluate was concentrated and purified by size-exclusion chromatography using a HiLoad 16/600 Superdex 75 prep-grade column (GE Healthcare) in gel-filtration buffer (20 m*M* Tris–HCl pH 8, 500 m*M* NaCl, 0.5 m*M* TCEP) at 277 K. Fractions were analysed on a 14% SDS–PAGE gel by electrophoresis and those containing the IRF4 DBD were pooled and concentrated to 10 mg ml^{-1}. Oligonucleotides containing an interferon response element (ISRE), 5′-AATAAAAGAAACCGAAAGTAA-3′ and 5′-TTTACTTTCGGTTTCTTTTAT-3′ (Eurofins Genomic), were annealed and incubated in a 1.2:1 DNA:protein molar ratio for 1 h at 277 K prior to crystallization.

2.2. Crystallization

The IRF4 DBD–ISRE complex was used to screen 384 conditions using the sitting-drop vapour-diffusion method. Initial hits appeared within a week and were optimized with an additive screen (JBScreen Plus HTS). The best crystals grew in 0.1 *M* sodium acetate pH 5.2, 5% PEG 4000, 10 m*M* EDTA at 293 K. Crystals were harvested using sample holders designed specifically for experiments on the in-vacuum I23 beamline and were successfully cryoprotected in 25% glycerol by flash-cooling in liquid nitrogen.

2.3. Data collection and processing

Diffraction data from two crystals of the IRF4 DBD–DNA complex were collected on a PILATUS 12M detector (Dectris) at ∼60 K on the long-wavelength beamline I23 at Diamond Light Source, Didcot, UK (Wagner *et al.*, 2016). From each crystal, four data sets of 360° (rotation increment 0.1°, exposure 0.1 s) were collected with different κ and φ angles at a wavelength of 2.7552 Å (energy 4.5 keV). The eight data sets were each processed independently with *XDS* and then merged together with *XSCALE* (Kabsch, 2010) in space group $C222_1$. Intensities were subsequently scaled to amplitudes in *AIMLESS* (Evans & Murshudov, 2013).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	2.755
No. of crystals	2
Resolution range (Å)	64.02–2.60 (2.69–2.60)
Space group	$C222_1$
a, *b*, *c* (Å)	77.9, 112.4, 140.7
α, β, γ (°)	90, 90, 90
Total No. of reflections	1562438 (150892)
Unique reflections	19282 (1865)
Overall multiplicity	81.0 (66.1)
Completeness (%)	99.24 (98.10)
Mean $I/\sigma(I)$	41.49 (1.65)
R_{merge}	0.103 (2.760)
R_{meas}	0.104 (2.781)
$CC_{1/2}$	1 (0.75)
R_{work}/R_{free}	0.212/0.241
R.m.s.d., bond lengths (Å)	0.010
R.m.s.d., angles (°)	1.24
Ramachandran statistics	
Favoured (%)	98.5
Allowed (%)	1.5
Outliers (%)	0
Average *B* factors (Å^2)	
Protein	
Chain *A*	87.3
Chain *B*	112.4
Chain *C*	167.9
DNA	
Chain *D*	94.3
Chain *E*	94.9

2.4. Structure solution and refinement

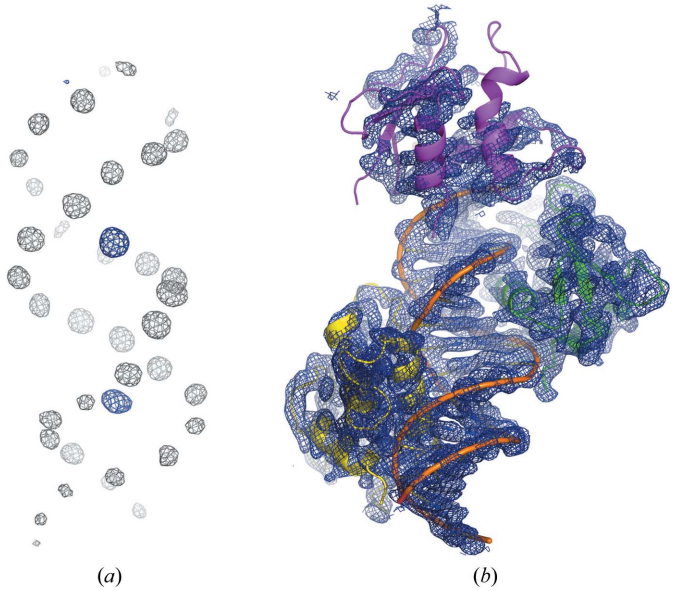
Structure solution was performed using native SAD techniques. The automatic experimental phasing pipeline *Crank*2 (Skubák & Pannu, 2013) using *PRASA* with 20 000 trials and a resolution cutoff of 3.2 Å found a substructure of 39 atoms with an occupancy of at least 25%. The pipeline provided an interpretable electron-density map and a starting model in which three IRF4 DBD molecules could be identified. The electron-density map quality and the location of the phosphorus sites allowed the manual building of the double-stranded DNA, since the pipelines is not yet able to build nucleic acids, and improvement of the IRF4 DBD molecules in *Coot* (Emsley *et al.*, 2010). Refinement was carried out with *phenix.refine* (Liebschner *et al.*, 2019) with a strategy consisting of positional, individual *B* factor, TLS and NCS refinement. The final IRF4 DBD–DNA complex structure was refined to 2.6 Å with an R_{work} and R_{free} of 21.2% and 24.1%, respectively, and was validated with *MolProbity* (Chen *et al.*, 2010). The final refined structure is composed of three molecules of IRF4 (residues 21–134, 22–130 and 19–130, respectively) and the 21 base pairs of ISRE DNA. Data-collection and refinement details are presented in Table 1.

3. Results

The human IRF4 DBD domain was expressed, purified and co-crystallized with 21-mer DNA with an AT 5′ overhang containing an ISRE element. Diffraction data were initially collected at a wavelength of 0.9795 Å on beamline I04 at Diamond Light Source. A complete data set was collected to a

resolution limit of 2.75 Å from a crystal belonging to space group $C222_1$, with unit-cell parameters a = 78.2, b = 112.5, c = 139.4 Å. The identification of the content of the crystal asymmetric unit via analysis of the Matthews coefficient was not unambiguous. The most likely oligomeric state, as suggested by previous studies (Ochiai *et al.*, 2013), is that of an IRF4 homodimer bound to one ISRE element, suggesting a molecular weight for the complex of about 44.8 kDa. The volume of the crystal asymmetric unit is compatible with the presence of either one (V_M = 3.4 Å^3 Da^{-1}, solvent content 64%) or two (V_M = 1.7 Å^3 Da^{-1}, solvent content 28%) copies of such a complex. Initial attempts to solve the structure by molecular replacement using the NMR structure of the IRF4 DBD (PDB entry 2dll; RIKEN Structural Genomics/Proteomics Initiative, unpublished work) to search for either one or two copies of the complex were unsuccessful. Automatic molecular-replacement programs such as *Phaser.MRage* (Bunkóczi *et al.*, 2013), where the asymmetric unit content can be left for the program to establish even when the number of copies of a single component are unknown, were also unsuccessful. Several reasons including conformational differences between the model and the data or inherent inaccuracies in the NMR model could account for the failure of this approach.

Taking advantage of the dedicated long-wavelength beamline I23 at Diamond Light Source, data were collected at a wavelength of 2.7552 Å with the aim of solving the structure of the complex *de novo* using native intrinsic phosphorus and sulfur SAD methods. The wavelength choice, guided by the experience of previous successful experiments on beamline I23, is a compromise between anomalous signal strength and absorption effects that decrease the data quality. Absorption increases with the cube of the wavelength and although in a high-vacuum environment there is no air absorption, absorption by the crystal, the sample holder and the surrounding mother liquor together can have a severe impact on the recorded intensities at long wavelengths. The impact on intensities is further exacerbated if the X-ray path length varies significantly depending on the crystal orientation. The limitations of native SAD phasing experiments that use wavelengths longer than 3 Å have previously been described (Basu *et al.*, 2019). At a wavelength of 2.7552 Å, S and P atoms contribute with anomalous differences f'' of 1.6 e and 1.3 e, respectively (Fig. 1). The final high-multiplicity (∼80) data set to a resolution of 2.6 Å (space group $C222_1$, unit-cell parameters a = 77.9, b = 112.4, c = 140.7 Å) was obtained by merging eight data sets collected from two crystals: four data sets from each crystal.



(*a*) (*b*)

Figure 2
Difference Fourier anomalous map and experimental electron-density map for the IRF4 DBD. (*a*) Difference Fourier anomalous map contoured at 5σ generated by *Crank*2 from the partially built model (no nucleic acids built). The grey electron density corresponds to P atoms from the DNA molecule, and the blue electron density, in the major DNA groove, to the S atoms from the IRF4 DBD. The weaker electron density at the very top of the picture corresponds to the S atom from the third IRF4 DBD molecule. (*b*) Experimental electron-density map generated by *Crank*2. The final model of the IRF4 DBD is fitted in the map to assess the map quality. This figure was prepared with *PyMOL* (version 2.0; Schrödinger).

PRASA, as part of the automatic structure-determination pipeline *Crank*2, was able to locate 39 atoms of the substructure within which the DNA double helix could be readily recognized (Fig. 2*a*). The substructure was used to produce an interpretable electron-density map in which, surprisingly, three IRF4 DBD molecules were identified (Fig. 2*b*). The quality of the electron-density map and the phosphorus sites were instrumental in the manual building of the DNA oligonucleotide since the *Crank*2 pipeline does not support the automatic building of nucleic acids. Iterative cycles of manual model building with *Coot* and refinement with *phenix.refine* allowed full structure determination. The final model contained three molecules of the IRF4 DBD bound to one ISRE element, which fits well in the crystallographic asymmetric unit (V_M = 2.5 Å^3 Da^{-1}, solvent content 51%) based on an estimated molecular mass of ∼61 kDa. When calculating phased anomalous difference maps with *ANODE* (Thorn & Sheldrick, 2011), the three IRF4 DBD sulfur sites gave anomalous peaks that were stronger on average than the DNA phosphorus sites (∼11σ versus ∼9σ); however, one of the sulfur sites had a much lower peak height when compared with the other two sites (∼4σ compared with ∼18σ and 13σ) (Fig. 2*a*). The corresponding IRF4 DBD molecule displays poorly defined electron density and higher B factors when compared with the other two IRF4 DBD molecules in the asymmetric unit (167 Å^2 when compared with 87 and 112 Å^2), as shown in Fig. 3.

Previous studies using electrophoretic mobility shift assays suggested that IRF4 binds the ISRE element as a homodimer with low affinity (Ochiai *et al.*, 2013). The finding of three molecules bound to ISRE is unexpected, and a full structural and biophysical analysis of binding affinities is currently under way.

4. Discussion

Although native SAD remains a challenging method for the solution of nucleic acid crystal structures (Harp *et al.*, 2016),

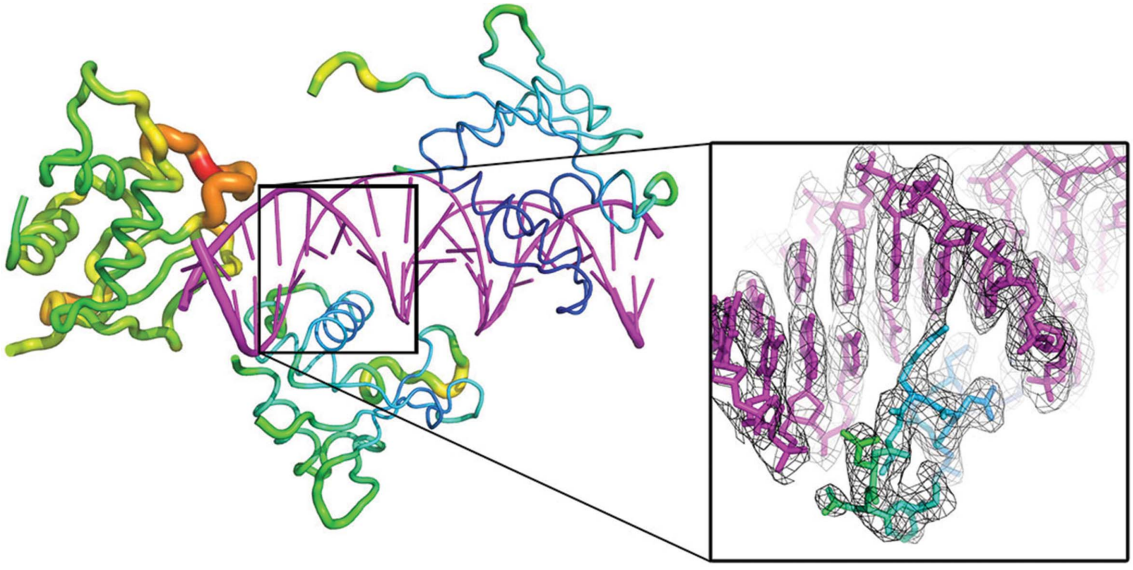


Figure 3
Crystal structure of the IRF4 DBD on ISRE DNA. The IRF4 DBD is in a *B*-factor putty cartoon representation, where the cartoon thickness and colour reflect the relative C^{α} *B* factors within the molecule. The ISRE DNA is coloured magenta. The σ_A-weighted $2F_o - F_c$ refined map is shown (grey mesh) at a contour level of 1.5σ. The map, focusing on one of the recognition helices, was carved around the atomic model of the IRF4 DBD with a border of 2 Å to improve clarity. This figure was prepared with *PyMOL* (version 2.0; Schrödinger).

this is not the case for protein–nucleic acid complexes. Native intrinsic phosphorus and sulfur SAD was chosen as a fast and elegant method for the determination of the IRF4 DBD–DNA complex structure. This technique does not rely on selenomethionine substitution or heavy-atom derivatization, but instead measures the anomalous signal from light atoms that are naturally present in proteins and nucleic acids. As the error associated with the measurement decreases with the square of the number of observations, a high-multiplicity data set was obtained by collecting and merging eight diffraction data sets collected from two different crystals. The individual data sets were collected at different κ and φ angles to minimize systematic error due to the experimental setup. As the I23 beam is unfocused, the beam flux is reduced, allowing multiple sweeps of 360° of data to be collected at low dose using the settings described in Section 2. Data collection is brought to an end when signs of radiation damage are detected, either via a decrease in the anomalous signal resolution or a decrease in the number of reflections recorded during data collection, by using *Diffraction Image Screening Tool and Library* (*DISTL*) software plots (Zhang *et al.*, 2006). The merging of the data sets increases the Bijvoet multiplicity at the same time as limiting the radiation damage. The higher redundancy increases the accuracy of the data and the strength of the anomalous signal to noise of the data set (Liu *et al.*, 2012).

Of the 45 anomalous scatterers in the asymmetric unit (42 P atoms in the double-stranded DNA and one S atom per IRF4 DBD molecule), 39 could be initially identified by *PRASA*, providing sufficient anomalous signal to phase the whole complex. Of the three S atoms, however, one produced a very weak anomalous signal when compared with the other two. This sulfur belongs to an IRF4 DBD molecule with poorly defined electron density and higher *B* factors. With only two strong anomalous sulfur sites, it could be argued that the structure of this specific complex could have been solved using the phosphorus substructure alone. A further advantage of solving the phosphorus substructure was that the DNA double helix was readily recognisable and the electron-density map for the nucleic acid portion of the structure was strong. The phosphorus sites were used as well defined guides for fitting and building the DNA double-helix model, which is important when, as in this case, the DNA departs from the canonical B form (Fig. 3).

Native intrinsic SAD phasing is particular helpful when homologous models for molecular replacement are not available, when molecular replacement is not successful and/or when the initial electron-density maps are not suitable for model building. At the time of this study, only an NMR model of the IRF4 DBD domain was available as a molecular-replacement model and it did not lead to a clear phasing solution. The molecular-replacement procedure was confounded by the unexpected oligomerization state of the complex: a heterotetramer with three IRF4 DBD molecules bound to one ISRE element. Furthermore, one of the three IRF4 DBD molecules in the asymmetric units displayed poor electron density and high *B* factors (Fig. 3), which might also explain the difficulty in solving the structure of the complex by molecular replacement.

Despite the challenges associated with the technique, native SAD phasing is on the brink of becoming the routine method of choice for *de novo* structure determination (Rose *et al.*, 2015). The availability of dedicated long-wavelength beamlines to increase the anomalous scattering signal of intrinsic light atoms has been instrumental in the increasing popularity of the method. Protein–DNA complexes are especially good candidates for native SAD phasing at long wavelengths since

the technique is particularly suited for sulfur and phosphorus substructure detection. To conclude, this work suggests that by using long-wavelengths beamlines, such as I23 at Diamond Light Source, this method could be generally applicable to a large number of nucleic acid–protein complexes.

Acknowledgements

We would like to thank Diamond Light Source for beamtime allocation and access. Author contributions were as follows: AA, KEO and RD performed experiments; KEO solved the crystal structures; AA, KEO, AW and EJM wrote the manuscript; EJM supervised the overall study.

Funding information

This research was supported by a Wellcome Trust Institutional Strategic Support Fund award (204833/Z/16/Z). For the purpose of open access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

References

Basu, S., Olieric, V., Leonarski, F., Matsugaki, N., Kawano, Y., Takashi, T., Huang, C.-Y., Yamada, Y., Vera, L., Olieric, N., Basquin, J., Wojdyla, J. A., Bunk, O., Diederichs, K., Yamamoto, M. & Wang, M. (2019). *IUCrJ*, **6**, 373–386.

Bunkóczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D. & Read, R. J. (2013). *Acta Cryst.* D**69**, 2276–2286.

Bushweller, J. H. (2019). *Nat. Rev. Cancer*, **19**, 611–624.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

Dyson, H. J. & Komives, E. A. (2012). *IUBMB Life*, **64**, 499–505.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Fuxreiter, M., Simon, I. & Bondos, S. (2011). *Trends Biochem. Sci.* **36**, 415–423.

Harp, J. M., Pallan, P. S. & Egli, M. (2016). *Crystals*, **6**, 125.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature*, **290**, 107–113.

Hudson, W. H. & Ortlund, E. A. (2014). *Nat. Rev. Mol. Cell Biol.* **15**, 749–760.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Lee, T. I. & Young, R. A. (2013). *Cell*, **152**, 1237–1251.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.

Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.

Liu, Z.-J., Vysotski, E. S., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.

Luo, Z., Dauter, M. & Dauter, Z. (2014). *Acta Cryst.* D**70**, 1790–1800.

Munshi, S., Gopi, S., Asampille, G., Subramanian, S., Campos, L. A., Atreya, H. S. & Naganathan, A. N. (2018). *Nucleic Acids Res.* **46**, 8700–8709.

Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., Chong, A. S., Klein, U., Dinner, A. R., Singh, H. & Sciammas, R. (2013). *Immunity*, **38**, 918–929.

Raiber, E. A., Murat, P., Chirgadze, D. Y., Beraldi, D., Luisi, B. F. & Balasubramanian, S. (2015). *Nat. Struct. Mol. Biol.* **22**, 44–49.

Rose, J. P. & Wang, B.-C. (2016). *Arch. Biochem. Biophys.* **602**, 80–94.

Rose, J. P., Wang, B.-C. & Weiss, M. S. (2015). *IUCrJ*, **2**, 431–440.

Schuermann, J. P. & Tanner, J. J. (2003). *Acta Cryst.* D**59**, 1731–1736.

Skubák, P. & Pannu, N. S. (2013). *Nat. Commun.* **4**, 2777.

Tateishi-Karimata, H. & Sugimoto, N. (2020). *Chem. Commun.* **56**, 2379–2390.

Terwilliger, T. C., Bunkóczi, G., Hung, L.-W., Zwart, P. H., Smith, J. L., Akey, D. L. & Adams, P. D. (2016). *Acta Cryst.* D**72**, 346–358.

Thorn, A. & Sheldrick, G. M. (2011). *J. Appl. Cryst.* **44**, 1285–1287.

Varadi, M., Zsolyomi, F., Guharoy, M. & Tompa, P. (2015). *PLoS One*, **10**, e0139731.

Wagner, A., Duman, R., Henderson, K. & Mykhaylyk, V. (2016). *Acta Cryst.* D**72**, 430–439.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Zhang, Z., Sauter, N. K., van den Bedem, H., Snell, G. & Deacon, A. M. (2006). *J. Appl. Cryst.* **39**, 112–119.

5. Chapter 5. Investigating the binding mechanism of interferon regulatory factor 4 to DNA in the context of Multiple Myeloma

Investigating the binding mechanism of interferon regulatory factor 4 to DNA in the context of Multiple Myeloma

Alessandro Agnarelli[1], Kamel El Omari[2,3], Aaron Alt[1], Leanne Milton-Harris[1], Daniel Adrian Epuran[1], David Wood[1], Timothy Chevassut[5], Michelle J West[1] and Erika J Mancini[1]*

[1]School of Life Sciences, University of Sussex, Falmer, Brighton, United Kingdom,

[2]Diamond Light Source, Harwell Science and Innovation Campus, Didcot, United Kingdom, [3]Research Complex at Harwell, Rutherford Appleton Laboratory, United Kingdom

[5]Brighton and Sussex Medical School, University of Sussex, Falmer, Brighton, United Kingdom,

*Corresponding author

E-mail: erika.mancini@sussex.ac.uk (EJM)

Abstract

Interferon regulatory factor 4 (IRF4) is a transcription factor that plays essential roles in the regulation of immune cells, including B and T cells. IRF4 has also emerged to have a vital role in the incurable haematological cancer multiple myeloma (MM) and previous studies have reported that IRF4 might bind to interferon sequence response elements (ISREs) DNA sequence to act as transcription factor in MM. We solved the crystal structure of IRF4 DNA binding domain (DBD) together with different ISRE motifs. We showed that IRF DBD bound to ISRE motifs because it recognizes a specific DNA shape with no evidences of protein-protein interactions. We also show that IRF4 contacts both consensus and non-consensus sequences. Additionally, we report that IRF4 affinity to different ISREs motifs is very poor. Particularly, IRF4 affinity to different ISRE motifs is influenced both by the sequences upstream of ISRE core sequences but also by the spacing of the nucleotides between ISRE sequences. Together, these data provide detailed insights into the interactions of IRF4 with ISRE motifs which will be key to future structure-guided drug discovery of inhibitors of IRF4 in MM.

1. Introduction

Interferon regulatory factor 4 (IRF4) is a transcription factor that plays a crucial role in the immune system. IRF4 regulates multiple stages of B cell development and it is essential for plasma cell differentiation [1] [2]. According to its protein level, IRF4 can form heterodimers or homodimers to specific DNA sequences activating the expression of genes related to germinal center (GC) B cell or plasma cell differentiation. In particular, an increase of IRF4 protein expression determines the homodimerization of IRF4 to interferon sequence response elements ISREs (GAAANNGAAA) DNA element activating the expression of genes responsible for the differentiation of B cell into plasma cells [3]. On the other hand, IRF4 binds as a heterodimer to Ets-IRF composite elements EICEs (GGAANN(N)GAAA) with PU.1 or AP-1-IRF composite elements AICEs (GAAATGAGTCA or GAAANNNNTGAGTCA) with proteins of the AP-1 family such as BATF in order to activate the expressions of genes related to GC B cell [4] [5]. IRF4 is also implicated in haematological malignancies. In particular, IRF4 overexpression has been shown to have a pivotal role in multiple myeloma (MM), an incurable malignancy of plasma cells [6]. IRF4 promotes MM cell viability by inhibiting the expression of pro-apoptotic genes *Bcl2 Modifying Factor* (*BMF*) and *BCL2L11* [7]. Moreover, knockdown experiments of IRF4 have shown that this transcription factor is essential for the viability of MM cells [6]. In MM cells IRF4 directly activates *MYC* expression. Conversely, MYC transactivates *IRF4* expression creating a positive autoregulatory feedback loop that results in an aberrant proliferation of MM cells and IRF4 overexpression [8] [6]. A ChIP-seq study on NCI-H929 MM cell lines showed that IRF4 mainly binds to ISRE motifs (GAAANNGAAA or GAAA) in order to activate the expression of genes [9]. Since IRF4 is overexpressed in MM, its interaction with DNA could occur through homodimerization to ISRE motifs. At present there are no X-ray crystal structures showing the interaction of IRF4 full length protein with ISRE DNA motifs. Compared to other IRF proteins, IRF4 is characterized by low affinity to DNA [10]. IRF4 affinity to DNA has been shown to

increase during protein-protein interaction. In particular, IRF4 interaction with Ets and AP-1 proteins results in an higher IRF4 affinity to EICEs and AICEs motifs [3] [4]. On the other hand, IRF4 binding to ISRE motifs is characterized by lower affinity, suggesting that IRF4 binds ISRE sequences when expressed at high amounts [11] [3] [12]. IRF4 consists of a highly conserved N-terminal DNA binding domain (DBD) connected to C-terminal interferon activation domain (IAD), critical in mediating protein-protein interactions via a linker domain (LKD) [10]. IRF4 DBD is characterized by tryptophan pentad repeat allowing it to form helix-loop-helix motif that facilitates DNA binding [13]. The low IRF4 binding affinity to DNA has been attributed to an auto-inhibitory region (AR) located in the last 30 residues of the IAD domain [14] [4]. Structural studies have shown that the AR is a flexible unstructured peptide that prevents the DBD from binding to DNA [10]. Moreover, IRF4-DBD interactions with transcription factors would release AR inhibition [4]. This hypothesis does not illustrate how release of the inhibition would occur when IRF4 binds as homodimers to ISRE sequences. In Chapter 4, we showed that three IRF4 DBD molecules bind an ISRE DNA motif. This result is unexpected as previous electrophoretic mobility shift assays suggested that IRF4 binds the ISRE element as a homodimer with low affinity [3]. In order to obtain further insights into the stoichiometry of this interaction we performed a full structural and biophysical analysis of IRF4 in complex with a variety of different ISRE motifs. In particular, in order to understand how IRF4 binds to ISRE motifs, we have co-crystallized IRF4 DBD with different ISRE sequences. In particular, the ISRE motifs analysed include two canonical ISRE motifs that differ from each other in the sequence upstream of GAAA motifs and conserved noncoding sequence 9 (CNS-9) region of *Prdm1* (encoding Blimp1) [3] [2]. Our study shows that the spacing of two nucleotides between two consensus IRF (GAAA) recognition sequences is essential to increase IRF4 binding affinity to ISRE motifs. Furthermore, as already shown by Sundararaj (2021) no intermolecular interactions were observed between the interacting DNA-binding domains. Therefore, the IRF4 binding causes a conformational change to DNA that allows IRF4 to bind ISRE motifs. In our study we also show

that IRF4 affinity to ISRE motifs is higher when compared to single consensus sequence GAAA or double consensus sequences separated by more than 2 nucleotides. These data confirm that the spacing of 2 nucleotides between GAAA consensus sequences plays a fundamental role for the IRF4-DBD binding to ISRE motifs.

2. Materials and Methods

2.1. Protein Expression and Purification

IRF4 DBD (a.a 20-139) was expressed and purified as described previously [15]. IRF4 DBD were pooled and concentrated to 10 mg ml^{-1}. Oligonucleotides containing an interferon response element (ISRE) 5'-ATAACTGAAACCGAAAGTAC-3', 5'-TGTACTTTCGGTTTCAGTTA-3' (canonical ISRE 2), 5'-AATAAAAGAAACCGAAAGTAA-3', 5'-TTTACTTTCGGTTTCTTTTAT-3'(canonical ISRE 1) and 5'- TCAACTGAAACCGAGAAAGC-3', 5'- AGCTTTCTCGGTTTCAGTTG-3' (CNS-9 region) (Eurofins Genomic) were annealed and incubated in a 1.2:1 molar ratio for 1 h at 277K prior to crystallisation.

2.2. Crystallization

The IRF4 DBD–ISRE complexes were used to screen 384 conditions using the sitting-drop vapour diffusion method. Initial hits appeared within a week. Regarding IRF4 DBD- canonical ISRE 2 complex, the best crystals grew in 0.1M Bis Tris pH 5.5, 0.2M ammonium acetate, 25% w/v PEG 3350 at 293K. IRF4 DBD-CNS-9 region complex was crystallized in 0.1M Tris HCl pH 8.5, 30% w/v PEG4000, 0.2M lithium sulfate. IRF4 DBD-canonical ISRE 1 was crystallised as described previously [15]. Crystals were successfully cryoprotected in 25% glycerol by flash cooling in liquid nitrogen.

2.3. Data collection and processing

Diffraction data from IRF4 DBD-canonical ISRE 2 complex were collected on DIAMOND BEAMLINE I04-1 at the Diamond Light Source Ltd (Didcot, Oxfordshire). These data were indexed, scaled and merged with xia2 dials. The space group was determined to be P 21 21 21 with unit cell dimensions a= 64.32, b= 66.05, c= 201.89 Å, and α,β,γ = 90°C. Diffraction data from IRF4 DBD-CNS-9 complex were collected on DIAMOND BEAMLINE I24 at the Diamond Light Source Ltd (Didcot, Oxfordshire). These data were indexed, scaled and merged with xia2 3dii [16] [17] [18] [19] [20]. The space group was determined to be P 21 21 21 with unit cell dimension a= 64.78, b= 83.25, c= 88.53 Å, and *α,β,γ = 90*°C. Diffraction data from IRF4 DBD-canonical ISRE 1 complex were collected as described previously [15].

2.4. Structure Solution and Refinement

Structure solutions were performed by molecular replacement using the PHASER- Expert Mode Molecular Replacement software in CCP4i2 [21]. IRF4 DBD-canonical ISRE 2 structure was solved using as reference model the structure of IRF4 DBD-CNS-9 complex. IRF4 DBD- DNA CNS-9 structure was solved using as reference model the structure of IRF4-PU1-EICE complex [22]. IRF4 DBD-canonical ISRE 1 structure was solved as described previously [15]. Iterative model building was performed with the program Autobuild [23]. Subsequent refinement cycles were performed first with REFMAC5 [24] and then PHENIX REFINE [25], taking care of keeping the same R_{free} test set in both programs. Toward the end of refinement, TLS (Translation/Libration/Screw) vibrational motion refinement was used and water molecules added giving a final R_{work}/R_{free} of 20/23% for IRF4 DBD-CNS-9 and 21/24% for IRF4 DBD-canonical ISRE 2. Data collection and refinement details are listed in Table 1 and Table 2, respectively. For structure details of IRF4 DBD-canonical ISRE 1 see chapter 4 and [15].

2.5. Microscale Thermophoresis (MST)

The MST assays were performed using the Monolith NT.115 machine. Different ISRE DNAs were labelled at 5'-ends using the dye AF647 (Eurofins Genomic). Serial dilutions of IRF4 DBD protein in MST buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20) were mixed with 20 nM of 5'-AF647 ISRE DNA and loaded into standard glass capillaries (Monolith NT.115 Capillaries, NanoTemper Technologies). Thermophoresis analysis was performed on a Monolith NT.115 instrument (20% LED, 40/60% MST power) at 22°C. The MST curves were fitted using NT Analysis software (NanoTemper Technologies) to obtain EC50 and Hill coefficient values for binding.

DNA used in MST assay:

Canonical ISRE 2 (artificial sequence): 5'-TAACTGAAACCGAAAGTAC-3', 5'-GTACTTTCGGTTTCAGTTA-3'.

CNS-9 region: 5'- CAACTGAAACCGAGAAAGC-3', 5'- GCTTTCTCGGTTTCAGTTG-3'

Canonical ISRE 1 (artificial sequence): 5'-ATAAAAGAAACCGAAAGTAA-3', 5'-TTACTTTCGGTTTCTTTTAT-3'

MYC promoter sequence 2: 5'- TGAAACCTGGCTGAGAAAT-3', 5'-ATTTCTCAGCCAGGTTTCA-3'

MYC promoter sequence 1: 5'- TCTCTTCTGAAACCTGGCT-3', 5'-AGCCAGGTTTCAGAAGAGA-3'

2.6. Electrophoretic Mobility Shift Assay (EMSA)

8% native acrylamide EMSA gels were allowed to polymerise for 2 hours before being pre-run for 1 hour at 120 V at 4°C in EMSA running buffer (0.5x Tris-Borate-EDTA (TBE). EMSA samples were prepared using purified protein and 5'-AF647 fluorescently labelled ISRE DNA (Eurofins

Genomic) in EMSA buffer (20 mM HEPES (pH 7.5), 5 mM $MgCl_2$, 150 mM NaCl). Each EMSA sample contained a final concentration of 100 nM ISRE DNA mixed with various protein concentrations (for CNS-9 and ISRE2: 0.1, 0.3, 0.6, 1, 6, 20, 40, 100μM IRF4 DBD final concentrations; ISRE1, MYC promoter sequence 1 and 2: 1.5, 3, 5, 30, 100, 200, 500, 700 μM and 1mM IRF4 DBD final concentrations). Samples were incubated on ice for 1 hour to allow formation of protein-DNA complexes before being loaded onto the gels. Control samples were made using the non-specific competitor (500 ng Salmon Sperm DNA (Invitrogen)) or the specific competitor (40 μM unlabelled ISRE DNA) mixed with the highest protein concentration. Samples were left on ice for 1 hour before the addition of the fluorescent probe (labelled ISRE DNA) followed by another hour on ice. All samples were loaded into the polymerised EMSA gel alongside an EMSA reference dye (40% sucrose, 0.2% Orange G, 50 mM HEPES pH 7.5). The gel was run again for 1 hour at 120 V at 4°C. The gel was visualised using an Odyssey Fc (LI-COR) machine (2 min exposure at wavelength 700 nm). EMSA quantification was performed using one site-specific binding with Hill slope by GraphPad Prism Version 6.01.

DNA used in EMSAs:

Canonical ISRE 2 (artificial sequence): 5'-TAACTGAAACCGAAAGTAC-3', 5'-GTACTTTCGGTTTCAGTTA-3'.

CNS-9 region: 5'- CAACTGAAACCGAGAAAGC-3', 5'- GCTTTCTCGGTTTCAGTTG-3'

Canonical ISRE 1 (artificial sequence): 5'-ATAAAAGAAACCGAAAGTAA-3', 5'-TTACTTTCGGTTTCTTTTAT-3'

MYC promoter sequence 2: 5'- TGAAACCTGGCTGAGAAAT-3', 5'-ATTTCTCAGCCAGGTTTCA-3'

MYC promoter sequence 1: 5’- TCTCTTCTGAAACCTGGCT-3’, 5’- AGCCAGGTTTCAGAAGAGA-3’

3. Results

3.1. Analysis of IRF4-DNA interactions

The IRF4 DNA-binding domain predominantly interacts with the DNA through a series of phosphate contacts resulting in the position of the recognition helix (α3) in the major groove and loop L1 in the minor groove (Fig. 1a, 1b, 1c). In particular, contacts with both ISRE core sequence GAAA of canonical ISRE 1 sequence (Fig. 1a), canonical ISRE 2 sequence (Fig. 1b) and CNS-9 region (Fig. 1c) are mediated by the C-terminal region of the α3 helix. Regarding the B chain of IRF4 DBD, Arg 98, Cys 99, Asn 102 and Lys 103 mainly mediated the interactions with GAAA sequence (Fig. 1b, 2b, 3b, 4). Specifically, for DNA CNS-9 region Arg 98 forms a hydrogen bond with the OP2 of the first base whereas in the canonical ISRE 1 and ISRE 2 sequences Arg 98 mediates direct contact with the guanine through an hydrogen bond (Fig. 1b, 2b, 3b, 4). The first adenine is then recognized by Cys 99 that mediates a hydrogen bond with the N6 of the adenine base (Fig. 1b, 2b, 3b, 4). Asn 102 interacts with the OP2 of the first base through an hydrogen bond (Fig. 1b, 2b, 3b, 4). In the DNA CNS-9 region and canonical ISRE 1 sequence, Lys 103 interacts with the last adenine by forming a van der Waals contact whereas for the canonical ISRE 2 sequence, Lys 103 forms a hydrogen bond with the N6 atom of the adenine (Fig. 1b, 2b, 3b, 4). Moreover, in the canonical ISRE 2 sequence Lys 103 mediates an interaction with the second adenine by forming a hydrogen bond with the N6 atom of the base (Fig. 2b, 4b). Similarly to the

chain B of IRF4 DBD, also in the chain A Arg 98, Cys 99, Asn 102 and Lys 103 interact with GAAA core sequence of canonical ISRE 1 and ISRE 2 sequences (Fig. 1c, 2c, 4a, 4b). In particular, Arg 98 interacts directly with the first base through a hydrogen bond with the N7 of the guanine and Asn 102 forms a hydrogen bond with the OP2 of the first base (Fig. 1c, 2c, 4a, 4b). Cys 99 mediates the interaction with the first adenine by forming a hydrogen bond with the N6 of the base(Fig. 1c, 2c, 4a, 4b). Recognition of the second adenine is performed by Lys 103 through a hydrogen bond with the N7 of the base (Fig. 1c, 2c, 4a, 4b). Moreover, only in the canonical ISRE 2 sequences Lys 103 interacts with the last adenine by forming a hydrogen bond with N6 of the adenine(Fig. 2c, 4b). Regarding the DNA CNS-9 region, Lys 103 of the chain A is the only residue that mediates contact with the GAAA core sequence by forming a hydrogen bond with the N7 atom of the guanine (Fig. 3c, 4c). Lys 103 interacts also with the first adenine by van der Waals contact (Fig. 3c, 4c). Moreover, in the DNA CNS-9 region, the first two bases upstream of the recognition sequence (GAGAAA) interacts with residues of IRF4 DBD that are the same interacting with the GAAA core sequences (Fig. 3c, 4c). In particular, Arg 98 interacts with the guanine by forming an hydrogen bond with the N7 atom (Fig. 3c, 4c). Asn 102 forms a hydrogen bond with the OP2 of the guanine (Fig. 3c). The adenine is recognized by Cys 99 through a hydrogen bond with the N6 of the base (Fig. 3c, 4c). Lys 103 also contacts the adenine through a hydrogen bond with C8 of the base (Fig. 3c, 4c).

3.2. ISRE DNA conformational changes upon IRF4 binding

Binding of the IRF4 DBD domain induced a significant bend of the ISRE DNA duplex. A quantitative analysis of the DNA conformational parameters was performed using the software Curves+ [26]. As shown in table 3, IRF4 DBD binding to CNS-9 region distorted the DNA backbone by 19.1° that is larger than that observed for the PU.1/IRF4/DNA heterodimer (15.6°). Moreover, IRF4 DBD binding to Canonical ISRE 1 and Canonical ISRE 2 induced a DNA

distortion of 5.3° and 8.4°, respectively. The homodimer IRF4 DBD-CNS-9 structure showed a mean axial rise per turn of 3.30 Å that was comparable to that of PU.1/IRF4/DNA heterodimer (3.23 Å), IRF4 DBD-Canonical ISRE 1 (3.32 Å) and IRF4 DBD-Canonical ISRE 2 (3.26 Å). The mean propeller twist and the helical twist per base pair were also comparable among the different structures as shown in table 4. Moreover, the base pair tilt of IRF4 DBD-CSN-9 structure was 0.4 Å whereas IRF4 DBD-Canonical ISRE 1 and IRF4 DBD-Canonical ISRE 2 showed a tilt of 0.1 Å and -0.6 Å, respectively.

3.3. IRF4 DBD is characterized by a different affinity to different ISREs motifs

To measure the IRF4 DBD affinity to different ISREs DNA sequences we performed EMSA and MST assays. IRF4 DBD showed very low affinity to the canonical ISRE 1 sequence (K_d= 18.26 μM) (Fig. 5a). In order to test if the overhang sequences could have an impact on IRF4 DBD affinity to canonical ISRE 1 sequence, we tested IRF4 DBD affinity to the canonical ISRE 2 sequence that differs from canonical ISRE 1 sequence only for the overhang sequences. IRF4 DBD binds to the canonical ISRE 2 sequence with a K_d= 0.6 μM, showing an increase affinity of about 30 times compared to the canonical ISRE 1 sequence (Fig. 5b). We then measured the IRF4 DBD affinity to *bona fide* ISREs sequences. Previous studies showed that IRF4 binds the conserved noncoding sequences 9 (CNS-9) region of *Prdm1* (encoding Blimp-1) and 5'-CAACTGAAACCGAGAAAGC-3' is one of the over-represented target sequences [3] [2]. Our results showed that IRF4 DBD bound to 5'- CAACTGAAACCGAGAAAGC-3' ISRE sequence with a K_d= 3.95 μM (Fig. 5c). This affinity was similar to the canonical ISRE 2 sequence and raised the question if IRF4 DBD interacted with GAAANNGAGAAA or GAAANNNNGAAA. In order to answer the question, we solved the crystal structure of the IRF4 DBD bound to the CNS-9 region and we analysed the specific IRF4-DNA interactions. In order to determine the IRF4 DBD affinity to single ISRE motifs, we analysed its affinity to 2 specific *MYC* promoter sequences obtained from IRF4 ChIP-seq data of H929 MM cell lines [9]. IRF4 DBD is

characterized by a low affinity both to *MYC* promoter sequences 1 and *MYC* promoter sequences 2 with a K_d= 612.7 μM and 521 μM, respectively (Fig. 6a, 6b). In order to have a more quantitative analysis of IRF4 DBD affinity to ISRE motifs, we performed MST assay. MST results confirmed the EMSA data, showing that IRF4 DBD displayed the highest affinity for the canonical ISRE 2 sequence with a K_d= 2.66 μM (Fig. 7). On the contrary, the affinity for *MYC* promoter sequences 1 and *MYC* promoter sequences 2 was low with (K_d= 2.50mM and 1.26 mM respectively) (Fig. 7). The IRF4 DBD affinity for CNS-9 region and canonical ISRE 1 sequence by MST were K_d= 5.32 μM and 122 μM, respectively (Fig. 7).

4. Discussion

IRF4 is a lymphoid transcription factor that regulates the expressions of genes both as homodimers or heterodimer interacting with other DNA-binding proteins [3] [22]. Previous studies have suggested that high concentrations of this transcription factor induces homodimerisation onto ISRE motifs in order to activate the expression of specific target genes related to plasma cell differentiation [3]. IRF4 plays a vital role in the incurable haematological cancer MM where it is overexpressed and possibly interacting with ISRE motifs [6] [9]. Our study provides a structural view of IRF4 DBD interacting with different ISRE motifs. The IRF4 DNA homodimeric complexes are exclusively through protein-DNA contacts and there is no evidence of protein-protein interaction. The IRF4 co-operativity in binding to ISRE motifs is mainly caused by the allosteric effects transmitted through the DNA with no contribution from the interacting DBDs. These data confirmed the IRF4 DBD crystal structure solved by Sundararaj *et al.* [12] and highlighted the differences with the IRF4-PU.1-EICE heterodimeric complex where there was the presence of protein-protein interaction [22]. IRF4-DBD-canonical ISRE 1 structure showed three molecules interacting with the DNA. In particular, the third IRF DBD molecule interacts with non-specific ISRE sequences. The presence of this molecule was

because of the overhang sequence used that created a spurious ISRE binding site. In addition, in the IRF4 DBD-DNA CNS-9 region structure, the highly conserved residues (Arg 98, Cys 99 and Asn 102) interact with both consensus and non-consensus DNA sequence elements. These data confirm that IRF transcription factors are highly versatile in binding to their target DNA. Indeed, the IRF7 DBD in the crystal structure of the IRF-3/IRF-7/NFkB complex bound to the positive regulatory domains (PRDs) of the IFN-β enhancer showed that the conserved Arg 98 interacts with bases upstream to the consensus sequence [27]. Our study confirms that IRF4 is characterized by low affinity to DNA with a K_d in the order of μM [4] [14]. In particular, our data shows that IRF4 DBD is characterized by different affinities when interacting to different ISREs motifs. IRF4 DBD displays higher affinity with ISRE canonical 2 than CNS-9 region, ISRE canonical 1 and *MYC* promoter sequences. These data suggest that the nucleotides spacer length between the ISRE motifs has an impact on the affinity of IRF4 DBD to ISRE sequences and also that IRF4 could bind *MYC* ISRE motifs probably through interactions with other proteins. A similar result has been observed with other IRF proteins. In particular, IRF3, IRF5 and IRF7 affinity was significantly decreased when interacting with ISRE motifs characterized by a space of three nucleotides between the two ISRE sequences [28]. Moreover, EMSA results showed that IRF3 binding to ISRE sequence with a space of eight nucleotides, displayed an affinity that was at least 100-fold weaker than to ISRE sequence with a space of two nucleotides [28]. The difference in affinity between canonical ISRE 1 and canonical ISRE 2 sequences suggest that also the overhang sequences upstream to GAAA motifs affect IRF4 DBD affinity to ISRE sequences. Our results show that IRF4 binding to different ISRE DNA sequences induces a significant bend in the DNA duplex. Previously Zeiske *et al.* reported that DNA binding sites characterized by conformational changes upon protein binding displayed lower affinity than binding sites with more optimal conformations prior to binding [29]. This suggests that the poor IRF4 DBD affinity to ISRE sequences could also be affected by the DNA distortion caused by IRF4 binding to DNA. Our data provide molecular and structural insights into the IRF4

interaction to ISRE motifs which will be key to towards the targeting of this transcription factor for inhibition in MM therapeutic strategies.

References

1. Klein, U., *Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination.* Nature Immunol., 2006. **7**: p. 773-782.
2. Sciammas, R., et al., *Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation.* Immunity, 2006. **25**(2): p. 225-36.
3. Ochiai, K., et al., *Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4.* Immunity, 2013. **38**(5): p. 918-29.
4. Brass, A.L., A.Q. Zhu, and H. Singh, *Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers.* Embo j, 1999. **18**(4): p. 977-91.
5. Tussiwand, R., et al., *Compensatory dendritic cell development mediated by BATF-IRF interactions.* Nature, 2012. **490**(7421): p. 502-7.
6. Shaffer, A.L., et al., *IRF4 addiction in multiple myeloma.* Nature, 2008. **454**(7201): p. 226-31.
7. Fedele, P.L., et al., *The transcription factor IRF4 represses proapoptotic BMF and BIM to licence multiple myeloma survival.* Leukemia, 2020.
8. Lopez-Girona, A., et al., *Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response.* Br J Haematol, 2011. **154**(3): p. 325-36.
9. Care, M.A., et al., *SPIB and BATF provide alternate determinants of IRF4 occupancy in diffuse large B-cell lymphoma linked to disease heterogeneity.* Nucleic Acids Res, 2014. **42**(12): p. 7591-610.
10. Remesh, S.G., V. Santosh, and C.R. Escalante, *Structural Studies of IRF4 Reveal a Flexible Autoinhibitory Region and a Compact Linker Domain.* J Biol Chem, 2015. **290**(46): p. 27779-90.
11. Krishnamoorthy, V., et al., *The IRF4 Gene Regulatory Module Functions as a Read-Write Integrator to Dynamically Coordinate T Helper Cell Fate.* Immunity, 2017. **47**(3): p. 481-497.e7.
12. Sundararaj, S., et al., *Structural determinants of the IRF4/DNA homodimeric complex.* Nucleic Acids Research, 2021. **49**(4): p. 2255-2265.
13. Escalante, C.R., et al., *Structure of IRF-1 with bound DNA reveals determinants of interferon regulation.* Nature, 1998. **391**(6662): p. 103-106.
14. Brass, A.L., et al., *Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1.* Genes Dev, 1996. **10**(18): p. 2335-47.
15. Agnarelli, A., et al., *Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4.* Acta Crystallogr F Struct Biol Commun, 2021. **77**(Pt 7): p. 202-207.
16. Winter, G., *xia2: an expert system for macromolecular crystallography data reduction.* Journal of Applied Crystallography, 2010. **43**(1): p. 186-190.
17. Winter, G., et al., *DIALS: implementation and evaluation of a new integration package.* Acta Crystallographica Section D, 2018. **74**(2): p. 85-97.
18. Kabsch, W., *XDS.* Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.
19. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D, 2011. **67**(4): p. 235-242.
20. Evans, P., *Scaling and assessment of data quality.* Acta Crystallographica Section D, 2006. **62**(1): p. 72-82.
21. Potterton, L., et al., *CCP4i2: the new graphical user interface to the CCP4 program suite.* Acta crystallographica. Section D, Structural biology, 2018. **74**(Pt 2): p. 68-84.

22. Escalante, C.R., et al., *Crystal structure of PU.1/IRF-4/DNA ternary complex.* Mol Cell, 2002. **10**(5): p. 1097-105.
23. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
24. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta crystallographica. Section D, Biological crystallography, 2011. **67**(Pt 4): p. 355-367.
25. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.
26. Lavery, R., et al., *Conformational analysis of nucleic acids revisited: Curves+.* Nucleic acids research, 2009. **37**(17): p. 5917-5929.
27. Panne, D., T. Maniatis, and S.C. Harrison, *An atomic model of the interferon-beta enhanceosome.* Cell, 2007. **129**(6): p. 1111-23.
28. Andrilenas, K.K., et al., *DNA-binding landscape of IRF3, IRF5 and IRF7 dimers: implications for dimer-specific gene regulation.* Nucleic Acids Res, 2018. **46**(5): p. 2509-2520.
29. Zeiske, T., et al., *Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites.* Cell Rep, 2018. **24**(9): p. 2221-2230.

Table 1

IRF4 DBD-CNS-9

DATA COLLECTION			**REFINEMENT**	
Source	DIAMOND BEAMLINE I24		Resolution (Å)	44.27-2.25 (2.35-2.25)
Wavelength (Å)	0.9688		Reflections used	23133 (2197)
Space Group	P 21 21 21		Reflections for R-free	1116 (95)
Unit cell a, b, c (Å) *α*,*β*,*γ* (deg)	64.78 83.25 88.53 90.00, 90.00, 90.00		Non-hydrogen atoms	2755
Resolution (Å)	44.30-2.25 (2.29-2.25)		Protein residues	223
Unique reflections	23181 (1086)		Water	76
Redundancy	6.0 (3.7)		R-work	0.20 (0.31)

Completeness (%)	99.3(95.9)		R-free	0.23 (0.30)
R-merge	0.112(1.136)		RMS deviations Bond lengths (Å)	0.009
<I/σ>	11.1(1.1)		Bond Angles (°)	1.04
Wilson B-factor (Å^2)	45.62		Average B-Factor (Å^2)	54.73
			Ramachandran plot (%) Favoured/allowed/disallowed	96.79/3.21/0.00
			MolProbity score	1.34

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

Table 2

IRF4 DBD-canonical ISRE 2

DATA COLLECTION			REFINEMENT	
Source	**DIAMOND BEAMLINE I04-1**		**Resolution (Å)**	**62.81-2.37**
Wavelength (Å)	0.9119		Reflections used	34807(2635)
Space Group	P 21 21 21		Reflections for R-free	1726(128)
Unit cell a, b, c (Å) *α*,*β*,*γ* (deg)	 64.32 66.05 201.89 90.00, 90.00, 90.00		Non-hydrogen atoms	5432
Resolution (Å)	64.35-2.37(2.41-2.37)		Protein residues	456
Unique reflections	35957(1740)		Water	76
Redundancy	6.3(5.5)		R-work	0.21(0.43)
Completeness (%)	100(100)		R-free	0.25(0.42)
R-merge	0.189(3.397)		RMS deviations Bond lengths (Å)	 0.010
<I/σ>	5.8(0.2)		Bond Angles (°)	1.25
Wilson B-factor (Å^2)	63.06		Average B-Factor (Å^2)	76.85
			Ramachandran plot (%) Favored/allowed/disallowed	 97.10/2.90/0.00
			MolProbity score	1.31

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

Table 3

	Total DNA bend °	Propel Twist °	Rise Å	Tilt Å	Helical twist °
IRF4 DBD-CNS-9	19.1	-12.5	3.30	0.4	35.2
IRF4 DBD-Canonical ISRE 1	5.30	-10.4	3.32	0.1	34.1
IRF4 DBD-Canonical ISRE 2	8.40	-13.7	3.26	-0.6	35.4
PU.1/IRF4/DNA heterodimer	15.6	-12.2	3.23	0	34.1

Figure Legends

Figure 1. **Overall structure of IRF4 DBD-ISRE complexes**. (a) IRF4 DBD interaction with canonical ISRE 1. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, IRF4 not interacting with ISRE motifs: orange. (b) IRF4 DBD interaction with canonical ISRE 2. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red. (c) IRF4 DBD interaction with CNS-9 region. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red.

Figure 2. **IRF4 DBD interactions to first ISRE motif.** (a) Cartoon representation showing IRF4 B chain recognition helix/ canonical ISRE 1 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow. (b) Cartoon representation showing IRF4 B chain recognition helix/ canonical ISRE 2 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow, carbon atoms: magenta. (c) Cartoon representation showing IRF4 B chain recognition helix/ CSN-9 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow.

Figure 3. **IRF4 DBD interactions to second ISRE motif.** (a) Cartoon representation showing IRF4 B chain recognition helix/ canonical ISRE 1 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow, carbon atoms: magenta. (b) Cartoon representation showing IRF4 B chain recognition helix/ canonical ISRE 2 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow, carbon atoms: magenta. (c) Cartoon representation showing IRF4 B chain recognition helix/ CSN-9 DNA interaction. IRF4: green, IRF4 recognition helix: blue, ISRE motifs: red, oxygen atom: red, nitrogen atom: blue, sulphur atom: yellow.

Figure 4. **An overview of IRF4 DBD-DNA interaction.** (a,b,c) Schematic diagram of IRF4 DBD- canonical ISRE 1, canonical ISRE 2 and CNS-9 DNA interaction, respectively. IRF4 chain A and chain B are coloured yellow and purple, respectively. Green lines represent bonds with the sugar and the phosphate. Red dashed lines represent bonds directly with the bases.

Figure 5. **Representative EMSA of IRF4 DBD-ISREs**. (a,b,c) EMSA analysis of IRF4 DBD-ISREs interactions showing an increase amount of IRF4 DBD binding to the ISREs probes. NS= non-specific.

Figure 6. **Representative EMSA of IRF4 DBD- *MYC* ISREs.** (a,b) EMSA analysis of IRF4 DBD-*MYC* ISREs interactions showing an increase amount of IRF4 DBD binding to the ISREs probes. NS= non-specific.

Figure 7. **Microscale thermophoresis analysis of IRF4 DBD-ISREs interactions.** Curves showing the normalised fluorescence data from MST experiments with IRF4 DBD and canonical ISRE 2 (K_d = 122 µM, green curve), CNS-9 (K_d = 5.32 µM, red curve), canonical ISRE 1 (K_d= 122 µM, cyan curve). IRF4-DBD displays the lowest affinity to *MYC* promoter sequences 1 (K_d= 2.50 mM, purple curve) and to *MYC* promoter sequences (K_d= 1.26 mM, orange). The error bars represent the standard deviation of each data point calculated from three independent experiments.

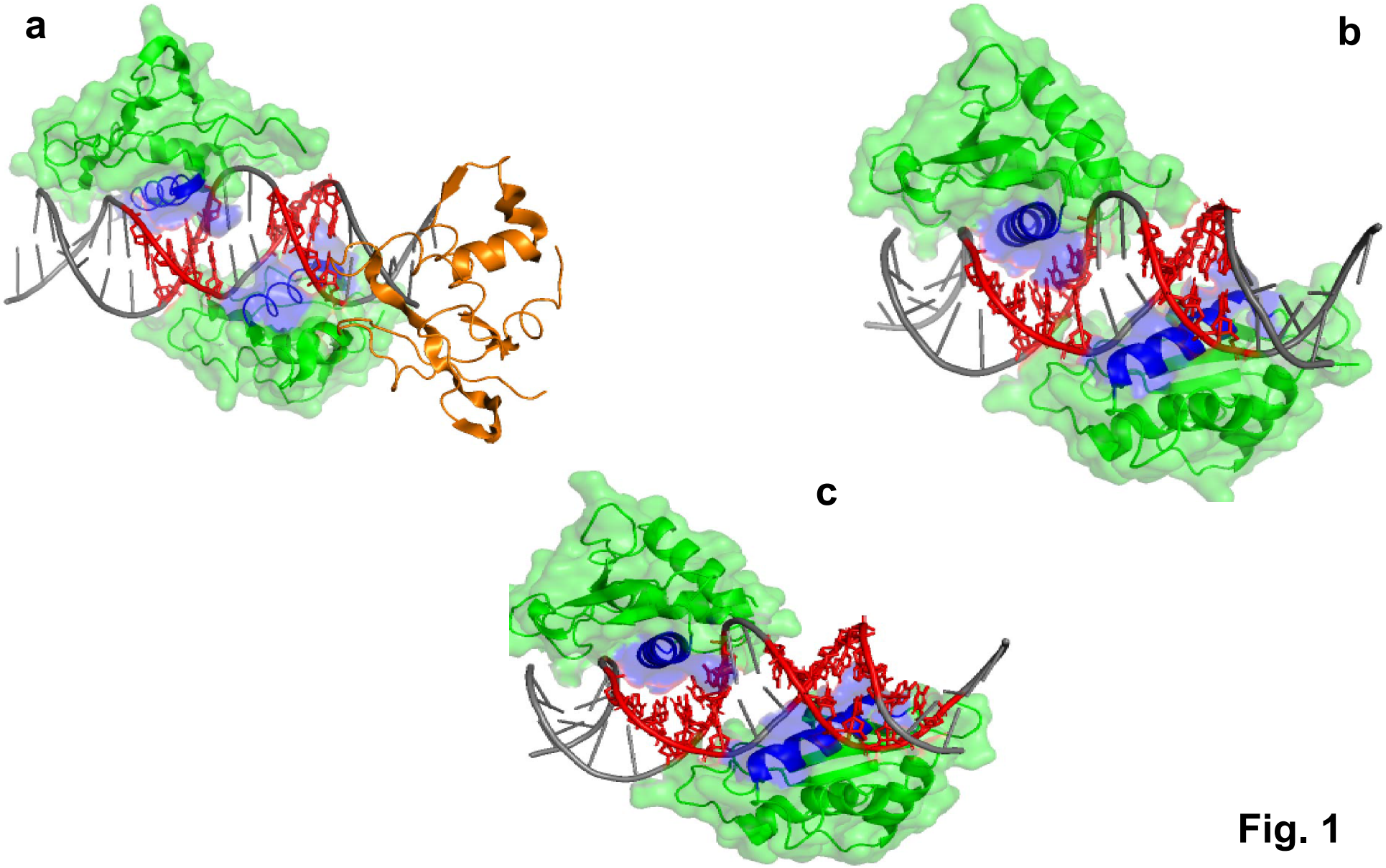


Fig. 1

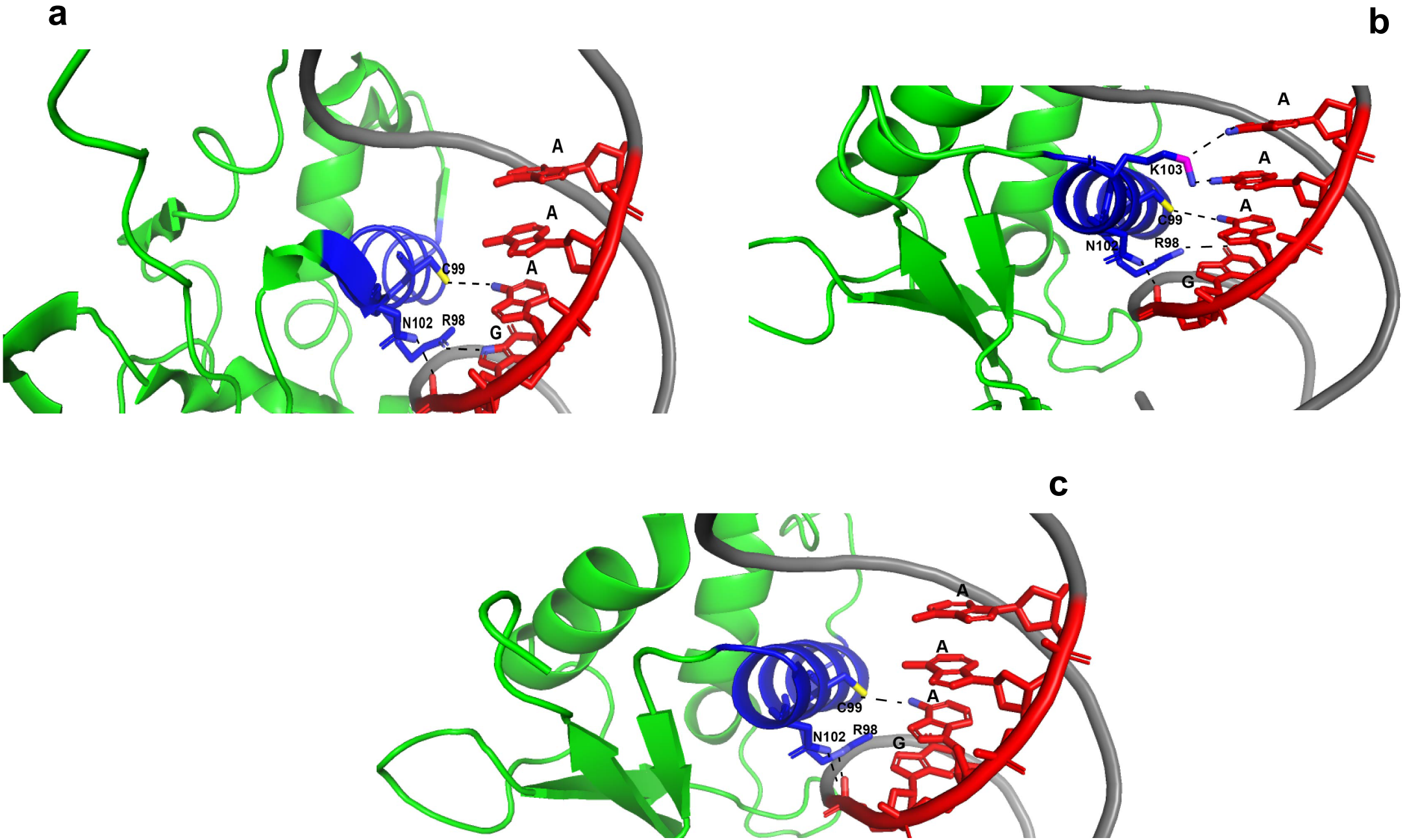


Fig. 2

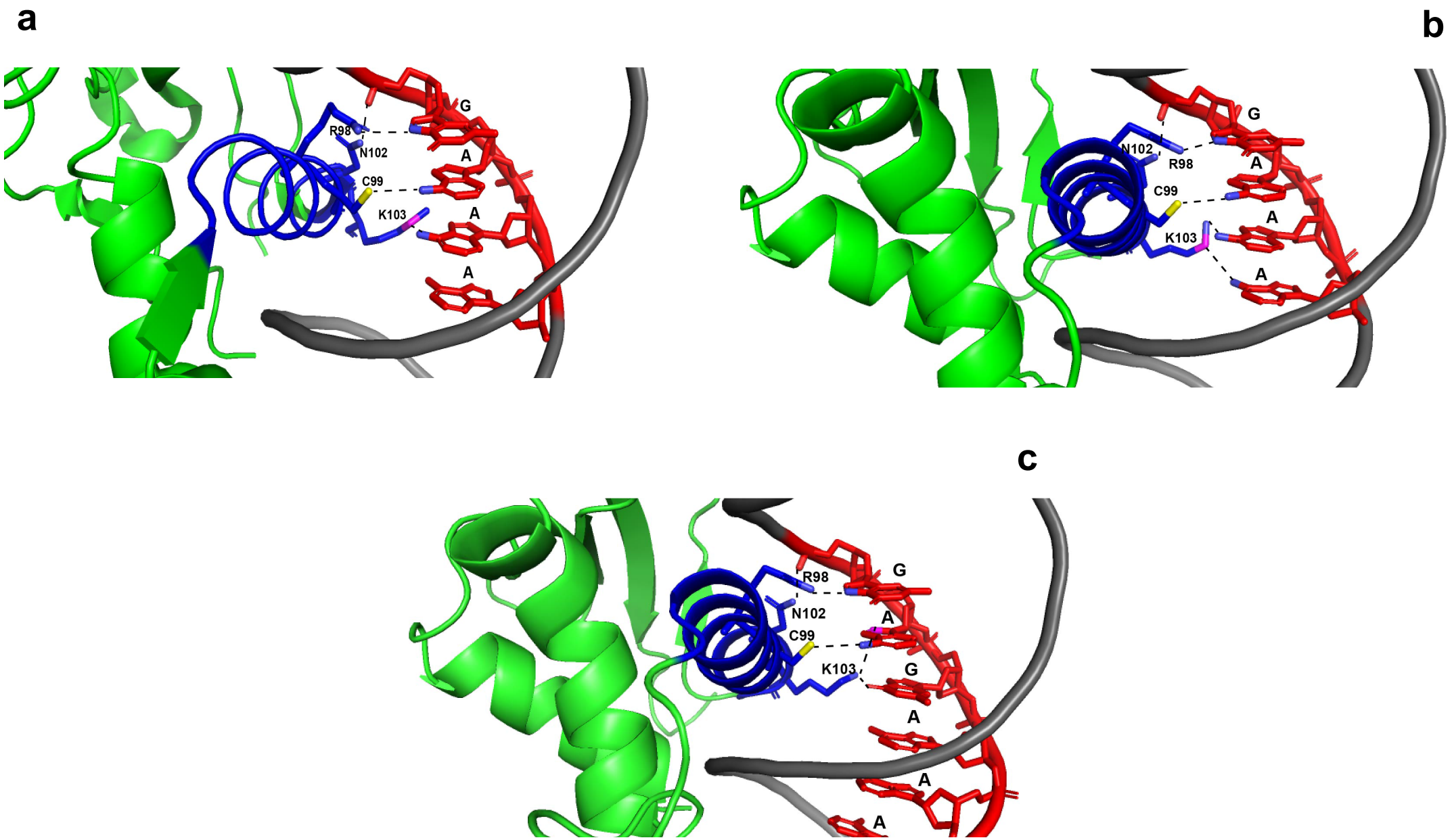

a
b
c
R98
N102
C99
K103
G
A
A
A
R98
N102
C99
K103
G
A
A
A
R98
N102
C99
K103
G
A
G
A
A
A

Fig. 3

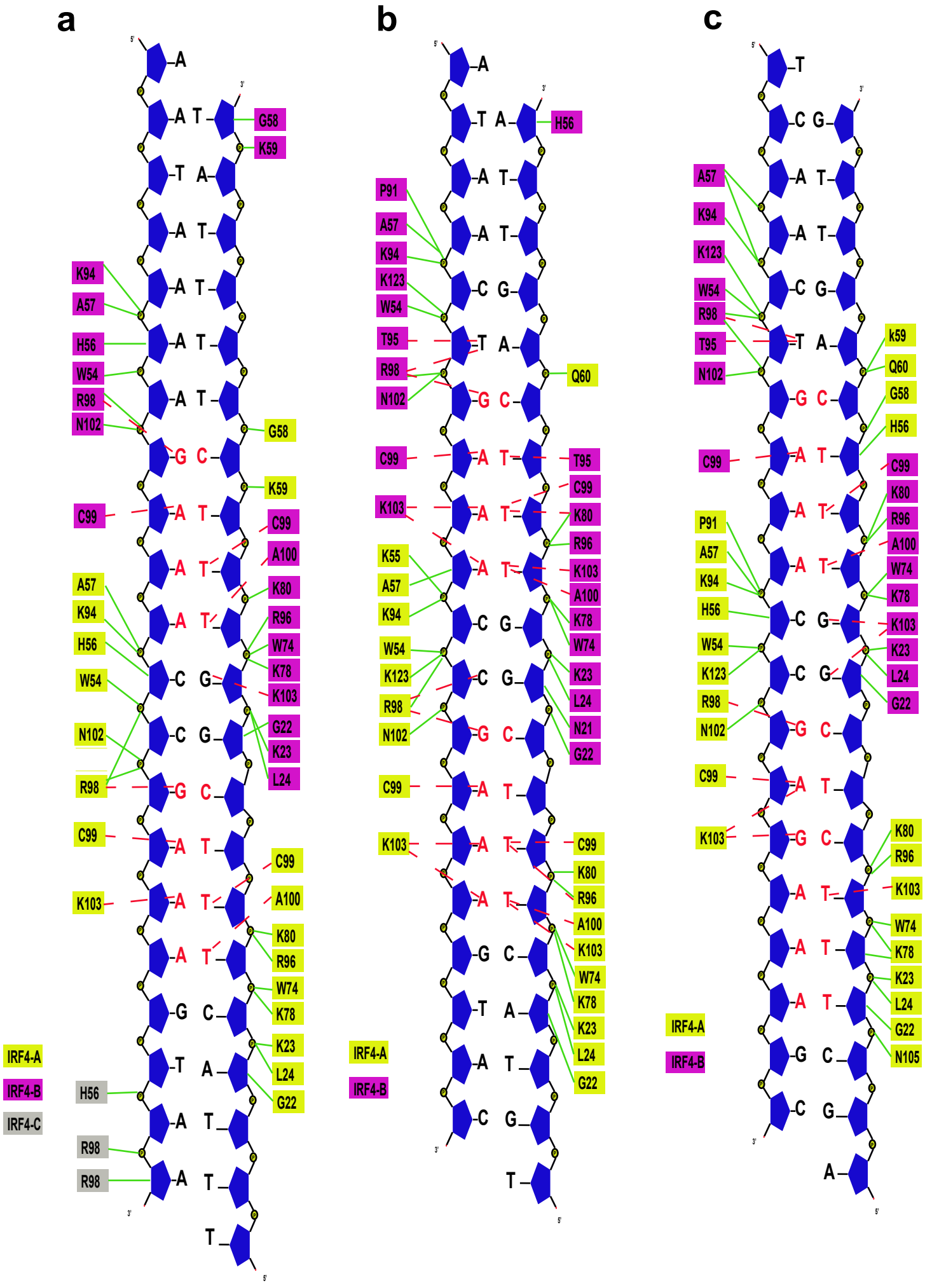
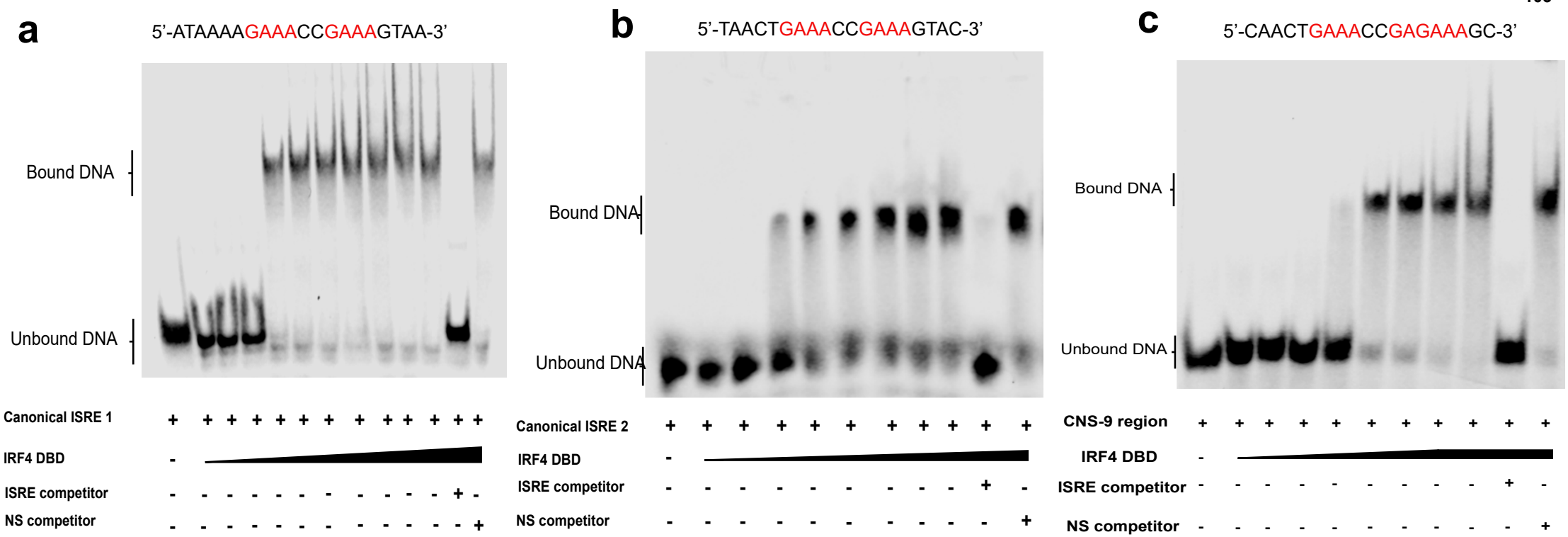

a
b
c
IRF4-A
IRF4-B
IRF4-C
H56
R98
R98
G58
K59
K94
A57
H56
W54
R98
N102
G58
K59
C99
C99
A100
K80
R96
W74
K78
K103
G22
K23
L24
N102
R98
C99
C99
K103
A100
K80
R96
W74
K78
K23
L24
G22
H56
P91
A57
K94
K123
W54
T95
R98
N102
Q60
C99
T95
C99
K103
K80
R96
K55
A57
K94
K103
A100
K78
W54
K123
R98
N102
W74
K23
L24
N21
G22
C99
K103
C99
K80
R96
A100
K103
W74
K78
K23
L24
G22
A57
K94
K123
W54
R98
T95
N102
k59
Q60
G58
H56
C99
C99
K80
R96
A100
W74
K78
K103
K23
L24
G22
P91
A57
K94
H56
W54
K123
R98
N102
C99
K103
K80
R96
K103
W74
K78
K23
L24
G22
N105

Fig. 4


a
5'-ATAAAAGAAACCGAAAGTAA-3'
Bound DNA
Unbound DNA
Canonical ISRE 1
IRF4 DBD
ISRE competitor
NS competitor
b
5'-TAACTGAAACCGAAAGTAC-3'
Bound DNA
Unbound DNA
Canonical ISRE 2
IRF4 DBD
ISRE competitor
NS competitor
c
5'-CAACTGAAACCGAGAAAGC-3'
Bound DNA
Unbound DNA
CNS-9 region
IRF4 DBD
ISRE competitor
NS competitor

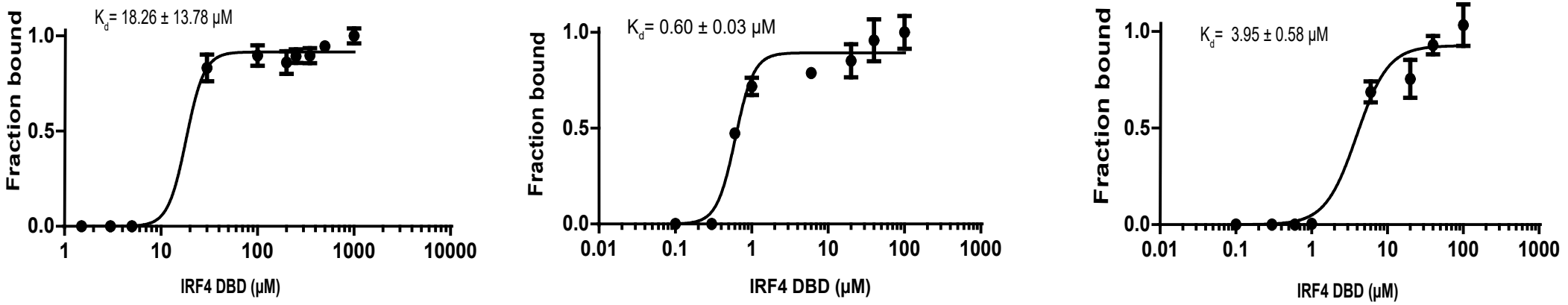
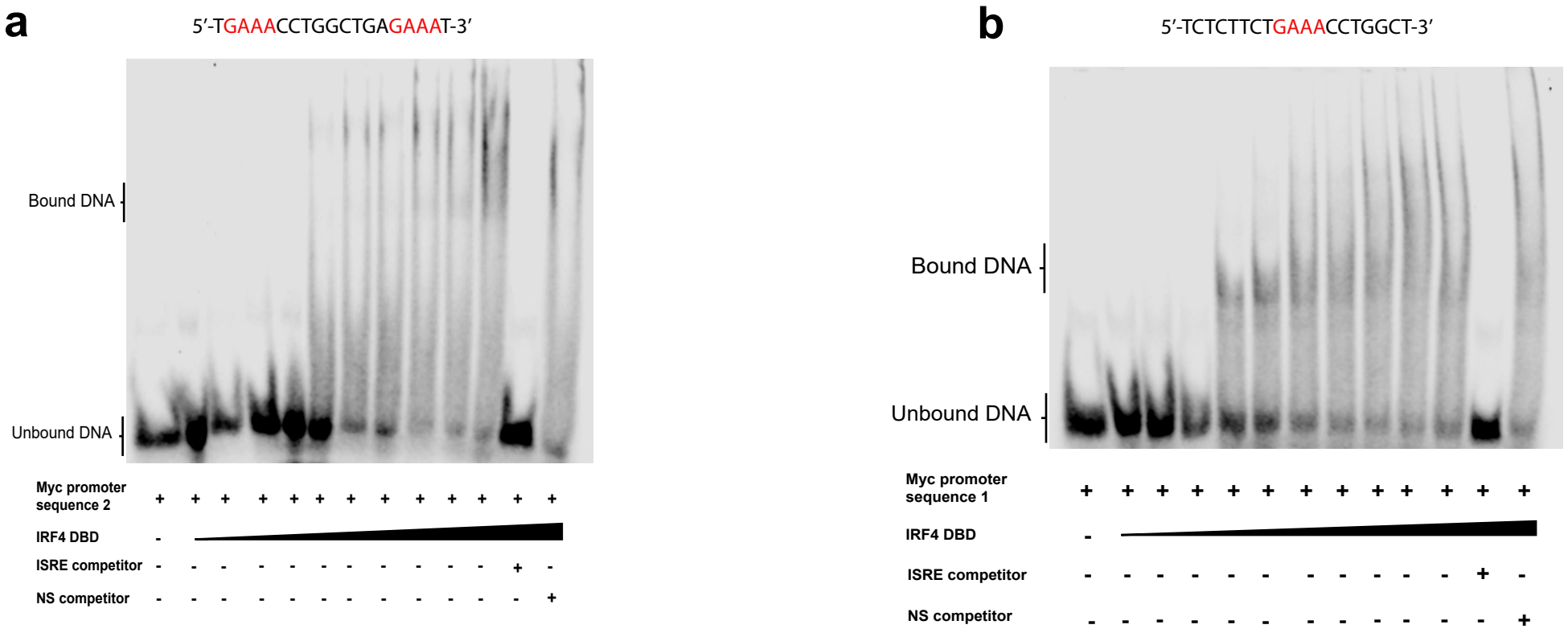
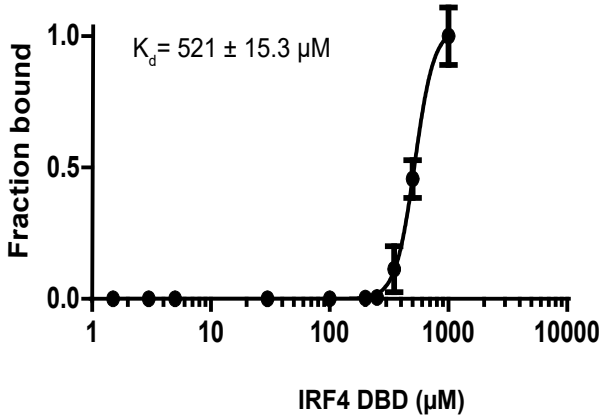

K_d= 18.26 ± 13.78 μM
Fraction bound
IRF4 DBD (μM)
K_d= 0.60 ± 0.03 μM
Fraction bound
IRF4 DBD (μM)
K_d= 3.95 ± 0.58 μM
Fraction bound
IRF4 DBD (μM)

Fig. 5


a
5′-TGAAACCTGGCTGAGAAAT-3′
Bound DNA
Unbound DNA
Myc promoter sequence 2
IRF4 DBD
ISRE competitor
NS competitor
b
5′-TCTCTTCTGAAACCTGGCT-3′
Bound DNA
Unbound DNA
Myc promoter sequence 1
IRF4 DBD
ISRE competitor
NS competitor


Fraction bound
1.0
0.5
0.0
K_d= 521 ± 15.3 μM
1
10
100
1000
10000
IRF4 DBD (μM)

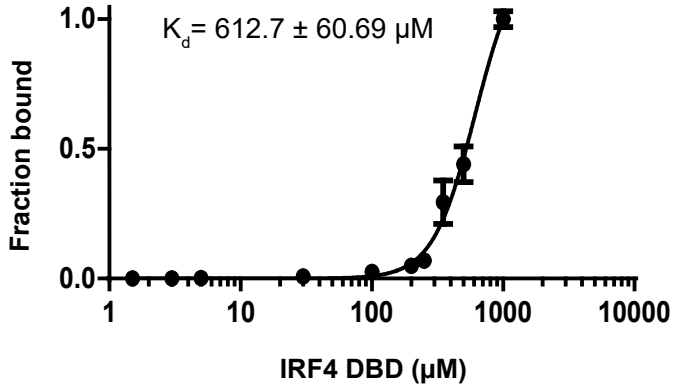
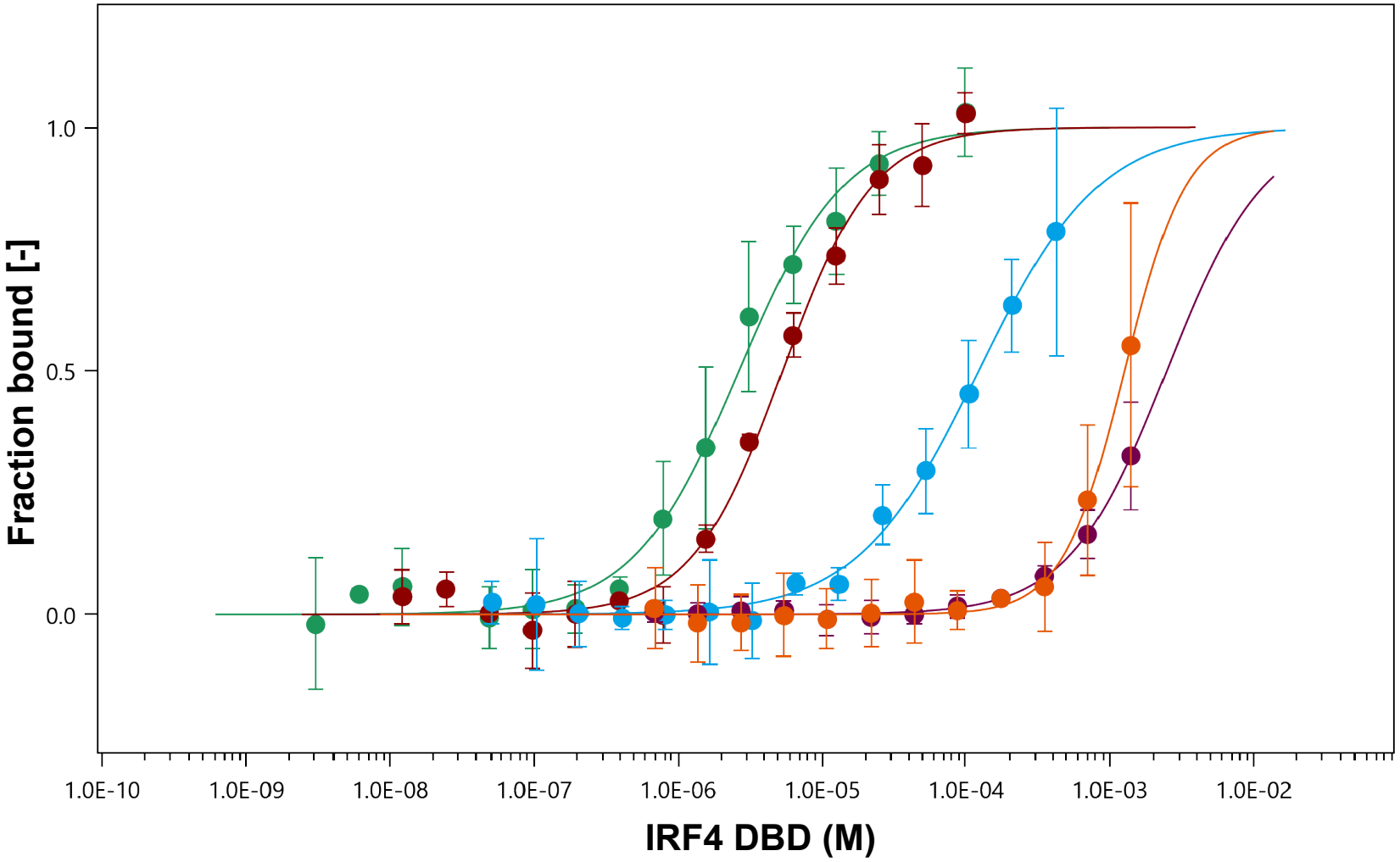

Fraction bound
1.0
0.5
0.0
K_d= 612.7 ± 60.69 μM
1
10
100
1000
10000
IRF4 DBD (μM)

Fig. 6



canonical-ISRE 2	K_d= 2.66μM
CNS-9	K_d= 5.32μM
canonical-ISRE 1	K_d= 122μM
MYC promoter sequence 1	K_d= 2.50mM
MYC promoter sequence 2	K_d= 1.26mM

Fig. 7

6. Chapter 6. Discussion and Future work

6.1. General Discussion

The main aim of this thesis was to lay the groundwork towards the targeting of IRF4 to subvert MM. The third chapter of this thesis showed the indirect targeting of IRF4 by its upstream epigenetic regulators. Targeting epigenetic dysregulation can be an attractive therapeutic strategy in MM [126]. Inhibition of the histone methyltransferase DOTL1 causes cell cycle arrest and apoptosis in MM cell lines [127]. In particular, DOTL1 inhibition targets the IRF4-MYC axis causing a repression of IRF4 and its target genes like *MYC*, *PRDM1* and *KLF2* in MM cells [127]. Moreover, the dual inhibition of the histone methyltransferases EZH2 and G9a reduced MM cell viability exerting a strong antitumor effect [128]. Particularly, in MM cell lines the dual inhibition of EZH2 and G9a blocks the IRF4-MYC axis causing a greater suppressive effects of important oncogenes like *IRF4*, *MYC*, *PRDM1* and *KLF2* than inhibition of either enzyme alone [128] [129]. Treatment of MM cells lines with the epigenetic drugs JQ1, OTX015, SGC-CBP30, ISOX-DUAL caused a significant decrease of MM cell viability and MYC expression. These results are in line with published studies that demonstrated that targeting BET protein family members decreases the expression of MYC and has significant antimyeloma activity [132] [122]. Despite the autofeedback positive loop between IRF4 and MYC in MM [52], our data showed that IRF4 protein levels did not decrease after treatment with specific epigenetic drugs. As shown by our results, this is due to the fact that IRF4 protein is characterised by a very long half-life. Moreover, transcription factor network modelling of MM confirmed this hypothesis and revealed additional and yet uncovered regulatory interactions within the IRF4 network. In conclusion, the epigenetic regulators tested were likely to kill MM cells by affecting MYC and its transcriptional pathway, not IRF4. The clinical utility of most BET inhibitors evaluated to date has been limited due to unexpected toxicities. In particular, the first clinical trials incorporating BET inhibitors tested OTX015 against both haematopoietic and solid cancers showed patients displayed severe dose limited toxicities including gastrointestinal disorders, anaemia, thrombocytopenia, hyperbilirubinaemia, fatigue, headache and back pain [117] [118] [133]. These results together with our data and to the fact that

IRF4 plays a main role in MM biology [52], suggest that an IRF4 direct targeting could be more efficacious to subvert MM. Based on our data, IRF4-direct targeting could be performed by the design of a drug that inhibits IRF4 interaction to ISRE motif. Structural insights about IRF4 FL interaction to ISRE sequences would be also very helpful to target IRF4. In fact, if the IRF4 FL molecules homodimerize during the interaction to ISRE, a drug inhibiting this homodimerization prior to the interaction with ISRE could act as IRF4-direct inhibitor. The fourth chapter of the thesis showed that by using long-wavelengths beamlines, such as I23 at Diamond Light Source, native SAD phasing method can be used to solve the crystal structures of protein-DNA complexes since the technique is particularly suited for sulfur and phosphorus substructure detection. When homologous models for molecular replacement are not available or when molecular replacement is not successful and/or when the initial electron-density maps are not suitable for model building, native intrinsic SAD phasing is particular helpful. The sole reliance on native intrinsic anomalous scattering elements that belong to DNA-protein complexes renders the method of general applicability to a large number of such protein complexes that cannot be solved by molecular replacement or by other phasing methods [134]. The fifth chapter of the thesis showed that IRF4 DBD bound to ISRE sequences without forming any homodimers suggesting that the homodimerization to ISREs might involve other domains of IRF4. Similarly to the structure published by Sundararaj *et al*. (2021), two IRF4 DBD molecules interact with CNS-9 motifs and there is no evidence of protein homodimerization. Our data showed also that IRF4 DBD interaction to canonical ISRE 2 (5'-TAACTGAAACCGAAAGTAC-3') is characterised by two different molecules and the co-operativity in binding to ISRE motifs is due to the allosteric effects along the DNA with no contribution from the interacting DBDs. This suggests that IRF4 adopts a conserved DNA-binding mode when recognizing its DNA targets regardless of difference in ISRE sequences. In the structure presented by Sundararaj *et al*. (2021), the connecting loop L1 showed the greatest RMS deviation (~2.8 Å) of all the structural components suggesting that this loop is inherently flexible. In contrast, superimposition of our three structures with IRF4 NMR structure and apo IRF4 DBD showed structural similarities in all the structural components including L1 loop. Based on our results L1 loop could not be inherently flexible. Our data showed also that the interaction to

canonical ISRE 1 (5’ ATAAAAGAAACCGAAAGTAA-3’) is characterised by three IRF4 DBD molecules. Similarly to the other two structures and to the one published by Sundararaj *et al*. (2021), there is no evidence of protein dimerization and the interaction with ISRE motifs is mediated exclusively by protein-DNA contacts. The presence of the third molecule in this structural analysis is unusual because previous studies suggested the IRF4 binds to ISRE as homodimer [43]. Analysing the IRF4 DBD-canonical ISRE 1 crystal structure, we can see that a pseudo-continuous DNA helix is generated by the crystal lattice which then create a major-groove that the ‘third’ molecule can bind to. This is not present in the other crystal structures where only two IRF4 DBD molecules bind to ISRE motifs. The presence of the third molecule in the IRF4 DBD-canonical ISRE 1 crystal structure could be also depend on the ISRE sequence used. In fact, the overhang sequence upstream to the ISRE motifs might have been created an additional ISRE binding site where IRF4 DBD has bound. Moreover, our data confirmed that IRF4 DBD is characterised by a low affinity to different ISRE motifs and it is DNA shape recognition that drives IRF4 interaction to different ISREs [31] [47]. In particular the very low affinity of IRF4 to single ISRE motifs showed that the spacing between ISRE sequences affects IRF4 affinity to DNA. Our data showed that IRF4 binding to different ISREs induced a significant bend in the DNA duplex. According to literature, DNA binding sites characterised by conformational changes upon protein binding displayed lower affinity than binding sites with more optimal conformations prior to binding [14]. This suggests that the poor IRF4 DBD affinity to ISRE sequences is also affected by the DNA distortion caused by IRF4 binding to DNA. These data provide key insights into the ISRE binding specificity and affinity in the context of MM.

6.2 Future work

6.2.1. Structure-guided fragment-based drug discovery

Based on the data showed in this thesis, there are different future experiments that can be performed. Future work could focus on the targeting of IRF4 DBD-ISRE complexes by using fragment-based drug discovery technique (FBDD). FBDD is an approach to develop potent compounds from fragments. In order to perform a fragment-screening experiment different steps are usually required like selecting a compound library, setting up a method for hit identification, solve the structures of fragment-target complexes, performing an assay for analysing structure-activity relationship (SAR) and designing a strategy to grow the fragment into a potent inhibitor [135]. X-ray crystallography is a powerful tool that plays essential roles in structure-based drug discovery [136]. Indeed, co-crystal structures provide direct and clear information to understand SAR and mechanism of action of the developed compounds [137]. X-ray structures offer structural information to understand binding modes of the inhibitors that bind to the active site of a target, inhibit the target through allosteric mechanisms and form covalent bonds with the target [138] [139]. X-ray crystallography is also a robust method that can be applied in fragment hit identification and confirmation [140]. By performing X-ray crystallography, next experiments should aim to solve the crystal structure of IRF4 DBD-ISRE complex together with a specific chemical fragment. The IRF4 DBD-ISRE crystal structures that I obtained are at medium/high resolution. These crystal structures could be optimal to perform a FBDD. Preliminary results showed that the crystals diffracted after incubation with 5% DMSO suggesting promising results for FBDD. The next step would be to measure the affinity of the specific inhibitor to IRF4 DBD-ISRE complex by using biophysical methods like MST. Further experiments will aim to test the viability of MM cell lines after treatment with the specific inhibitor and performing molecular biology techniques like western blot analysis and qRT-PCR in order to measure the expression of IRF4 protein and IRF4 mRNA levels. These different kinds of experiments will give an initial idea on how successful is the targeting of IRF4 DBD-ISRE complex. Finally *in vivo* drug target validation will be performed first in

animal disease models and after that the potential drug will be tested by different clinical trial phases.

6.2.2. Targeting IRF4 through ROCK2 kinase inhibitors

IRFs are characterised by specific phosphorylated residues in the C-terminal IAD [30]. This specific region is known as SRR and regulate the stabilization of IRF dimers and interaction with DNA [30]. ROCK2 and its only other family member, ROCK1, are highly homologous serine-threonine kinases that serve as major downstream effectors of the Rho subfamily of small GTPases, which includes RhoA [141]. Like other small GTPases, RhoA cycles between an inactive and an active state, a process controlled by Rho-guanine-nucleotide exchange factors (GEFs) [142]. Upon RhoA binding, the ROCKs undergo a conformational change resulting in kinase activation [141]. Previous studies have shown that IRF4 is directly phosphorylated by ROCK2 on two distinct phosphorylation sites, S446 and S447, which are located in the C-terminal IAD [50]. SLx-2119 is a direct inhibitor of ROCK2 [143]. Based on this information, another future experiment should focus on the direct targeting of ROCK2 in order to inhibit the IRF4 translocation to the nucleus and thus causing the inhibition of IRF4 transcriptional activity in MM.

6.2.3. Analysis of the IRF4 full length protein structure and function

Other future work should focus on the IRF4 full length protein (IRF4 FL). Currently there are no crystal structures of IRF4 FL protein. In addition to the DBD, IRF4 FL is characterised by other different domains including a C-terminal IAD domain, critical in mediating protein-protein interactions, connected to the DBD domain via a linker (LK) domain [31]. Because of the presence of flexible and intrinsically disordered regions, IRF4 FL could not crystallise. Alphafold is a very recent software that can predict protein 3-D structures based solely on their amino acid sequence [144]. This computational method can regularly predict protein structures with atomic accuracy even where no similar structure is known [144]. Alphafold could be used to predict the IRF4 FL structure. It would be possible to have an idea on how the different structural domains are

organized and by the superimposition of this predicted structure with IRF4 DBD-ISRE complex structure, it would be possible to hypothesize the interaction of IRF4 FL to ISRE sequences. Moreover, since cryogenic electron microscopy (cryo-EM) does not require protein crystals, it can be used to solve the structure of IRF4 FL. According to literature, IRF4 FL should homodimerize when interacting to ISRE sequences [43]. If this is the case, then future analysis should focus on the specific domains that are responsible for IRF4 FL homodimerization to ISRE sequences. Furthermore, by performing FBDD, it will be possible to find chemical fragments targeting not only IRF4 FL interaction with ISRE sequences but also IRF4 FL homodimerization. It will be then interesting to measure the affinity of IRF4 FL to different ISRE sequences by using MST. Since IRF4 FL is characterised by more domains than IRF4 DBD, potentially this could result in a different affinity of IRF4 FL to ISRE sequences when compared to that of IRF4 DBD.

Bibliography

1. Mitsis, T., et al., *Transcription factors and evolution: An integral part of gene expression (Review).* World Acad Sci J, 2020. **2**(1): p. 3-8.
2. Powell, R.V., et al., *Lineage specific conservation of cis-regulatory elements in Cytokinin Response Factors.* Scientific Reports, 2019. **9**(1): p. 13387.
3. Rebeiz, M. and M. Tsiantis, *Enhancer evolution and the origins of morphological novelty.* Current Opinion in Genetics & Development, 2017. **45**: p. 115-123.
4. Yesudhas, D., et al., *Proteins Recognizing DNA: Structural Uniqueness and Versatility of DNA-Binding Domains in Stem Cell Transcription Factors.* Genes (Basel), 2017. **8**(8).
5. Pabo, C.O., E. Peisach, and R.A. Grant, *Design and selection of novel Cys2His2 zinc finger proteins.* Annu Rev Biochem, 2001. **70**: p. 313-40.
6. Ellenberger, T.E., et al., *The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex.* Cell, 1992. **71**(7): p. 1223-37.
7. Teichmann, M., H. Dumay-Odelot, and S. Fribourg, *Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes.* Transcription, 2012. **3**(1): p. 2-7.
8. Jones, S., *An overview of the basic helix-loop-helix proteins.* Genome Biology, 2004. **5**(6): p. 226.
9. Lohse, M.B., et al., *Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains.* Proceedings of the National Academy of Sciences, 2013. **110**(19): p. 7660-7665.
10. Theobald, D.L., R.M. Mitton-Fry, and D.S. Wuttke, *Nucleic acid recognition by OB-fold proteins.* Annual review of biophysics and biomolecular structure, 2003. **32**: p. 115-133.
11. Cheatle Jarvela, A.M. and V.F. Hinman, *Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks.* EvoDevo, 2015. **6**(1): p. 3.
12. Guertin, M.J. and J.T. Lis, *Chromatin landscape dictates HSF binding to target DNA elements.* PLoS Genet, 2010. **6**(9): p. e1001114.
13. Farley, E.K., K.M. Olson, and M.S. Levine, *Regulatory Principles Governing Tissue Specificity of Developmental Enhancers.* Cold Spring Harb Symp Quant Biol, 2015. **80**: p. 27-32.
14. Zeiske, T., et al., *Intrinsic DNA Shape Accounts for Affinity Differences between Hox-Cofactor Binding Sites.* Cell Rep, 2018. **24**(9): p. 2221-2230.
15. Crocker, J., et al., *Low affinity binding site clusters confer hox specificity and regulatory robustness.* Cell, 2015. **160**(1-2): p. 191-203.
16. Tillo, D., et al., *High nucleosome occupancy is encoded at human regulatory sequences.* PLoS One, 2010. **5**(2): p. e9129.
17. Morgunova, E. and J. Taipale, *Structural perspective of cooperative transcription factor binding.* Current Opinion in Structural Biology, 2017. **47**: p. 1-8.
18. Long, H.K., S.L. Prescott, and J. Wysocka, *Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution.* Cell, 2016. **167**(5): p. 1170-1187.
19. Amoutzias, G.D., et al., *Choose your partners: dimerization in eukaryotic transcription factors.* Trends in Biochemical Sciences, 2008. **33**(5): p. 220-229.
20. Zaret, K.S. and J.S. Carroll, *Pioneer transcription factors: establishing competence for gene expression.* Genes Dev, 2011. **25**(21): p. 2227-41.
21. Kuvardina, O.N., et al., *Hematopoietic transcription factors and differential cofactor binding regulate PRKACB isoform expression.* Oncotarget, 2017. **8**(42): p. 71685-71698.
22. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology.* Cell, 2008. **132**(4): p. 631-44.
23. Iwasaki, H. and K. Akashi, *Hematopoietic developmental pathways: on cellular basis.* Oncogene, 2007. **26**(47): p. 6687-6696.
24. A. Rad and Mikael Häggström, M.D. *Hematopoiesis simple*. 2009; Available from: https://en.wikipedia.org/wiki/Haematopoiesis#/media/File:Hematopoiesis_simple.svg.
25. Ochiai, K., et al., *Zinc finger-IRF composite elements bound by Ikaros/IRF4 complexes function as gene repression in plasma cell.* Blood Adv, 2018. **2**(8): p. 883-894.

26. Huber, M. and M. Lohoff, *IRF4 at the crossroads of effector T-cell fate decision.* Eur J Immunol, 2014. **44**(7): p. 1886-95.
27. Nam, S. and J.-S. Lim, *Essential role of interferon regulatory factor 4 (IRF4) in immune cell development.* Archives of Pharmacal Research, 2016. **39**(11): p. 1548-1555.
28. Battistini, A., *Interferon regulatory factors in hematopoietic cell differentiation and immune regulation.* J Interferon Cytokine Res, 2009. **29**(12): p. 765-80.
29. Hiscott, J., et al., *Convergence of the NF-kappaB and interferon signaling pathways in the regulation of antiviral defense and apoptosis.* Ann N Y Acad Sci, 2003. **1010**: p. 237-48.
30. Thompson, C.D., B. Matta, and B.J. Barnes, *Therapeutic Targeting of IRFs: Pathway-Dependence or Structure-Based?* Front Immunol, 2018. **9**: p. 2622.
31. Remesh, S.G., V. Santosh, and C.R. Escalante, *Structural Studies of IRF4 Reveal a Flexible Autoinhibitory Region and a Compact Linker Domain.* J Biol Chem, 2015. **290**(46): p. 27779-90.
32. Andrilenas, K.K., et al., *DNA-binding landscape of IRF3, IRF5 and IRF7 dimers: implications for dimer-specific gene regulation.* Nucleic Acids Res, 2018. **46**(5): p. 2509-2520.
33. Antonczyk, A., et al., *Direct Inhibition of IRF-Dependent Transcriptional Regulatory Mechanisms Associated With Disease.* Frontiers in Immunology, 2019. **10**(1176).
34. Escalante, C.R., et al., *Structure of IRF-1 with bound DNA reveals determinants of interferon regulation.* Nature, 1998. **391**(6662): p. 103-106.
35. Fujii, Y., et al., *Crystal structure of an IRF-DNA complex reveals novel DNA recognition and cooperative binding to a tandem repeat of core sequences.* The EMBO journal, 1999. **18**(18): p. 5028-5041.
36. Furui, J., et al., *Solution structure of the IRF-2 DNA-binding domain: a novel subgroup of the winged helix-turn-helix family.* Structure, 1998. **6**(4): p. 491-500.
37. De Ioannes, P., C.R. Escalante, and A.K. Aggarwal, *Structures of apo IRF-3 and IRF-7 DNA binding domains: effect of loop L1 on DNA binding.* Nucleic acids research, 2011. **39**(16): p. 7300-7307.
38. Wathelet, M.G., et al., *Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo.* Mol Cell, 1998. **1**(4): p. 507-18.
39. Lin, R., Y. Mamane, and J. Hiscott, *Multiple regulatory domains control IRF-7 activity in response to virus infection.* J Biol Chem, 2000. **275**(44): p. 34320-7.
40. Schmid, S., D. Sachs, and B.R. tenOever, *Mitogen-activated protein kinase-mediated licensing of interferon regulatory factor 3/7 reinforces the cell response to virus.* The Journal of biological chemistry, 2014. **289**(1): p. 299-311.
41. Hagman, J., *Critical Functions of IRF4 in B and T Lymphocytes.* J Immunol, 2017. **199**(11): p. 3715-3716.
42. Shukla, V. and R. Lu, *IRF4 and IRF8: Governing the virtues of B Lymphocytes.* Front. Biol. (Beijing), 2014. **9**(4): p. 269-282.
43. Ochiai, K., et al., *Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4.* Immunity, 2013. **38**(5): p. 918-29.
44. Brass, A.L., A.Q. Zhu, and H. Singh, *Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers.* Embo j, 1999. **18**(4): p. 977-91.
45. Tussiwand, R., et al., *Compensatory dendritic cell development mediated by BATF-IRF interactions.* Nature, 2012. **490**(7421): p. 502-7.
46. Escalante, C.R., et al., *Crystal structure of PU.1/IRF-4/DNA ternary complex.* Mol Cell, 2002. **10**(5): p. 1097-105.
47. Sundararaj, S., et al., *Structural determinants of the IRF4/DNA homodimeric complex.* Nucleic Acids Research, 2021. **49**(4): p. 2255-2265.
48. Chen, W. and W.E. Royer, Jr., *Structural insights into interferon regulatory factor activation.* Cell Signal, 2010. **22**(6): p. 883-7.
49. Cheng, T.F., et al., *Differential activation of IFN regulatory factor (IRF)-3 and IRF-5 transcription factors during viral infection.* J Immunol, 2006. **176**(12): p. 7462-70.
50. Biswas, P.S., et al., *Phosphorylation of IRF4 by ROCK2 regulates IL-17 and IL-21 production and the development of autoimmunity in mice.* J Clin Invest, 2010. **120**(9): p. 3280-95.

51. Wang, L., et al., *Gene expression profiling identifies IRF4-associated molecular signatures in hematological malignancies.* PLoS One, 2014. **9**(9): p. e106788.
52. Shaffer, A.L., et al., *IRF4 addiction in multiple myeloma.* Nature, 2008. **454**(7201): p. 226-31.
53. Rui, L., et al., *Malignant pirates of the immune system.* Nature Immunology, 2011. **12**(10): p. 933-940.
54. Gualco, G., L.M. Weiss, and C.E. Bacchi, *MUM1/IRF4: A Review.* Applied Immunohistochemistry & Molecular Morphology, 2010. **18**(4): p. 301-310.
55. Xu, D., et al., *Interferon regulatory factor 4 is involved in Epstein-Barr virus-mediated transformation of human B lymphocytes.* J Virol, 2008. **82**(13): p. 6251-8.
56. Ramos, J.C., et al., *IRF-4 and c-Rel expression in antiviral-resistant adult T-cell leukemia/lymphoma.* Blood, 2007. **109**(7): p. 3060-8.
57. Sharma, S., et al., *Regulation of IFN regulatory factor 4 expression in human T cell leukemia virus-I-transformed T cells.* J Immunol, 2002. **169**(6): p. 3120-30.
58. Iida, S., et al., *Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma.* Nat Genet, 1997. **17**(2): p. 226-30.
59. Feldman, A.L., et al., *Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas.* Leukemia, 2009. **23**(3): p. 574-80.
60. Di Bernardo, M.C., et al., *A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia.* Nat Genet, 2008. **40**(10): p. 1204-10.
61. !!! INVALID CITATION !!! {}.
62. Dimopoulos, M.A., et al., *Multiple Myeloma: EHA-ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-up.* Hemasphere, 2021. **5**(2): p. e528.
63. Manier, S., et al., *Genomic complexity of multiple myeloma and its clinical implications.* Nat Rev Clin Oncol, 2017. **14**(2): p. 100-113.
64. Bradner, J.E., D. Hnisz, and R.A. Young, *Transcriptional Addiction in Cancer.* Cell, 2017. **168**(4): p. 629-643.
65. Holien, T., et al., *Addiction to c-MYC in multiple myeloma.* Blood, 2012. **120**(12): p. 2450-3.
66. Smittenaar, C.R., et al., *Cancer incidence and mortality projections in the UK until 2035.* Br J Cancer, 2016. **115**(9): p. 1147-1155.
67. Cancer Research UK. *Projections of incidence for myeloma*. 2016; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma/incidence#heading-Three.
68. van de Donk, N., C. Pawlyn, and K.L. Yong, *Multiple myeloma.* Lancet, 2021. **397**(10272): p. 410-427.
69. diag2tec. *Multiple Myeloma*. Available from: https://www.diag2tec.com/our-expertise/multiple-myeloma/.
70. Kumar, S.K. and S.V. Rajkumar, *The multiple myelomas - current concepts in cytogenetic classification and therapy.* Nat Rev Clin Oncol, 2018. **15**(7): p. 409-421.
71. Rajkumar, S.V. and S. Kumar, *Multiple myeloma current treatment algorithms.* Blood Cancer Journal, 2020. **10**(9): p. 94.
72. Rajkumar, S.V., *Multiple myeloma: 2020 update on diagnosis, risk-stratification and management.* Am J Hematol, 2020. **95**(5): p. 548-567.
73. Walker, B.A., et al., *Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms.* Leukemia, 2014. **28**(2): p. 384-390.
74. Robiou du Pont, S., et al., *Genomics of Multiple Myeloma.* J Clin Oncol, 2017. **35**(9): p. 963-967.
75. Wikipedia. *Tumour heterogeneity linear vs branched*. February 27, 2014; Available from: https://en.wikipedia.org/wiki/File:Tumour_heterogeneity_linear_vs_branched.pdf.
76. Corre, J., et al., *Multiple myeloma clonal evolution in homogeneously treated patients.* Leukemia, 2018. **32**(12): p. 2636-2647.
77. Lohr, J.G., et al., *Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy.* Cancer Cell, 2014. **25**(1): p. 91-101.
78. Wikipedia. *Tumour heterogeneity treatment bottleneck*. February 27, 2014; Available from: https://en.wikipedia.org/wiki/File:Tumour_heterogeneity_treatment_bottleneck.pdf.

79. Agnarelli, A., T. Chevassut, and E.J. Mancini, *IRF4 in multiple myeloma-Biology, disease and therapeutic target.* Leuk Res, 2018. **72**: p. 52-58.
80. Agnarelli, A., et al., *Cell-specific pattern of berberine pleiotropic effects on different human cell lines.* Scientific Reports, 2018. **8**(1): p. 10599.
81. Köhl, U., et al., *CAR T Cells in Trials: Recent Achievements and Challenges that Remain in the Production of Modified T Cells for Clinical Applications.* Hum Gene Ther, 2018. **29**(5): p. 559-568.
82. Carpenter, R.O., et al., *B-cell maturation antigen is a promising target for adoptive T-cell therapy of multiple myeloma.* Clin Cancer Res, 2013. **19**(8): p. 2048-60.
83. Seckinger, A., et al., *Target Expression, Generation, Preclinical Activity, and Pharmacokinetics of the BCMA-T Cell Bispecific Antibody EM801 for Multiple Myeloma Treatment.* Cancer Cell, 2017. **31**(3): p. 396-410.
84. Ghermezi, M., et al., *Serum B-cell maturation antigen: a novel biomarker to predict outcomes for multiple myeloma patients.* Haematologica, 2017. **102**(4): p. 785-795.
85. Jagannath, S., et al., *KarMMa-RW: A study of real-world treatment patterns in heavily pretreated patients with relapsed and refractory multiple myeloma (RRMM) and comparison of outcomes to KarMMa.* Journal of Clinical Oncology, 2020. **38**(15_suppl): p. 8525-8525.
86. Munshi, N.C., et al., *Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma.* N Engl J Med, 2021. **384**(8): p. 705-716.
87. Teoh, P.J. and W.J. Chng, *CAR T-cell therapy in multiple myeloma: more room for improvement.* Blood Cancer Journal, 2021. **11**(4): p. 84.
88. García-Guerrero, E., B. Sierro-Martínez, and J.A. Pérez-Simón, *Overcoming Chimeric Antigen Receptor (CAR) Modified T-Cell Therapy Limitations in Multiple Myeloma.* Frontiers in immunology, 2020. **11**: p. 1128-1128.
89. Danhof, S., M. Hudecek, and E.L. Smith, *CARs and other T cell therapies for MM: The clinical experience.* Best Pract Res Clin Haematol, 2018. **31**(2): p. 147-157.
90. Rafiq, S., C.S. Hackett, and R.J. Brentjens, *Engineering strategies to overcome the current roadblocks in CAR T cell therapy.* Nature Reviews Clinical Oncology, 2020. **17**(3): p. 147-167.
91. Sonneveld, P., *Management of multiple myeloma in the relapsed/refractory patient.* Hematology Am Soc Hematol Educ Program, 2017. **2017**(1): p. 508-517.
92. Belluti, S., G. Rigillo, and C. Imbriano, *Transcription Factors in Cancer: When Alternative Splicing Determines Opposite Cell Fates.* Cells, 2020. **9**(3): p. 760.
93. Bushweller, J.H., *Targeting transcription factors in cancer — from undruggable to reality.* Nature Reviews Cancer, 2019. **19**(11): p. 611-624.
94. Arkin, Michelle R., Y. Tang, and James A. Wells, *Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality.* Chemistry & Biology, 2014. **21**(9): p. 1102-1114.
95. Tovar, C., et al., *MDM2 small-molecule antagonist RG7112 activates p53 signaling and regresses human tumors in preclinical cancer models.* Cancer Res, 2013. **73**(8): p. 2587-97.
96. Graves, B.J., et al., *Autoinhibition as a transcriptional regulatory mechanism.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 621-9.
97. Chen, Y.N., et al., *Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases.* Nature, 2016. **535**(7610): p. 148-52.
98. Uversky, V.N., *Intrinsic Disorder, Protein-Protein Interactions, and Disease.* Adv Protein Chem Struct Biol, 2018. **110**: p. 85-121.
99. Dyson, H.J. and P.E. Wright, *Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300.* J Biol Chem, 2016. **291**(13): p. 6714-22.
100. Zhang, Y., H. Cao, and Z. Liu, *Binding cavities and druggability of intrinsically disordered proteins.* Protein Sci, 2015. **24**(5): p. 688-705.
101. Xu, Y. and C.R. Vakoc, *Targeting Cancer Cells with BET Bromodomain Inhibitors.* Cold Spring Harb Perspect Med, 2017. **7**(7).
102. Filippakopoulos, P., et al., *Histone recognition and large-scale structural analysis of the human bromodomain family.* Cell, 2012. **149**(1): p. 214-31.

103. Dhalluin, C., et al., *Structure and ligand of a histone acetyltransferase bromodomain.* Nature, 1999. **399**(6735): p. 491-496.
104. Wu, S.Y. and C.M. Chiang, *The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation.* J Biol Chem, 2007. **282**(18): p. 13141-5.
105. Bisgrove, D.A., et al., *Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13690-5.
106. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
107. Perez-Salvia, M. and M. Esteller, *Bromodomain inhibitors and cancer therapy: From structures to applications.* Epigenetics, 2017. **12**(5): p. 323-339.
108. Filippakopoulos, P., et al., *Selective inhibition of BET bromodomains.* Nature, 2010. **468**(7327): p. 1067-73.
109. Nicodeme, E., et al., *Suppression of inflammation by a synthetic histone mimic.* Nature, 2010. **468**(7327): p. 1119-23.
110. Mirguet, O., et al., *Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains.* J Med Chem, 2013. **56**(19): p. 7501-15.
111. Zuber, J., et al., *RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia.* Nature, 2011. **478**(7370): p. 524-8.
112. Mertz, J.A., et al., *Targeting MYC dependence in cancer by inhibiting BET bromodomains.* Proc Natl Acad Sci U S A, 2011. **108**(40): p. 16669-74.
113. Soodgupta, D., et al., *Small Molecule MYC Inhibitor Conjugated to Integrin-Targeted Nanoparticles Extends Survival in a Mouse Model of Disseminated Multiple Myeloma.* Mol Cancer Ther, 2015. **14**(6): p. 1286-1294.
114. Ott, C.J., et al., *BET bromodomain inhibition targets both c-Myc and IL7R in high-risk acute lymphoblastic leukemia.* Blood, 2012. **120**(14): p. 2843-52.
115. Wadhwa, E. and T. Nicolaides, *Bromodomain Inhibitor Review: Bromodomain and Extra-terminal Family Protein Inhibitors as a Potential New Therapy in Central Nervous System Tumors.* Cureus, 2016. **8**(5): p. e620.
116. Noel, J.K., et al., *Abstract C244: Development of the BET bromodomain inhibitor OTX015.* Molecular Cancer Therapeutics, 2013. **12**(11 Supplement): p. C244-C244.
117. Amorim, S., et al., *Bromodomain inhibitor OTX015 in patients with lymphoma or multiple myeloma: a dose-escalation, open-label, pharmacokinetic, phase 1 study.* Lancet Haematol, 2016. **3**(4): p. e196-204.
118. Berthon, C., et al., *Bromodomain inhibitor OTX015 in patients with acute leukaemia: a dose-escalation, phase 1 study.* Lancet Haematol, 2016. **3**(4): p. e186-95.
119. Cochran, A.G., A.R. Conery, and R.J. Sims, *Bromodomains: a new target class for drug development.* Nature Reviews Drug Discovery, 2019. **18**(8): p. 609-628.
120. Hay, D.A., et al., *Discovery and Optimization of Small-Molecule Ligands for the CBP/p300 Bromodomains.* Journal of the American Chemical Society, 2014. **136**(26): p. 9308-9319.
121. Chekler, E.L., et al., *Transcriptional Profiling of a Selective CREB Binding Protein Bromodomain Inhibitor Highlights Therapeutic Opportunities.* Chem Biol, 2015. **22**(12): p. 1588-96.
122. Conery, A.R., et al., *Bromodomain inhibition of the transcriptional coactivators CBP/EP300 as a therapeutic strategy to target the IRF4 network in multiple myeloma.* Elife, 2016. **5**.
123. Filippakopoulos, P. and S. Knapp, *Targeting bromodomains: epigenetic readers of lysine acetylation.* Nature Reviews Drug Discovery, 2014. **13**(5): p. 337-356.
124. Shorstova, T., W.D. Foulkes, and M. Witcher, *Achieving clinical success with BET inhibitors as anti-cancer agents.* British Journal of Cancer, 2021. **124**(9): p. 1478-1490.
125. Ohguchi, H., et al., *The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival.* Nat Commun, 2016. **7**: p. 10258.
126. Issa, M.E., et al., *Epigenetic strategies to reverse drug resistance in heterogeneous multiple myeloma.* Clin Epigenetics, 2017. **9**: p. 17.
127. Ishiguro, K., et al., *DOT1L inhibition blocks multiple myeloma cell proliferation by suppressing IRF4-MYC signaling.* Haematologica, 2019. **104**(1): p. 155-165.

128. Ishiguro, K., et al., *Dual EZH2 and G9a inhibition suppresses multiple myeloma cell proliferation by regulating the interferon signal and IRF4-MYC axis.* Cell Death Discovery, 2021. **7**(1): p. 7.
129. Pawlyn, C., et al., *Overexpression of EZH2 in multiple myeloma is associated with poor prognosis and dysregulation of cell cycle control.* Blood Cancer Journal, 2017. **7**(3): p. e549-e549.
130. Potterton, L., et al., *CCP4i2: the new graphical user interface to the CCP4 program suite.* Acta crystallographica. Section D, Structural biology, 2018. **74**(Pt 2): p. 68-84.
131. Brass, A.L., et al., *Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1.* Genes Dev, 1996. **10**(18): p. 2335-47.
132. Delmore, J.E., et al., *BET bromodomain inhibition as a therapeutic strategy to target c-Myc.* Cell, 2011. **146**(6): p. 904-17.
133. Lewin, J., et al., *Phase Ib Trial With Birabresib, a Small-Molecule Inhibitor of Bromodomain and Extraterminal Proteins, in Patients With Selected Advanced Solid Tumors.* J Clin Oncol, 2018. **36**(30): p. 3007-3014.
134. Agnarelli, A., et al., *Phosphorus and sulfur SAD phasing of the nucleic acid-bound DNA-binding domain of interferon regulatory factor 4.* Acta Crystallogr F Struct Biol Commun, 2021. **77**(Pt 7): p. 202-207.
135. Li, Q., *Application of Fragment-Based Drug Discovery to Versatile Targets.* Frontiers in Molecular Biosciences, 2020. **7**(180).
136. Thomas, S.E., et al., *Structure-guided fragment-based drug discovery at the synchrotron: screening binding sites and correlations with hotspot mapping.* Philos Trans A Math Phys Eng Sci, 2019. **377**(2147): p. 20180422.
137. Carvalho, A.L., J. Trincão, and M.J. Romão, *X-ray crystallography in drug discovery.* Methods Mol Biol, 2009. **572**: p. 31-56.
138. Anantharajan, J., et al., *Structural and Functional Analyses of an Allosteric EYA2 Phosphatase Inhibitor That Has On-Target Effects in Human Lung Cancer Cells.* Mol Cancer Ther, 2019. **18**(9): p. 1484-1496.
139. Zhong, W., et al., *Targeting the Bacterial Epitranscriptome for Antibiotic Development: Discovery of Novel tRNA-(N(1)G37) Methyltransferase (TrmD) Inhibitors.* ACS Infect Dis, 2019. **5**(3): p. 326-335.
140. Glöckner, S., A. Heine, and G. Klebe, *A Proof-of-Concept Fragment Screening of a Hit-Validated 96-Compounds Library against Human Carbonic Anhydrase II.* Biomolecules, 2020. **10**(4).
141. Julian, L. and M.F. Olson, *Rho-associated coiled-coil containing kinases (ROCK).* Small GTPases, 2014. **5**(2): p. e29846.
142. Jaffe, A.B. and A. Hall, *RHO GTPASES: Biochemistry and Biology.* Annual Review of Cell and Developmental Biology, 2005. **21**(1): p. 247-269.
143. Boerma, M., et al., *Comparative gene expression profiling in three primary human cell lines after treatment with a novel inhibitor of Rho kinase or atorvastatin.* Blood Coagulation & Fibrinolysis, 2008. **19**(7): p. 709-718.
144. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021.
145. Drott, D., *Overcoming the codon bias of E. coli for enhanced protein expression.* inNovations, 2001. **12**.
146. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* J Mol Biol, 1975. **94**(3): p. 441-8.

Appendix

7. Extended Materials and Methods

7.1. Cell Culture

Multiple Myeloma cell lines KMS-12-BM, NCI-H929, SKMM-1 and acute myeloid leukemia OCI-AML3 were grown at 37°C, 5% CO_2 as follows. KMS-12-BM, SKMM-1 and OCI-AML3: RPMI medium with 20% FBS, 1% Penicillin-Streptomycin-L-glutamine, 1% Sodium-Pyruvate and 0.05% Thioglycerol. NCI-H929 RPMI medium with 20% FBS, 1% Penicillin Streptomycin-L-glutamine, 1% Sodium Pyruvate, 0.05% β-mercaptoethanol.

7.1.2. Cell Titer-Blue viability assay

Cells were plated at 20000 cells per well in 96 well plates and treated with DMSO or indicated compounds for 48 hours at 37°C (full concentration range used 0.01-10 µM). Cell viability was analysed by performing CellTiter-Blue® Cell Viability Assay following manufactures instructions (Promega). After 48 hours treatment, 20 µL CellTiter-Blue® Reagent (Promega G8080) was added to each 96 wells plate. The plates were shacked for 10 seconds and then incubated for 2 hours at 37°C. After 2 hours the plates were shacked 10 seconds and fluorescence was recorded at 560/590nm. Cells were plate in triplicate/per condition and at least n=3 assays were performed. IC_{50} was determined using nonlinear regression (curve fit) with log(inhibitor) vs response—Variable slope (four parameters) by GraphPad Prism Version 6.01. Mean IC_{50} was analysed and statistical significance was performed by one-way ANOVA followed by Tukey's test to compare statistical differences among IC_{50s} of different cell lines treated with a specific drug.

7.1.3. Western Blotting

Cell lysates samples were prepared adding 100 µL 1x Gel Sample Buffer/ 10^6 cells. Gel sample buffer: 50 mM Tris pH 6.8, 4% SDS, 5% β-Mercaptoethanol, 0.01% Bromophenol blue, 10%

Glycerol, 1 mM EDTA. The lysates were then sonicated at 25% 5x 10 seconds with 10 seconds gaps on ice. Samples were separated by SDS-PAGE with Tris-Glycine 1X and SDS 1% running buffer for 90 minutes at 120 Volt. Proteins were transferred on to Protran nitrocellulose membranes (Schleicher and Schuell) for 90 minutes at 85 Volt. Membrane were blocked with 5% milk in PBS-Tween for 1 hour. Membranes were incubated with primary antibody overnight at 4°C. Membranes were then washed 3x10 minutes in PBS-Tween. After that, membranes were incubated with HRP-conjugated secondary antibody for 1 hour at room temperature. Membranes were then washed 3x10 minutes in PBS-Tween. Pierce™ ECL western blotting substrate (Thermo Fisher Scientific 32209) was added to the membrane for visualization at LI-COR machine. Stripping was performed by adding Tween® 20 [BP337-100, Thermo Fisher] to the membranes for 10 minutes. Membranes were then washed 3x10 minutes in PBS-Tween and block with 5% milk in PBS-Tween for 1 hour. Primary and secondary antibodies were added as described above. Primary antibodies specific for IRF4 (1:10000, Anti-MUM1 antibody [EP5699] (ab133590), Abcam), Myc (1:300, Antibody (9E10): sc-40, Santa Cruz Biotechnology), β-actin (1:5000, Anti-Actin antibody A2066, Sigma-Aldrich) and HRP-conjugated secondary antibodies anti-rabbit (1:3000, abcam ab205718), anti-mouse (1:5000, Cell signalling 7076S) were used.

7.1.4. RNA Extraction, cDNA synthesis and qRT-PCR

Total RNA was extracted using Monarch total RNA miniprep kit (T2010S). RNA concentrations were determined using a NanoDrop 2000 instrument (Thermo Scientific). cDNA was synthesized by using ImProm-II™ Reverse Transcription System kit with random primers (Promega A3800). RNA/primers mixes were prepared on ice with 1 µg RNA, 1 µl random primers and Nuclease-Free Water up to 5 µL. Reverse transcription reaction contained 4.0 µL ImProm-II™ 5X Reaction Buffer, 4 µL $MgCl_2$ (6.6 mM), 1.0 µL dNTP Mix, 0.5 µL Recombinant RNasin® Ribonuclease Inhibitor, 1.0 µL ImProm-II™ Reverse Transcriptase and Nuclease-Free Water up to 15 µL. 15 µl of Reverse transcription reaction were then mixed to 5 µL RNA/primers mixes in PCR tubes (Axygen® PCR-02-C). cDNA was synthesized by placing the PCR tubes first in a controlled-temperature heat block at 25°C for 5 minutes for the annealing reaction, then in a controlled-temperature heat block at 42°C for 1 hour for the extension reaction. The reverse transcriptase was then inactivated incubating the reaction tubes in a controlled-temperature heat block at 72°C for 15 minutes. cDNA was then used for PCR amplification.

Real time PCR was performed using an Applied Biosystems StepOnePlus PCR machine. In the Real time PCR reactions cDNAs represents 20% of the reaction volume. For each sample we used a reaction volume of 15 µL that was composed by 1X of GoTaq® qPCR Master mix, 2X (Promega A6002), 0.15 µM of each primer, Nuclease-Free Water and 3 µL cDNA. cDNA was amplified by heating samples to 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute prior to dissociation curve analysis. Serial dilutions of cDNA were used to generate standard curves for each primer set (Table 1).

Primer name	**Sequence (5'-3')**
IRF4 forward	AACAAACTGGAGAGAGACCAGACC
IRF4 reverse	CCTCTCCAAAGCATAGAGTCACC
MYC forward	CCTGGTGCTCCATGAGGAGAC
MYC reverse	CAGACTCTGACCTTTTGCCAGG
PRDM1 forward	TACATACCAAAGGGCACACG
PRDM1 reverse	TGAAGCTCCCCTCTGGAATA
KLF2 forward	AGACCTACACCAAGAGTTCGCATC
KLF2 reverse	CATGTGCCGTTTCATGTGCAGC
CDK4 forward	CTTCTGCAGTCCACATATGCAACA
CDK4 reverse	CAACTGGTCGGCTTCAGAGTTTC
hTERT forward	GGAGCAAGTTGCAAAGCATTG
hTERT reverse	TCCCACGACGTAGTCCATGTT
β-actin forward	TTCTACAATGAGCTGCGTGTG
β-actin reverse	GGGGTGTTGAAGGTCTCAAA

Table 1. Primers used to perform qRT-PCR.

7.1.5. Protein half-life

To analyse protein stability KMS-12-BM, NCI-H929, SKMM-1 cells were treated with 10 µg/mL cycloheximide in T25 flasks. Cells were incubated with cycloheximide for up to 72 hours. After 72 hours cells were pipetted into 50 mL tubes and spinned at 1300 rpm for 10 minutes at 277 K. The media was aspirated off and a same or half volume of PBS was added. An aliquot of cell was removed to count the total cell number. After that, the cells were span again (1300 rpm for 10 minutes at 277 K) and 1 mL/5x10^6 cells of PBS was added to each tube. The cells were then spinned at 13000 rpm for 10-20 seconds (pulse). PBS was aspirated off and the pellet was frozen on dry ice. Western blotting analysis was then performed. The half-life protein was quantified by using nonlinear regression (curve fit) with one phase decay GraphPad Prism Version 6.01.

7.1.6. Statistical analysis

Statistical analysis was carried out by using Statgraphics (version XVI) and GraphPad Prism (version 6.01) software. Data tabulation and descriptive statistics were performed by using Excel program (Office 2016). Data are expressed as a mean of three independent experiments with three replicates. Error bars represent standard error of the mean (SEM). Normality was tested by Shapiro-Wilk and Kolmogorov-Smirnov tests. Homoscedasticity was tested by Levene's test. For multiple comparisons of normally distributed data, one-way ANOVA analysis of variance with the Tukey's HSD post-hoc test was performed. P-values < 0.05 were considered to be statistically significant.

7.2. DNA handling and Polymerase Chain Reaction

IRF4 DBD construct was designed into the vector pCDFDuet™-1 so as to produce an His_6-tagged IRF4 protein. Plasmids was designed using the open source software SnapGene Viewer with primers designed using the same software. 50 µL polymerase chain reaction (PCR) reactions were carried out using Thermo Scientific Phusion Flash High-Fidelity PCR Master Mix (F548S). Each 50

µL mixture was made in nuclease free water and contained 25 µL 2X Phusion Flash High-Fidelity PCR Master Mix, 2.5 µL forward primer (10 µM), 2.5 µL reverse primer (10 µM), template plasmid DNA (20 ng). PCR primer oligonucleotides were supplied by Eurofins Genomics as in table 2.

Primer name	**Sequence (5'-3')**
IRF4 DBD forward	CGCGAAACAGCTTAAACCCTGGAATAATAA
IRF4 DBD reverse	TTATTATTCCAGGGTTTACAGCTGTTTCGCG

Table 2. Primers used to design IRF4 DBD construct.

Thermo cycling 2-step protocol was carried out as in table 3.

	Cycles	**Time**	**Temperature**
Denaturation	1	30 seconds	98°C
Denaturation	30	15 seconds	98°C
Annealing		2 minutes and 30 seconds	72°C
Extension		5 minutes	72°C
Final Extension	1	1 minute hold	4°C

Table 3. Thermo cycling conditions used for PCR.

After PCR reaction, digestion with DpnI restriction enzyme (R0176S, New England BioLabs) was performed in order to remove the plasmid template. 1 µL DpnI was added to PCR tubes at 310 K for 15 minutes. High quality DNA from PCR reaction was obtained by using the Monarch® PCR & DNA Cleanup Kit (5 µg) (T1030S, New England BioLabs). DNA was then quantified and transformation followed by plasmid purification were performed in order to send the samples for sequencing. 20 µL of 50 ng/µL samplese were sent for sequencing.

7.3. DNA Quantification

DNA concentration and purity was determined using a Nanodrop 1000 spectrophometer (Thermo Scientific) and corresponding software package. The absorbance of the DNA solution was measured at 260 and 280 nm. Ratios for 260/280 and 230/280 were calculated to give an indication of purity as DNA absorbs UV light at 260 and 280 nm, whereas aromatic proteins (contaminants) absorb significantly at 280 nm. A ratio greater than 1.8 was considered pure. Nucleic acid concentration (c) was calculated automatically from the absorbance at 260 nm using the Beer-Lambert law: $A = \varepsilon lc$ (A: absorbance, ε: extinction coefficient, l: pathlength).

7.4. Transformation of Bacteria by Heat-Shock

Depending on whether the bacteria were being used for producing DNA or expressing protein, different strains of *E. coli* were used. DH5α Competent Cells (Thermo Fisher) were selected as a host strain during cloning and amplification of plasmid DNA. Rosetta cells (Novagen) are derivatives of BL21 *E. coli* and they were selected as they offer enhanced expression of eukaryote proteins by supllying tRNAs for rare bacterial codons [145]. 50µL aliquots of chemical-competent cells were thawed on ice. 300 ng plasmid DNA was incubated with 50 µL cells on ice for 30 minutes. The cells underwent heatshock for 42 seconds at 315 K and then were immediately put on ice for 2 minutes. Following the addition of 450 µL LB broth (no antibiotic), the cells were incubated in a shaker (200 rpm) for 1 hour at 310 K (Innova 4230 refrigerated incubator shaker). 200 µL of the transformed cells were plated onto agar plates containing chloramphenicol and spectinomycin (Rosetta cells) and only spectinomycin (DH5α cells). The cells were incubated at 310 K overnight.

7.5. Plasmid DNA Purification

Succesfully transformed *E. coli* colonies were cultered overnight in 5mL LB media containing an appropriate selective antibiotic (e.g. spectinomycin). The *E. coli* cultures were pelleted (4000 rpm for 20 minutes) and plasmid DNA purified from the cell pellets using the standard procedures associated with the Monarch® Plasmid Miniprep Kit (T1010L, New England BioLabs).

7.6. DNA Sequencing

The DNA sequence of plasmids was confirmed commercially by Eurofins Genomics, using an automated Sanger process [146]. Primer DuetDOWN1 was used for sequence verification of IRF4 DBD construct.

7.7. Glycerol stocks

Transformed *E. coli* cultures were stored as glycerol stocks in order to simply revive them and grow fresh cultures. Glycerol stocks were prepared from the culture of a single colony of transformed *E. coli.* A single colony was picked up from a LB-agar plate and grown overnight in LB media with appropiate antibiotics. 0.7mL of the overnight culture was added to 0.7mL of 80% sterile glycerol in a sterile screw cap freezer tube and stored at 193 K.

7.8. Protein Expression

All expression was carried out in *E. coli* Rosetta cells (Merck Life Science UK Limited) in shaking incubators. Glycerol stocks or single colonies were swabbed and grown in an initial 100 mL LB starter culture which contained chloramphenicol and spectinomycin antibiotics and left shaking overnight at 310 K (Innova 4230 refrigerated incubator shaker). Stock concentrations of 33 mg/mL and 100 mg/mL for chloramphenicol and spectinomycin respectively were used at a 1:1000 dilution for all experiments. 11 mL of pre-culture was used to inoculate 1 L of LB broth with chloramphenicol and spectinomycin for protein expression. These cultures were left shaking at 200 rpm at 310 K until optimal density at 600 nm (OD_{600}) of 0.6 was reached (Thermo Scientific MaxQ 8000). The addition of 0.4mM Isopropyl β-D-1thiogalactopyranoside (IPTG) induced IRF4 DBD. The culture remained shaking (200 rpm) at 293 K overnight before cell pellets were collected by centrifugation at 6238 RCF (relative centrifugal force) for 10 minutes at 277 K and stored at 253 K.

7.9. IRF4 Protein Purification

Cell pellets were resuspended in lysis buffer (0.1 mM $MgCl_2$, 0.01% triton, 25 mM HEPES pH 7.4, 150 mM NaCl, 5 mM Imidazole, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), protease inhibitor cocktail (Roche)) and lysed by sonication on ice. The lysates were clarified by centrifugation at 26700 g for 45 min at 277 K. The supernatants were applied onto a HisPur Cobalt Resin column (Thermo Fisher) previously equilibrated with wash buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 5 mM Imidazole, 0.5 mM TCEP). Following a 10 min incubation at 227 K and the application of 5 column volumes (CV) of wash buffer, the proteins were eluted by addition of elution buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 150 mM Imidazole, 0.5mM TCEP). The collected eluates were concentrated and purified by size exclusion chromatography (SEC) at 277K using a HiLoad 16/600 Superdex 75 prep-grade column (GE Healthcare) in gel filtration buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 0.5 mM TCEP). IRF4 fractions were analysed on a 14% SDS-PAGE gel by electrophoresis.

7.9.1. Protein Acrylamide Gel Electrophoresis

Sodium dodecyl-sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) separates denatured proteins by molecular weight. Gels were prepared by diluting 30% (v/v) 37.5 : 1 acrylamide : bisacrylamide in 400 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.1% (v/v) Tetramethylrthylenediamine (TEMED). Poured resolving gels were overlaid with 70% ethanol and allowed to polymerise for at least 15 minutes. The ethanol was poured off and stacking gel added, which comprised 4% (v/v) acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.1% (v/v) TEMED, into which a comb (1mm thick) was placed to generate wells, which was left to polymerise for 15 minutes. Gels were stained for 30 minutes with Coomassie dye solution (Generon) and PageRuler™ Prestained Protein Ladder (Thermo Fisher) was used.

7.10. Electrophoretic Mobility Shift Assay (EMSA)

8% native acrylamide EMSA gels (Table 4) were prepared and allowed to polymerise for 2 hours before being pre-run for 1 hour at 120 V at 277 K in EMSA running buffer (0.5x Tris-Borate-EDTA (TBE)). EMSA samples were prepared using purified protein and 5'-AF647 fluorescently labelled ISRE DNA (Eurofins Genomic) in EMSA buffer (20 mM HEPES (pH 7.5), 5 mM $MgCl_2$, 150 mM NaCl). Each EMSA sample contained a final concentration of 100 nM ISRE DNA mixed with various protein concentrations. Samples were incubated on ice for 1 hour to allow formation of protein-DNA complexes before being loaded onto the gels. Control samples were made using the non-specific competitor (500 ng Salmon Sperm DNA (Invitrogen)) or the specific competitor (40 µM unlabelled ISRE DNA) mixed with the highest protein concentration. These samples were left on ice for 1 hour before the addition of the fluorescent probe (labelled ISRE DNA) followed by another hour on ice. All samples were loaded into the polymerised EMSA gel alongside an EMSA reference dye (40% sucrose, 0.2% Orange G, 50 mM HEPES pH 7.5). The gel was run again for 1 hour at 120 V at 277 K. The gel was visualised using an Odyssey Fc (LI-COR) machine (2 minutes exposure at wavelength 700 nm). EMSA quantification was determined using One site-Specific binding with Hill slope by GraphPad Prism Version 6.01. EMSA DNAs are shown in table 5.

30% Acrylamide/Bisacrylamide (29:1)	3.2mL
5x TBE	1.2mL
10% APS	80µL
MilliQ Water	7.5mL
TEMED	8µL

Table 4. 8% native acrylamide EMSA gel.

EMSA ISRE DNA	Sequence (5'-3')
ISRE canonical ISRE 2 forward	TAACTGAAACCGAAAGTAC
ISRE canonical ISRE 2 reverse	GTACTTTCGGTTTCAGTTA
CNS-9 region forward	CAACTGAAACCGAGAAAGC
CNS-9 region reverse	GCTTTCTCGGTTTCAGTTG
ISRE canonical ISRE 1 forward	ATAAAAGAAACCGAAAGTAA
ISRE canonical ISRE 1 reverse	TTACTTTCGGTTTCTTTTAT
MYC promoter sequence 2 forward	TGAAACCTGGCTGAGAAAT
MYC promoter sequence 2 reverse	ATTTCTCAGCCAGGTTTCA
MYC promoter sequence 1 forward	TCTCTTCTGAAACCTGGCT
MYC promoter sequence 1 reverse	AGCCAGGTTTCAGAAGAGA

Table 5. DNA used in EMSA assay.

7.11. Microscale Thermophoresis (MST)

The MST assays were performed using the Monolith NT.115 machine. Different ISRE DNAs were labelled at 5'-ends using the dye AF647 (Eurofins Genomic). Serial dilutions of IRF4 DBD protein in MST buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% Tween-20) were mixed with 20 nM of 5'-AF647 ISRE DNA and loaded into standard glass capillaries (Monolith NT.115 Capillaries, NanoTemper Technologies). Thermophoresis analysis was performed on a Monolith NT.115 instrument (20% LED, 40/60% MST power) at 295 K. The MST curves were fitted using NT Analysis software (NanoTemper Technologies) to obtain EC50 and Hill coefficient values for binding. DNA used in MST are the same as EMSA and are shown in table 5.