The Conformational Dynamics and Molecular Mechanism of KstR.

Ali Razzak

5268567

Supervisor: Dr. J. Shaun Lott

Co-supervisor: Dr. Jane Allison

A thesis submitted in partial fulfilment of the requirement for the degree of Master of Science (Biological Science). March 14, 2018.

Abstract

Mycobacterium tuberculosis (Mtb) contains a remarkable array of lipid and cholesterol metabolic processes. The expression of genes involved in cholesterol degradation is controlled by KstR, a transcriptional repressor in Mycobacteria. KstR has a pivotal role in facilitating the persistence of Mtb within human hosts through the regulation of cholesterol metabolisation.

KstR activity can be reduced to a three state model. The first state is apo KstR. The second state is DNA-bound KstR, presenting a contracted shape compared to apo that is structurally restrained due to its association with DNA. The third state is ligand-bound KstR, where it associates with metabolites produced early in the cholesterol degradation pathway, which bind to the ligand binding pocket. Ligand-bound KstR is more structurally constrained than apo KstR, locking KstR into a wider architecture, and preventing DNA binding. Modulation of the conformation of KstR is essential in informing its association with DNA, and thereby its ability to repress cholesterol degradation genes.

By starting with the resolved structures of apo KstR, DNA-bound KstR, and ligand-bound KstR, the activity of KstR was modelled through computational simulations which revealed potential KstR conformational trajectories. Molecular dynamics simulations have provided insight into which regions of KstR change conformation, and the extent of conformational deviation that different states of KstR can take on. This demonstrated distinct conformational paths between apo KstR, DNA-bound KstR, and ligand-bound KstR. This research has contributed evidence towards a mechanism of action of KstR.

Acknowledgements

I would like to express my utmost gratitude to my supervisor Dr. Shaun Lott for his support, guidance, and charm. The opportunities and resources he has disposed me with have had an ineffable impact on me.

I would also like to thank my co-supervisor Dr. Jane Allison. My everlasting appreciation goes to her ability to orchestrate what has been the most phenomenal and seamless working relationship I have been fortunate enough to be a part of.

All I know of Molecular Dynamics has been generously and exhaustively bequeathed unto me by Dr. Thomas Coiler of the Allison Laboratory. His immense patience, careful instruction and wealth of knowledge drove if not propelled much of the material which constitutes this project. It occurred to me this project aligned with his arrival, serendipity if I ever saw it.

Special thanks goes to the Dr. Thu Ho of the Lott Laboratory who possesses a more profound knowledge of science than any Google search would grace me with. I am pressed to recall a more kinder and warmer person than her.

I would also like to extend my appreciation to members of both the Lott Laboratory and Allison Laboratory. Without their ability to pull me back into society, even if momentarily, I'm not sure what sanatorium would take me.

Finally, I would like to thank my family. Suffice to say there is nothing without them. In a similar vein, thanks goes out to my friends who have consistently reciprocated in kind where I had at least been reticent. I hope to see more of you.

List of Figures

List of Tables

Abbreviations and Acronyms

3OChA-CoA	3-oxo-4-cholestenoic-CoA
4-BNC-CoA	3-oxo-23,24-bisnorchol-4-en-22-oyl-CoA
bp	base pair
CoA	Coenzyme-A
DNA	Deoxyribonucleic acid
DBD	DNA binding domain
DSSP	Define Secondary Structure of Proteins
FLB	Full ligand-bound
FrFLB	From full ligand-bound
FrLDB	From long DNA bound
H-bond	Hydrogen bond
HTH	Helix turn helix
INF	Interferon
LBD	Ligand binding domain
LoDB	Long DNA bound
LJ	Lennard-Jones
***R*g**	Radius of gyration
RNA	Ribonucleic acid
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
MD	Molecular dynamics
MDR	Multi drug resistant
Mtb	*Mycobacterium tuberculosis*
PDB	Protein database
SASA	Solvent accessible surface area
SLB	Short ligand-bound
SLDB	Shortened long DNA bound
TB	Tuberculosis
TetR	Tetracycline receptors
TFR	TetR family of regulators
VDW	Van der Waals
XDR	Extensively drug resistant

Contents

Chapter 1

Introduction

1.1 Tuberculosis

1.1.1 Burden on humanity

Tuberculosis (TB) causes more deaths annually than any other infectious disease. 10.4 million people were estimated to be afflicted by TB in 2016, of which 6.3 million people were diagnosed in 2016 [1] [6]. There were a total of 1.7 million deaths caused by TB or complicated by TB (by conditions such as HIV) over 2016. The disease affects different regions of the world disproportionately, with the most affected countries located in sub-Saharan Africa, South-East Asia, and the West Pacific, as illustrated on Figure 1.1.1 [1].

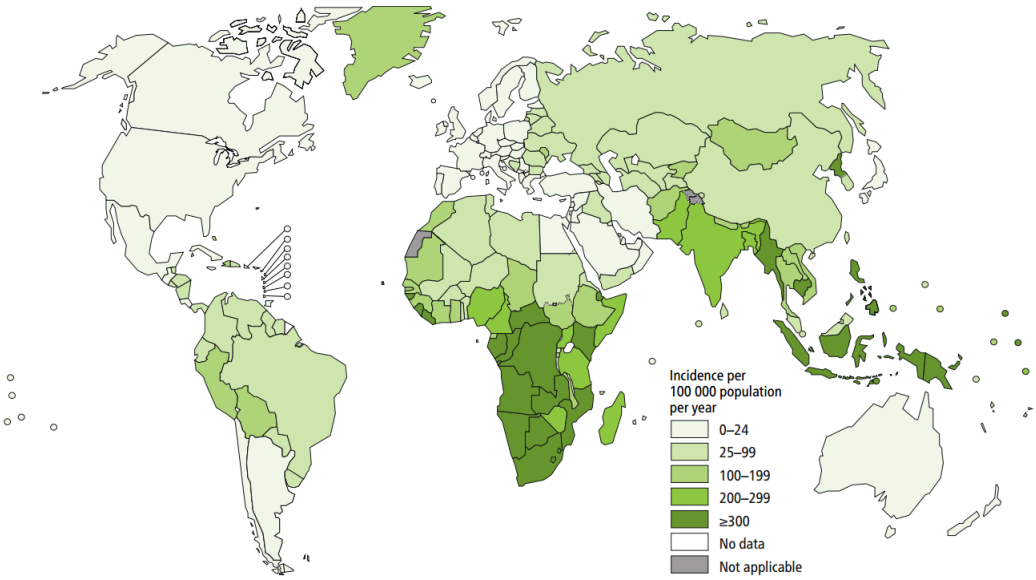

Figure 1.1.1: *Mapped estimates of TB incidence rates globally, hue of green indicating severity of incidence for assigned country, 2016. Reproduced with permission from WHO [1].*

1.1.2 Mycobacterium tuberculosis pathogenicity

The causative agent of TB is *Mtb*. *Mtb* is a pathogenic bacterial species in the Mycobacteriaceae family. Upon infecting a human host, *Mtb* targets the respiratory system, thriving in the lungs [2]. This can lead to lifelong infections that are diagnosed as chronic TB [7]. *Mtb* can engage in cycles of activity and quiescence within the host's lungs. The ability to fluctuate between activity and quiescence is essential to the virulence and persistence of *Mtb* within a host [8].

The virulence of *Mtb* is enhanced by its capacity to persist in a human host's system despite facing immense duress [9] [10]. *Mtb* subverts and exploits the body's innate immune system to its advantage [11]. An *Mtb* infection traumatises cells which line the lungs, activating the innate immune system, causing alveolar macrophages to migrate to the site. Macrophages phagocytose *Mtb*, impeding the spread of the bacteria within the host [2]. Although the metabolic activity of *Mtb* is suppressed, it can persist within the host, resulting in a active innate immune response [11] [12].

A chronic *Mtb* infection results in the iterative migration of macrophages to the site. This leads to the formation of a granuloma which isolates afflicted regions and divorces *Mtb* from nutrients essential for metabolic activity, as illustrated in Figure 1.1.2 [2]. Bacteria and foreign material within the granuloma are phagocytosed by macrophages, capturing them in phagocytic vesicles [13]. Phagosomes house hydrolytic enzymes and isolate *Mtb* cells within a acidic organelle before transporting their contents to lysosomes; all of which facilitates *Mtb* clearance [13]. However, *Mtb* possesses mechanisms which undermine the phagosomal activity of macrophages, allowing it to retain a pH of 6.2 - 6.3, access iron sources, and escape lysosomal degradation [14]. While the physiological conditions inside the granuloma are deleterious to *Mtb*, they also mask *Mtb* from the immune system. Together, these adaptations allow *Mtb* to proliferate intracellularly in macrophages [10] [15].

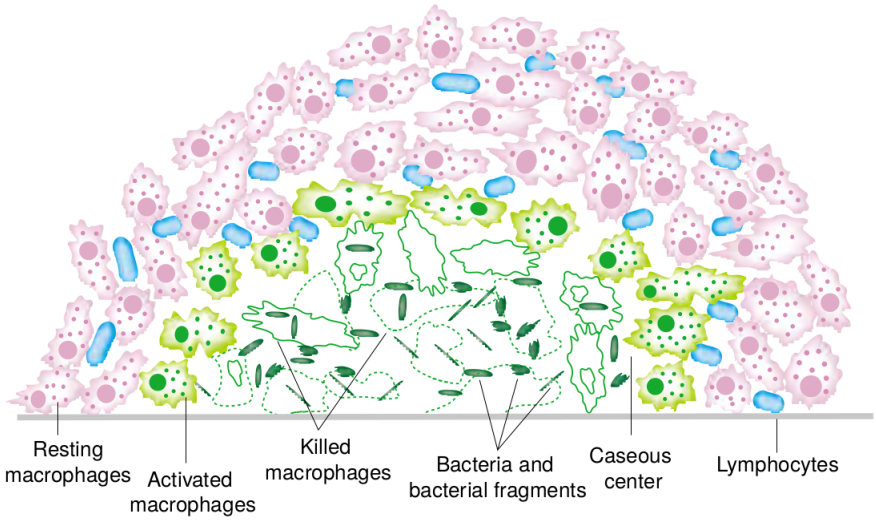

Figure 1.1.2: *Cartoon showing composition of granuloma and organisation of the immune cells surrounding Mtb infected tissue. The caseous centre (Dark green) of granuloma, composed of necrotised tissue, killed macrophages, bacteria and bacterial fragments, is surrounded by multi-nucleated activated giant macrophages. At the periphery of the granuloma there are immune cells such as resting macrophages (pink) and lymphocytes (light blue). Reproduced with permission from Bentrup et al [2].*

Employment of the adaptive immune system curtails the phagosomal resilience of *Mtb* by introducing cytokines, such as interferon-γ (INF-γ), that modify the actions of macrophages [11]. INF-γ activated macrophages produce nitric acid and harsher oxidative pressures within their phagosomes thereby enhancing their bacterial cell killing capacity [11] [13] [15] [16]. Despite the immune system's measures to suppress *Mtb* proliferation, the bacteria is capable of persisting within the granuloma environment by entering a metabolically active quiescent state [9].

Under conditions of extreme metabolic and physiological stress, like that of the interior of a macrophage's lysosome, *Mtb* takes on a quiescent state [15]. For *Mtb* to sustain itself in a quiescent state it seeks out alternative sources of nutrients [2] [17]. It has been observed that many members of the Mycobacteriaceae family, including *Mtb*, have conserved the apparatus required to catabolise cholesterol for essential lipids for the cell membrane, among other cell components [4] [18] [19]. Maintenance of its quiescent state is dependent on access to cholesterol, a substrate which *Mtb* can preferentially metabolise [15] [20] [21].

1.1.3 Treatment regimen

The standardised first line treatment regime of *Mtb* is directly observed therapy short-course (DOTS) [1] [22]. DOTS is prescribed following case detection by sputum smear and empirical treatment regimens. The treatment is composed of a cocktails of drugs (Isoniazid, Rifampicin, Pyrazinamide and Ethambutol) over the course of six months [23]. Rigorous applications of the DOTS program

has led to curing TB in the 85-90% of cases, however it is not successful in all settings [1] [23] [24].

The DOTS regimen is prescribed for six months as *Mtb*-infected hosts often relapse following initial treatment [24]. The treatment is slow at treating TB, as the drug only kills replicating *Mtb*. The long treatment time of ensures that infrequently replicating bacteria are allowed a chance to replicate and hence be killed. Resistance mutations occur spontaneously at a low but known frequency [1] [23] [25].

The shortcomings of DOTS is a consequence of it being slow acting, causing poor compliance. Where the DOTS regimen is compromised, the likelihood of resistant strains being selected for is enhanced, resulting in Mult-Drug-Resistant (MDR) strains [23] [24]. The combination of incomplete therapy and mono-therapy are thought to be the main contributors towards the generation of MDR and Extensively Drug Resistant (XDR) bacterial strains [1].

1.1.4 Antibiotic resistance

TB has been recognised as a dire international threat, with spending on TB prevention, diagnosis and treatment by both private and government institutions raising over the last decade, as detailed on Figure 1.1.3 [1]. TB has progressively become a more dangerous and difficult disease to treat due to its evolving antibiotic resistance, and propagating MDR strains [26]. This has contributed to the death toll of TB, causing it to be ranked as the number one killer among infectious diseases [1] [27].

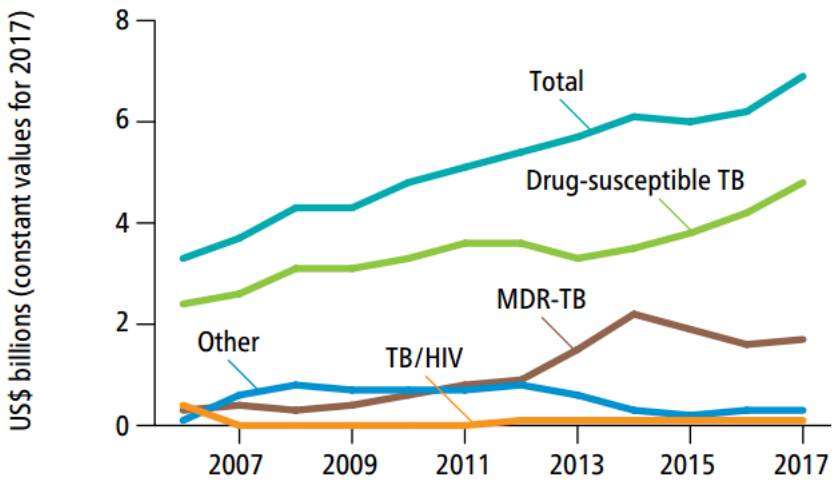

Figure 1.1.3: *Funding for TB prevention, diagnosis and treatment in total and by category of expenditure, from 2006 to 2017, across 118 countries with 97% of reported cases. Reproduced with permission from WHO [1].*

The antibiotic adaptations and drug resistance of *Mtb* makes TB a difficult disease to treat. Research into various antibiotic therapy strategies have been

pursued to combat the disease [2]. A variety of first line drugs that target essential *Mtb* processes are administered to combat TB including [11] [28].

Despite drug treatment, residual populations of *Mtb* can successfully persist and adapt to evade an antibiotic's cell-clearing action. Antibiotic drug exposure selects mutations which render the antibiotic's impotent, giving rise to antibiotic resistant variants of *Mtb*. [1] [8] [15]. The chronic nature of *Mtb* and the induction of persistent strains operate synergystically to produce an increasingly resistant latent reservoir of MDR *Mtb* [2]. MDR *Mtb* strains increasingly favour the selection of resistant mutations as they are continually exposed to new antibiotics. This can generate especially resistant strains, categorised as Extensively Drug Resistant (XDR) [1].

Poorly considered drug treatment regimen can increase the resistance capability of existing strains, heightening the prevalence of MDR and XDR *Mtb* [1] [8]. There were 153,119 cases of MDR TB identified in 2016 comprising 2.4% of all TB incidents that year. Furthermore, of cases tested there was a 31% increase in Rifampicin resistant strains in 2016 [1]. This has motivated intensive research to thoroughly examine the pathology of *Mtb* in mammalian hosts in hopes of identifying novel targets in order to control the TB pandemic.

Antibiotic drug regimes are used to treat TB by killing *Mtb*, but this can be ineffective against MDR and XDR strains of *Mtb* [29]. Drug resistant *Mtb* cases constitute among the greatest challenges to modern medicine due to dwindling pools of antibiotics and evolving anti-antibiotic adaptations that *Mtb* cases are mounting globally [1] [17]. There is a severe need to carefully examine the pathology of *Mtb* in the hope of identifying novel targets that may overcome drug resistant *Mtb* populations and slow the spread of TB.

1.1.5 Mycobacterium tuberculosis cholesterol catabolisation

Mtb is heavily dependent on lipid metabolism - both fatty acids and cholesterol can be used [9]. The bacterium can utilise cholesterol as its sole carbon source via β-oxidative and energy production metabolic pathways, and during chronic phases of infection *Mtb* is reliant on cholesterol [30] [31]. This is highlighted by its preference for invading lipid-dense foamy macrophages [2] [20].

There are approximately 250 distinct fatty acid degrading enzymes involved in the fatty acid oxidation processes of *Mtb* [30]. The fatty acid degradation potential of *Mtb* is demonstrated by its utilisation of the methylcitrate cycle, methylmalonyl pathway, and tricarboxylic acid cycle [20] [30] [32]. This metabolic arsenal provides *Mtb* with enzymes that prompt the breakdown of host membrane layers, manipulation of host cellular behaviour, synthesis of the lipid cell envelope, and detoxification of oxidative stress [2] [15].

Expression of cholesterol metabolising enzymes is predominantly regulated by KstR. Constituents of the KstR regulon include: 3-hydroxy-9,10-seconandrost-

1,3,5(10)-triene-9,17-dione hydroxylase reductase (HsaB), 3,4-DHSA hydrolase (HsaC), 4,9-DHSA hydrolase (HsaD), and 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase subunit (HsaA) among others [30] [31] [7] [33]. Mutational studies of genes within the *Mtb* KstR regulon showed that when its constituents were knocked out, survival of *Mtb* cells within the host's lungs was severely attenuated [21].

While cholesterol metabolism is essential during the quiescent phase of *Mtb*, the cholesterol degradation apparatus is not constitutively expressed, revealing that *Mtb* can modulate the expression of cholesterol degradation enzymes [20]. The expression of cholesterol metabolising enzymes within *Mtb* is dictated by nutrient restriction as imposed by INF-γ activated macrophages and granuloma entrapment [9]. The primary factor that governs *Mtb* cholesterol metabolism is KstR, a Tetracycline Receptor (TetR) protein that regulates the expression of cholesterol metabolising enzymes [34].

1.1.6 Overview of TetR family proteins

TetR is a bacterial protein that regulates the efflux of antibiotics [35] [36]. Almost every sequenced bacterium has been found to express at least one TetR-like protein, whereas bacteria with larger genomes have been found to possess multiple homologues [34]. The characteristics of TetR are observed frequently among other prokaryotic proteins giving rise to the Tetracyline Family of Regulators (TFR) [37] [38].

A member of the TFR is identified through common architectural characters, including: a conserved, predominantly α-helical structure, a ligand binding domain (LBD) and a DNA binding domain (DBD) [39]. An example of TFR architecture is presented in Figure 1.2.2. The TFR has been thoroughly researched, boasting 200,000 sequences along with approximately 200 structures available in public databases [34]. Many proteins within the TFR are responsible for antibiotic resistance and gene regulation, operating as single-component signal transduction systems that interact with a vast array of ligands [34] [39].

While antibiotic efflux regulators like TetR constitute 25% of the TFR, it is not limited to that function [39]. Different TFR members are responsible for other functions including: metabolic gene control, antibiotic production, and regulation of prokaryotic physiology [19] [29] [38]. Members of the TFR have been found to function as both activators or repressors, local or global regulators, and can interact with ligands and proteins of various sizes [39]. Despite the immense amount of research invested into the TFR central questions addressing the family's protein dynamics remain elusive.

1.2 KstR

1.2.1 KstR background

Investigations into what enables the persistence of *Mtb* revealed its dependence on cholesterol metabolisation [20] [40]. Bioinformatic analysis of microarray data led to the identification of KstR, a transcriptional repressor that predominantly modulates the expression of cholesterol degradation machinery in *Mtb* [3] [41].

Similar to other members of the TetR family of proteins, the activity of KstR relies on its homodimeric structure. Each monomer is composed of a DBD and a LBD [42]. The DBD binds to specific motifs along the *Mtb* genome. Early stage coenzyme-A (CoA) thioester metabolites of cholesterol, 3-oxo-4-cholestenoic-CoA (3OChA-CoA) and 3-oxo-23,24-bisnorchol-4-en-22-oyl-CoA (4-BNC-CoA), bind to a cavity in the LBD and prevent DNA binding [4]. KstR can be thought of as a cholesterol detector: in the absence of cholesterol it represses cholesterol metabolisation, whilst when cholesterol is present, it induces cholesterol metabolisation.

X-ray crystal structures have revealed that KstR occupies three distinctive conformations. When bound to DNA KstR takes on a contracted conformation and represses transcription. When bound to ligands it adopts a wider and flatter conformation that disallows DNA binding. If neither a ligand or DNA is bound it samples a conformation between the two [4] [18]. This is summarised in Figure 1.2.1. While these three states are well characterised, it is unclear how the conformation of KstR reorganises to transition between these three conformations.

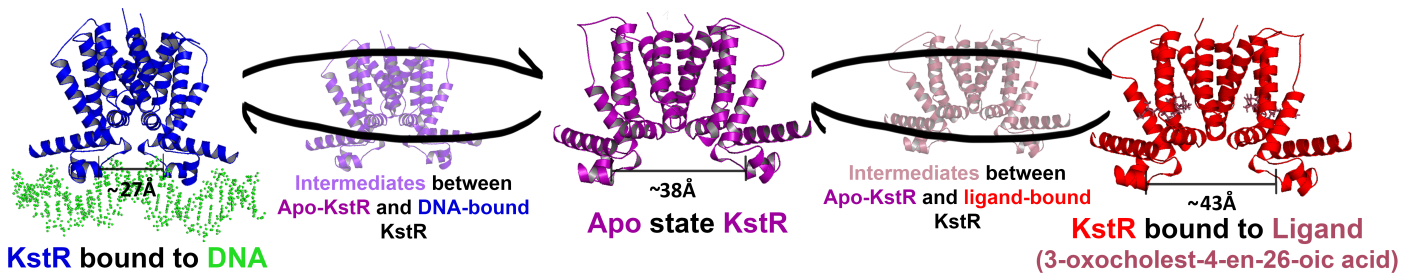

Figure 1.2.1: *Cartoon representation of proposed KstR transition model. DNA-bound KstR Protein Data Bank (PDB): 5UA2, apo KstR: 3MNL, ligand-bound KstR PDB: 5CW8. The intermediate structures were extracted from MD trajectories. The distance between the DBDs for the three crystal structures is labelled beneath in Å.*

1.2.2 KstR Structure

Similar to other TFRs, KstR is primarily composed of α-helices and endogenously forms a homodimeric complex [26] [42]. Each KstR monomer contains

nine α-helices connected by coiled regions. Critical to the transcriptional regulation function of KstR are the DBD and LBD within the structure of KstR. A linker sequence between the $\alpha4$ and $\alpha5$ helices connects the DBD and the LBD. The DBD is positioned at the N-terminus of the protein and is composed of helices $\alpha1$ - $\alpha3$, of which $\alpha2$ and $\alpha3$ form a helix-turn-helix (HTH) motif, characteristic of the TFR and also found in a wider range of DNA binding proteins [4] [34]. The *apo* KstR DBD has a larger average B-factor than the LBD, a product of its exposure to solvent and sparse interaction among its components. Helix $\alpha4$ operates as a conduit between the DBD and LBD, anchoring the linker sequence between $\alpha5$ and itself. Helices $\alpha4$ - $\alpha9$ scaffold the C-terminal LBD, within which the ligand docks into a deep hydrophobic pocket. The dimer interface between the two KstR monomers occupies 1099 $Å^2$ of the 9500 $Å^2$ total surface area of each KstR monomer, made up of contributions from the $\alpha6$ - $\alpha9$ helix bundle. The ligand binding pocket is formed by the triangular groove presented between helices $\alpha5$ - $\alpha7$, with hydrophobic residue contributions from $\alpha9$. The KstR homodimer produces a clamp-like structure. Figure 1.2.2, displays the DBD, LBD and α-helix assignment [4].

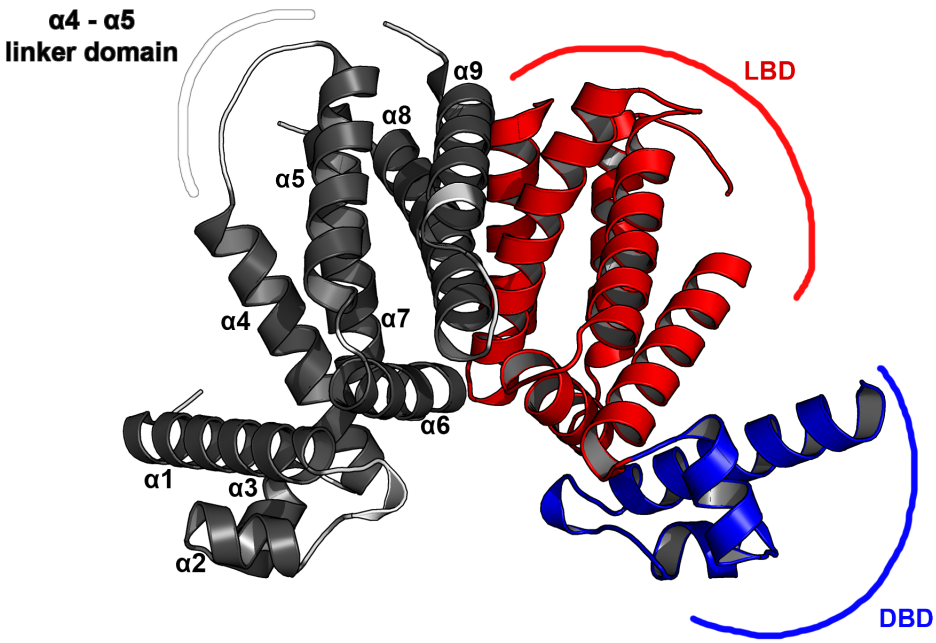

Figure 1.2.2: *Cartoon representation of KstR (PDB: 3MNL) dimer with α helices annotated in one subunit (black), and the DBD (blue) and LBD (red) regions indicated in the second subunit.*

1.2.3 Apo KstR

The free conformation of KstR informs how it associates with DNA and its mechanism of action [3]. *Apo* KstR is loosely restrained, allowing it to take on a diverse range of conformations [36] [43].

The flexibility of *apo* KstR is a consequence of multiple structural proper-

ties [36]. Firstly, the DBD is composed of the minority of α-helices (3 of 9 in the KstR monomer) and hydrophobic residues compared to the LBD, causing it to be less restrained by local chemical and structural forces [44]. Secondly, the DBD is positioned towards the periphery of the complex compared to the LBD, consequently it is more exposed to aqueous molecules. Thirdly, the DBD is tethered to the LBD through a long coiled linker domain between $\alpha 4$ and $\alpha 5$, increasing the mobility of the DBD [43].

While the binding dynamics between the *apo* KstR dimer and DNA are unknown, two possible pathways can be considered: induced fit or conformational selection, both portrayed in Figure 1.2.3 [45]. The induced fit model suggests a two-step binding procedure: initially a single DBD of the free KstR homodimer contacts and binds DNA, forming a partial DNA-bound KstR complex. The association with DNA structurally restricts the KstR dimer, favouring binding of the second DBD to the DNA, as illustrated in Figure 1.2.3A [46]. Alternatively, conformational selection proposes that among the dynamically moving KstR dimers, a cohort of those conformations are predisposed to associate with DNA, and will bind to DNA if within proximity, as portrayed in Figure 1.2.3B [36] [47]. Both models indicate a fluctuation between DNA-bound KstR and *apo* KstR. However, the models disagree about KstR transitions between the two states.

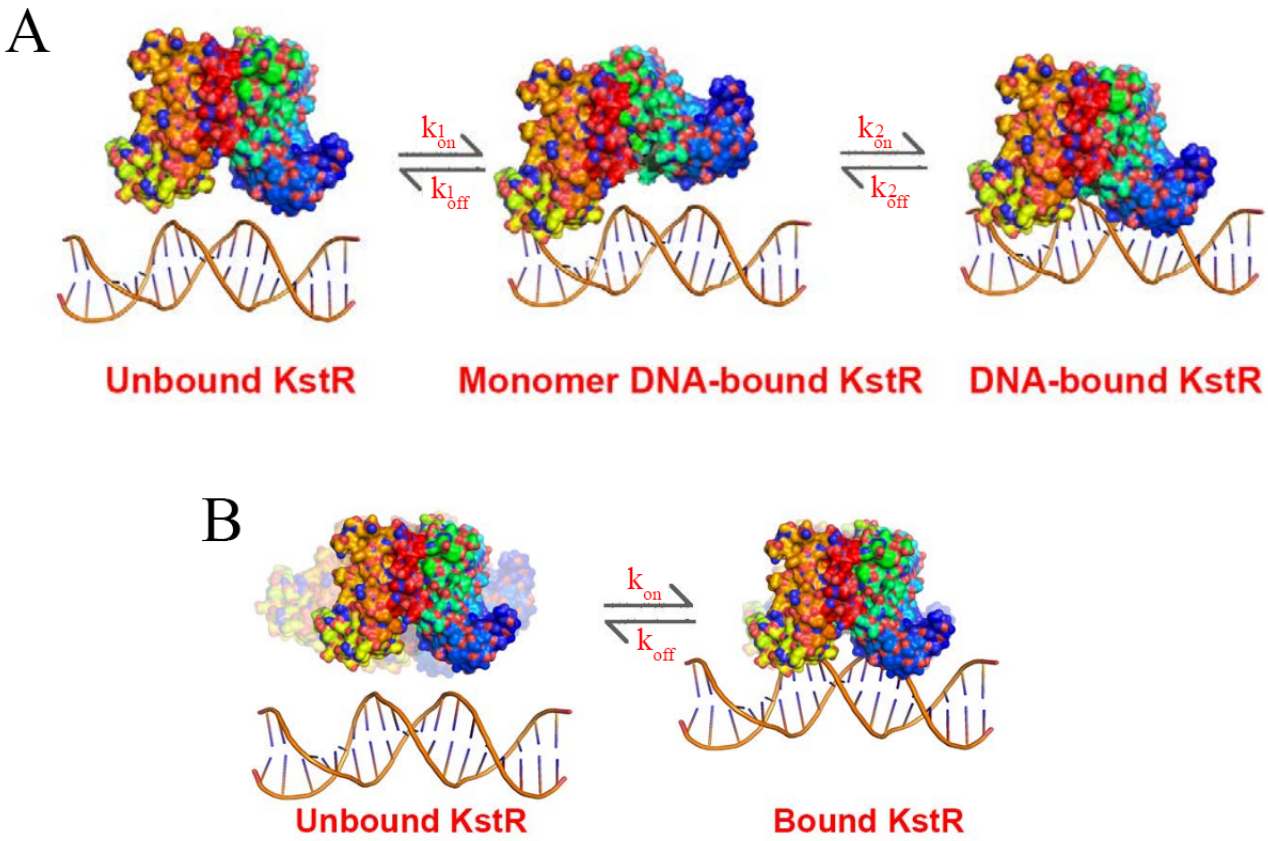

Figure 1.2.3: *Diagram showing different binding models: A, two-step Induced-fit binding model; B, conformational selection binding model (transparent structures indicating different possible conformational states).*

1.2.4 KstR gene control

KstR is a transcriptional repressor that binds to a short palindromic DNA sequence. KstR controls the expression of genes utilised in cholesterol and lipid degradation and uptake through its DNA binding activity [31] [48] [49]. DBD residues which constitute the $\alpha 2$ and $\alpha 3$ helices of KstR bind to a specific DNA sequence, namely [TnnAACnnGTTnnA], a 29 bp motif that is conserved among Mycobacterial species, as displayed in Figure 1.2.4 [3]. Bioinformatic investigations into the location of the motif within the *Mtb* genome established that it is frequently located at promoter regions 10 and 35 bps upstream of genes associated with cholesterol degradation enzymes. Electrophoretic mobility shift assays revealed that KstR bound with specificity to this DNA sequence as a dimeric complex [3] [48].

KstR binding at this motif occludes ribonuclic acid (RNA) polymerase access to the genes' promoter region, thus preventing expression of the gene itself. A KstR knock-out strain of *Mtb* showed that when KstR is absent, the genes downstream of the KstR motif were significantly over expressed compared to expression in wild-type strains. Additionally, KstR was found to regulate approximately 74 genes - many known to be induced by cholesterol and other

lipids [3] [7]. KstR modulates the expression of genes within the KstR regulon, which is comprised of genes expressing fatty acid metabolising and cholesterol metabolising enzymes as well as KstR itself [3] [18] [48].

```
                                   -10    -5          5     10
                                          ←————       ————→
MTB     TCGACTTGACGGTTGATCGTTAGCACTATTAGAACACGTTCTAGTGGGC-AAGAACGCTC 237
MB      TCGACTTGACGGTTGATCGTTAGCACTATTAGAACACGTTCTAGTGGGC-AAGAACGCTC 237
MM      CCAACTTGACG-TTGGTCGTCTGCACTATTAGAACACGTTCTAGTAAGT-ATGAACGGTT 236
MSM     TCGACTTGACGGTCGAACGCTGTCACTATTAGAACACGTTCTAGTGAG--AAGCACGGAC 172
MAP     TCGACTTGACGGTCGATCACTGGCACTATTAGAACACGTTCTAGTCTCCGGTGAGCGAGT 107
NFA     TTGTCTTGACTTCCGCCCGGCAGGGATATTAGAACATGTTCTAG------GTGTGGGAGT 208
            ******    *  *        ********* *******        *   *
```

Figure 1.2.4: *Alignment of KstR intergenic region among Mycobacterial species and other actinomycetes. Conserved inverted palindromic repeat present throughout species in bold, residues conserved in all genomes marked by an asterisk, [TnnAACnnGTTnnA] KstR binding motif outlined in red. Mtb: M. tuberculosis; MB: Mycobacterium bovis; MM: Mycobacterium marinum; MSM: M. smegmatis; MAP: Mycobacterium avium; NFA: Nocadia farinica. Reproduced with permission from Kendall et al [3].*

1.2.5 DNA-bound KstR

When bound to DNA each DBD associates with sequential major grooves in the DNA helix. Contacts between DNA and KstR are supported by hydrogen bonding (H-bonding) networks and van der Waals (VDW) interactions. The $\alpha 1$ helix of the DBD and the N-terminal loop of $\alpha 4$ contact the phosphate backbone at opposite sides of the DNA groove. The side chains of residues Arg-17, Thr-52, and Tyr-56 and backbone nitrogen of Ala-49 form hydrogen bonds with the phosphate groups of C(2) and T(3) on one side of the major groove as displayed in Figure 1.2.5. On the other side of the major groove, an intensive hydrogen network is formed around around phosphate groups C(10), G(11), and T(12) involving the side chains of Gln-38, Tyr-54 and Lys-60. Furthermore, a hydrophobic interaction is seen between side chain of Val-50 and T(12) pyrimidine rings. This binding network produces a dissociation constant K_d of 10.5 $\pm$ 0.6 nm between DNA and KstR. This demonstrates that there is a tight binding between KstR and DNA in the absence of ligand (*Ho et al. manuscript in preparation*).

The distance between the DBDs of a DNA-bound KstR dimer is approximately 32 Å indicating a contracted KstR complex compared to *apo* and ligand-bound KstR states. DNA binding produces changes in the protein conformation, with dislocations among the $\alpha 5$ - $\alpha 7$ bundle, $\alpha 4$ and the DBD. As a result of the contracted conformation, the ligand binding pocket formed by the LBD is contorted so that it is no longer accessible for CoA thioesters. Additionally, the architecture of KstR is warped due to DNA pulling the DBD regions of KstR together. Moreover, His-62 at the N-terminal end of helix $\alpha 4$ is rigidly connected to helices $\alpha 2$ - $\alpha 3$ of the HTH motif, forming H-bonds with

Ser-124 in the helix $\alpha 7$. The H-bonding at Ser-124 acts like a transmitter, pushing the $\alpha 6$ - $\alpha 7$ loop and N-terminal end of helix $\alpha 6$. Aside from contortions at the ligand binding pocket, the structure of helices $\alpha 5$ - $\alpha 9$ is well conserved between the *apo* and DNA-bound KstR states (*Ho et al. manuscript in preparation*).

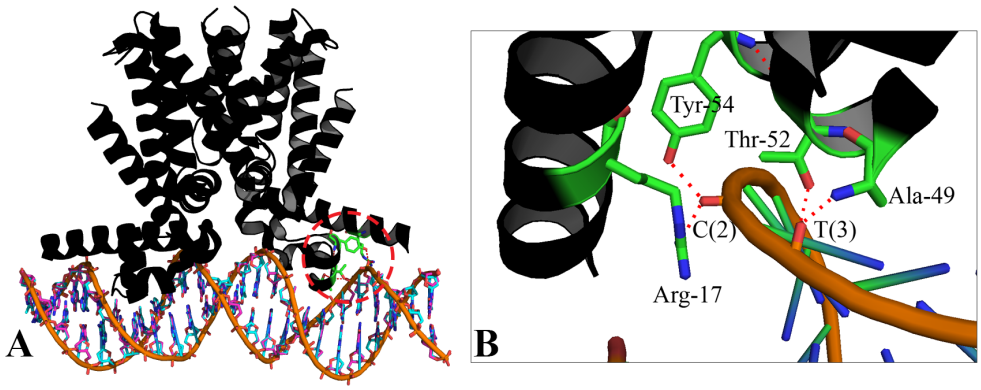

Figure 1.2.5: *Annotated cartoon KstR bound to DNA. A, KstR dimer interacting with DNA. Red dashed circle shows the location of H-bonding network between Arg-17, Thr-52, Tyr-56, and Ala-49 of KstR and C(2) and T(3) of DNA. B, Closer image of the same hydrogen bond network with polar contacts outlined. Red dashes represent polar contacts between residues of KstR and specific sites of the DNA phosphate backbone (Ho et al. manuscript in preparation).*

1.2.6 KstR ligands

Gene modulation of KstR is regulated by the presence of at least two specific metabolites produced early in cholesterol degradation. The two cholesterol metabolites discovered to have the highest affinity to KstR are 3OChA-CoA and 4-BNC-CoA. Of the two CoA thioesters, 3OChA-CoA has the higher affinity ($K_d = 60 \pm 20$ nm) [4].

The enzyme, 3-β-hydroxysteroid dehydrogenase (3-β-HSD), is responsible for the C3 cholesterol oxidation step early in the cholesterol degradation pathway. This produces the CoA thioester, 3OChA-CoA as displayed in Figure 1.2.6 [7] [19]. Although the expression of 3-β-HSD in *Mtb* is repressed by KstR it maintains a low baseline expression, thereby constitutively producing the metabolite 3OChA in the presence of cholesterol [18].

4-BNC-CoA was also proposed as a cholesterol metabolite that can bind to KstR ($K_d = 280 \pm 50$ nm) [4]. 3-ketosteroid 9a-hydroxylase (KshAB) is the enzyme responsible for the production of 4-BNC-CoA in the cholesterol degradation pathway [50]. Despite KstR expressed by *Mtb* repressing KshAB expression, a baseline expression persists in the absence of cholesterol [51]. It is speculated that 3OChA-CoA and 4-BNC-COA interact as ligand metabolites instead of cholesterol itself, as they are more soluble [19].

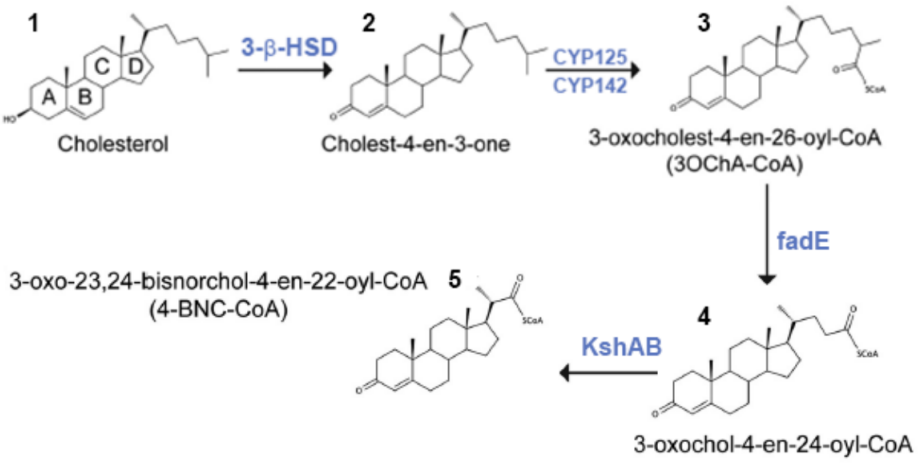

Figure 1.2.6: *Schematic of cholesterol degradation pathway towards 3OChA-CoA and 4-BNC-CoA in Mtb. Reproduced with permission from Hot et al [4].*

1.2.7 Ligand-bound KstR

The ligand binding pocket is lined almost entirely with hydrophobic amino acids and is sandwiched between the helices $\alpha4$, $\alpha5$, $\alpha7$ and $\alpha8$ of the LBD. The entrance of the pocket is flanked by $\alpha4$, $\alpha5$, and sections of the $\alpha9$ helix, organising hydrophobic residues Arg-102, Arg-106, Lys-99, Arg-70, Arg-74, and Lys-78 [4]. The interacting residues are primarily dependent on which ligand binds into the site, but encompass: Leu-68, Gln-71, Phe-72, Iso-75, Leu-100, Ala-103, Met-104, Asp-107, Leu-110, Thr-111, Met-114, Tyr-118, Iso-136, Arg-158, Ser-161, Trp-164, and Leu-168. Critical to binding is Arg-158, which is positioned at the end of the pocket. The ketone oxygen of 3OChA-CoA (PDB: 5CW8) and 4-BNC-CoA (PDB: 5CXI) interacts with Arg-158, as shown in Figure 1.2.7 [4].

The LBD operates as an allosteric binding site - ligand binding induces conformational changes across the entirety of KstR [52] [53]. Changes in conformation include a twist and a translational shift observed at the four-helix bundle composed of $\alpha6$ - $\alpha9$, shifts in the $\alpha5$ - $\alpha7$ helical bundle, the $\alpha4$ helix, and the DBD as a whole. The impetus for the conformational change of KstR is a perturbation in helix $\alpha5$ caused by switch-like flipping of Trp-164 in the hydrophobic pocket due to ligand interaction [4]. The distance between the DBDs is 44 Å as a result of these changes in ligand-bound KstR. Overall, the structure of ligand-bound KstR illustrates the ligands' ability to induce conformational movements at more peripheral regions, causing conformational changes that favour the dissociation of KstR from DNA.

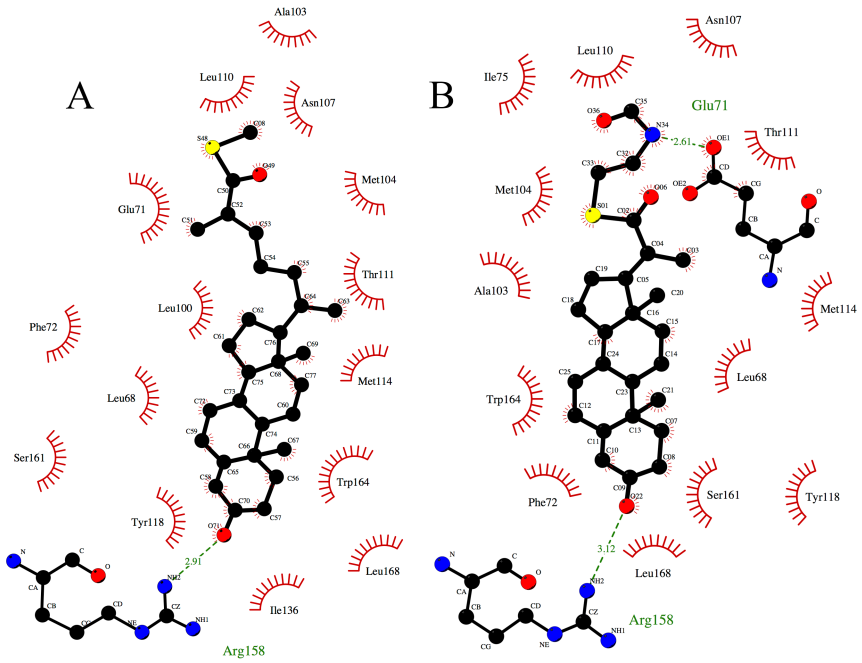

Figure 1.2.7: *Schematic of residues that line the KstR ligand pocket and their interaction with A) 3-OChA-CoA and B) 4BNC-CoA. Red represents hydrophobic contacts with ligand, dashed green line marks hydrogen bonding. Reproduced with permission from Hot et al [4].*

1.2.8 KstR mechanism

Despite the repression of the KstR regulon, a baseline expression of the occluded genes, including those encoding 3-β-HSD and KshAB expression, persists [54]. Therefore, when cholesterol is available to *Mtb* cells, CoA thioesters such as 3OChA-CoA and 4-BNC-CoA can still be produced. Cholesterol metabolites drive changes to the structure of KstR, which dictates the expression of cholesterol degradation genes and thus is a component of *Mtb* virulence. KstR binds to [TnnAACnnGTTnnA] motifs at promoter regions of DNA [18]. The malleability of *apo* KstR permits it to take on different conformational ensembles, allowing the structure to conform and bind to either DNA or ligand.

KstR crystal structures have shown that the distance between the DBDs of *apo* KstR is approximately 38 Å , which is wide enough to enable binding to DNA [4]. When KstR is bound to DNA it prevents RNA polymerase from associating with DNA promoter regions where the KstR motif is present. As a consequence KstR represses the expression of those genes, which are mainly involved in cholesterol degradation. Crystal structures of DNA-bound KstR suggest that it adopts the shortest possible distance between the DBDs of 32 Å . This interaction favours the fastening of KstR to DNA, restricting ligand binding. The crystal structures of early-stage cholesterol metabolites of ligand-bound KstR takes on a wide inter-DBD distance of 44 Å [44]. The presence of ligand biases the conformation of ligand-bound KstR, favouring a conformation with a wide distance between the DBDs [3] [4] [19]. Ligand-bound KstR is excluded from binding with DNA, thereby RNA polymerase can bind

to previously occluded promoter regions of the KstR regulon. This results in gene de-repression and cholesterol degradation, shown in Figure 1.2.8 [4] [18].

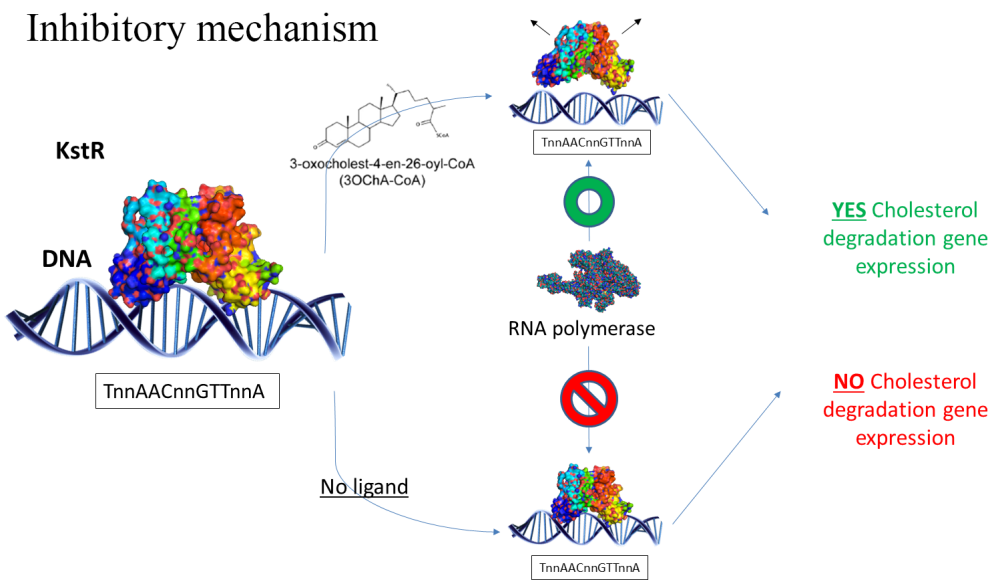

Figure 1.2.8: *Illustration summarising how KstR effects expression of genes in the presence and absence of ligand.*

It is clear that the crystal structures represent keys moments in the mechanism of action of KstR. What is unknown is how KstR transitions between DNA-bound KstR, *apo* KstR, and ligand-bound KstR states. Molecular dynamics (MD) is a technique which harnesses computational simulations to determine the kinetics of a molecular system. The mechanism of transition between the static states of KstR can be investigated by employing MD techniques.

1.3 Computational modelling

1.3.1 Molecular dynamics

MD was founded as a theoretical physics model for solid state systems in the late 1950s [55]. It was then introduced in a computational chemistry context to simulate dynamics of liquids by modelling them as hard spheres. In 1977 it was adapted for modelling biological molecules through the simulation of bovine pancreatic trypsin inhibitor [56]. While the system was limited in scope compared to contemporary standards, possessing 500 atoms and simulated for 9.2 ps, it was of key importance in showing the dynamic nature of proteins. Substantial research and development has produced advances in computer technology and algorithms used to run MD simulations have enhanced the capabilities of MD, growing it into a field in its own right [57]. The breadth to which MD techniques have been utilised has also expanded, integrating into a wider range of sciences [7] [58].

The MD simulation regime can be broken down into two components: a FF set to characterise an atom properties and a simulation engine to simulate the atoms behaviour. A variety of Force fields (FF) are available that represent interactions within different biomolecular system [59] [60] [61]. Popular MD software packages include: CHARMM, AMBER, NAMD, GROMACS [5]. While the FFs share some essential characters they differ through their parameterisation philosophy, sophistication of atomic description, and selection of theoretical or experimental data used to determine parameters.

1.3.2 Force fields

FFs represent a set of conditions that govern how atoms should interact with one another. In a simulation FFs treat atoms, not electrons, therefore bonding information must be provided for atom interactions in lieu of solving the electronic Schrodinger's equation at each step. A FF includes terms for organising atoms into molecules, this is achieved by describing bonds, angles and torsions between atoms [43].

FFs are composed of empirically parameterised sets of functions that describe the energy of a set of atoms dependent on their atomic coordinates. A FF includes bonded, E_{bonded}, and non-bonded, $E_{non-bonded}$, interaction potentials that sum to yield the energy, V, of a set of atoms, r (Eq. 1.3.1).

$$V(r) = E_{bonded} + E_{non-bonded} \qquad \text{(Eq. 1.3.1)}$$

The basis of all FFs is the assignment of atoms to a limited set of atom types by grouping those that are chemically and physically similar. Classical FFs have many different atom types for the same element. As a condition, over the course of a simulation an atom cannot change type. Consequently, the use of atom types is a compromise between having the ability to accurately describe each atom and the practical restriction of the number of atom types. Widely used biochemical FFs, including CHARMM, AMBER, GROMOS and OPLS, comprise similar potential functions, although their parameterisation may vary [62].

1.3.3 Two body harmonic bond stretching potential

The potential energy of bond stretching between two atoms is approximated by a simple harmonic oscillator, described by the following equation (Eq. 1.3.2):

$$U_b(r) = \frac{1}{2}k_b(r - r_0)^2 \qquad \text{(Eq. 1.3.2)}$$

where k_b represents a force constant between two atoms, r is the bond length and r_0 is the ideal bond length so that $(r - r_0)$ computes the deviation of the bond length from the ideal value. The contribution to the force on particle i from bond stretching as formulated in Eq. 1.3.3.

$$F_i(r) = -k_b(r - r_0) \qquad \text{(Eq. 1.3.3)}$$

1.3.4 Three body harmonic angle potential

A harmonic potential is also implemented when approximating the energy potential generated by angular vibrational motions which occur between three atoms (i, j, k) as described by the equation (Eq. 1.3.4) below:

$$U_a(\theta) = \frac{1}{2}k_a(\theta - \theta_0)^2 \qquad \text{(Eq. 1.3.4)}$$

In this equation k_a represents the harmonic spring constant, the current angle is θ, and θ_0 is an idealised angle value such that potential energy increases as the current angle deviates from the ideal value. The contribution to the force on each particle is given by the equations (Eq. 1.3.5/6/7):

$$F_i = -\frac{\delta U_a(\theta_{ijk})}{\delta r_i} \qquad \text{(Eq. 1.3.5)}$$

$$F_k = -\frac{\delta U_a(\theta_{ljk})}{\delta r_k} \qquad \text{(Eq. 1.3.6)}$$

$$F_j = -F_i - F_k \qquad \text{(Eq. 1.3.7)}$$

For CHARMM FFs the angle potential is extended to include the Urey-Bradley potential [63]. This second term describes a covalent spring between the outer atoms (i and k), where r_{ik} gives the distance between two atoms and r_{UB} is the equilibrium distance, and k_{UB} is a force constant which activates Urey-Bradley term, detailed in Eq. 1.3.8.

$$U_a(\theta) = \frac{1}{2}\epsilon_{angles}k_a(\theta - \theta_0)^2 + k_{UB}(r_{ik} - r_{UB})^2 \qquad \text{(Eq. 1.3.8)}$$

1.3.5 Four body improper dihedral angle potential

Improper dihedral angles ensure carbon rings and other planar groups remain planar and the ring structures do not excessively pucker or flip. Four atoms are used to define an dihedral angle and maintaining chirality in united-atom force fields. The improper dihedral terms are implemented using a harmonic potential (Eq. 1.3.9) as shown below:

$$U_{id}(\phi_{ijkl}) = \frac{1}{2}k_\phi(\phi - \phi_0)^2 \qquad \text{(Eq. 1.3.9)}$$

where ϕ is the dihedral angle and ϕ_0 is the equilibrium dihedral angle between four atoms i, j, k, and l.

1.3.6 Four body proper dihedral angle potential

For the torsional potential a periodic function improves the potential energy calculation as harmonic descriptions are periodic. Multiple terms can be applied to each torsion generating a Fourier series which better matches the true dihedral potential. Each dihedral is expressed in Eq. 1.3.10.

$$U_d(\phi_{ijkl}) = \epsilon\frac{1}{2}k_\phi[1 + cos(n\theta - \delta)] \qquad \text{(Eq. 1.3.10)}$$

The k_ϕ is a force constant proportional to the barrier to rotation, n is the periodicity indicating the number of minima in the function, while δ is the phase shift that determines which torsional angle ϕ corresponds to a minima.

1.3.7 Non-bonded terms

The contribution to the potential energy from interactions between non-bonded atoms is divided into VDW and electrostatic interactions [5]. Calculation of the non-bonded interaction takes up the lion's share of computational time due calculations for the distance between all pairs of atoms. In practice this is reduced in several ways [60]. The short distances between 1-2 and 1-3 bonded atom pairs cause very large non-bonded interaction terms between these pairs and are therefore excluded from the final energy term. The 1-4 non-bonded atom interactions can be be scaled down [5]. Finally, at large distances, the VDW between two atoms converge towards zero while long-range electrostatics tail off slower. This is exploited by only calculating non-bonded interactions out to some cut-off distance, which improves the computational efficiency [62].

1.3.8 Van der Waals interactions

VDWs interactions are composed of a repulsion and an attraction term that depend on the distance between the atoms. In CHARMM this is described by a simple Lennard-Jones (LJ) potential [59]. The general form of the LJ potential (Eq. 1.3.11) is expressed below:

$$U_{LJ}(\sigma_{ij}) = 4\varepsilon((\frac{\sigma_{ij}^{\theta_{ij}}}{\sigma_{ij}})^{12} - (\frac{\sigma_{ij}^{\theta_{ij}}}{\sigma_{ij}})^{6}) \qquad \text{(Eq. 1.3.11)}$$

where ε is the energy well depth and σ_{ij}^{eq} is the inter-atomic separation where the energy is smallest. the short-range repulsions are accounted for by the σ^{-12} terms at short distances, and the London dispersion-attraction terms are handled by the σ^{-6}. At short distances the repulsive term dominates thus preventing superposition of atoms. The LJ potential goes to zero as σ_{ij} increases. The cut-off distance is typically chosen to be a distance at which it is close to zero. The participating forces in relation to the distance between particle i and j are obtained by differentiating Eq. 1.3.11. The force on an atom i is given by Eq. 1.3.12,

$$F_i(\sigma_{ij}) = E(12(\frac{\sigma_{ij}}{R})^{11} - 12(\frac{\sigma_{ij}}{R})^{5}) \qquad \text{(Eq. 1.3.12)}$$

where σ_{ij} is a constant in the LJ potential.

Electrons are not explicitly present in atomic-level models. To account for the different electronegativities of atoms they are given partial i.e. fractional, charges [64]. The pairwise electrostatic interaction is a determined by the Coulomb's law as formulated below (Eq. 1.3.13):

$$u(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \qquad \text{(Eq. 1.3.13)}$$

where for the two atoms i and j separated by distance r_{ij}. The partial charges are given by q_i and q_j, and ϵ_0 is the free space permitivity.

Long range electrostatic interactions are calculated using schemes such as reaction-field, Ewald summation, and Particle-mesh Ewald (PME) summation [65]. PME is typically used with CHARMM FFs.

1.3.9 Bonded terms

Gas phase molecular structures obtained from microwave and electron spectroscopy data, and crystal structures from X-ray data, provide ideal values for bond and angle terms [66]. Force constant parameter values come from fitting to vibrational data generally sourced from gas-phase infrared and Raman spectroscopes with normal mode calculations although solution or crystal data can be used if needed [67]. Quantum mechanical calculations provide additional means of generating reference data.

Bond stretch, bond angle, dihedral and improper dihedral terms define a molecule's geometry as formulated in Eq. 1.3.14.

$$U_{bonded,i} = U_b(r_{ij}) + U_a(\theta_{ijk}) + U_{id}(\phi_{ijkl}) + U_d(\theta_{ijkl}) \qquad \text{(Eq. 1.3.14)}$$

The first term $U_b(r_{ij})$ provides the potential energy penalty for stretching the bond between two atoms i and j. The second term $U_a(\theta_{ijk})$ provides the potential energy function produced by angle bending between three atoms (i, j and k). The third term $U_{id}(\phi_{ijkl})$ provides the potential energy function for maintaining the chirality of tetrahedral carbon atoms (in united atom force fields) and planar structure of carbon rings and peptide bonds. The last potential energy $U_d(\theta_{ijkl})$ is associated with rotations about bonds [5] [68].

1.3.10 Propagation

A MD simulation evolves by solving Newton's equation of motion (Eq. 1.3.15) for a system of N interacting particles or atoms i.e. $i = 1, 2...N$:

$$F_i = m_i a_i = m\frac{\delta^2 r_i}{\delta t^2} \qquad \text{(Eq. 1.3.15)}$$

where F_i is the force on atom i, m_i is its mass, a_i its acceleration, and r_i its coordinates. Newton's equations of motion calculates the change in an atoms position in the system as a function of time. For a given particle or atom Newton's equation of motion is integrated in small time steps to ensure conservation of energy. The force on each atom is equal to the negative derivative of the potential energy function as seen below (Eq. 1.3.16):

$$F_i = -\frac{\delta U}{\delta r_i} \qquad \text{(Eq. 1.3.16)}$$

where the total potential energy, U, is the sum of contributions to the potential energy from bonded and non-bonded interactions. The potential energy and therefore the force on an atom depends on its position in relation to other atoms in the system, as defined by the FF. The combination of equation

1.3.15 and 1.3.16 results in the following equation (Eq. 1.3.17) that relates change in the position of an atom with a derivative of the potential energy.

$$-\frac{\delta U}{\delta r_i} = m_i \frac{\delta^2 r_i}{\delta t^2} \qquad \text{(Eq. 1.3.17)}$$

Given possession of initial coordinates and their velocities, the and the new position of an atom can be determined for the subsequent time step through integration of Newton's second equation of motion. Different numerical time integration algorithms have been developed for integration of the equations of motion with GROMACS utilising the leap-frog algorithm.

The leap-frog algorithm uses atom i at position r, r_i, at time, t, to update the positions and velocities, v_i of an atom using the force $F_i(t)$ as shown in Eq. 1.3.18).

$$-r(t+\Delta) = 2r(t) - r(t) + \Delta t v(t + \frac{1}{2}\Delta t) \qquad \text{(Eq. 1.3.18)}$$

The velocity v is not directly generated which is necessary for computing the kinetic and total energy of the system, and thereby is calculated with the following equation (1.3.19) at time t.

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} F(t) \qquad \text{(Eq. 1.3.19)}$$

Leap-frog produces trajectories whose update-relation is formulated in Eq. 1.3.20.

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \frac{1}{m} F(t)\Delta t^2 + O(\Delta t^4) \qquad \text{(Eq. 1.3.20)}$$

The main characters of a time integration algorithm for use in MD simulations are that it's computationally efficient, preserves the momentum of an atom and allows adoption of large time steps. Additionally, it is important for the time integration algorithm to aspire to calculating true trajectories as precisely as possible, however sources of errors, such as round errors and truncation of the Taylor expansions persist [69].

1.3.11 Spatial Boundary conditions

There are three general approaches to managing spatial boundaries: vacuum boundary conditions, fixed boundary conditions, and periodic boundary conditions. For periodic boundary conditions, the molecule is placed in a periodic box such that their are replicate images of the system in all directions [5]. In practice an atom that "leaves" one side of the box "reappears" on the opposite side of the box. Periodic boundary conditions are therefore preferable (Figure 1.3.1). This allows for the finite system to have similar behaviour of a infinite system.

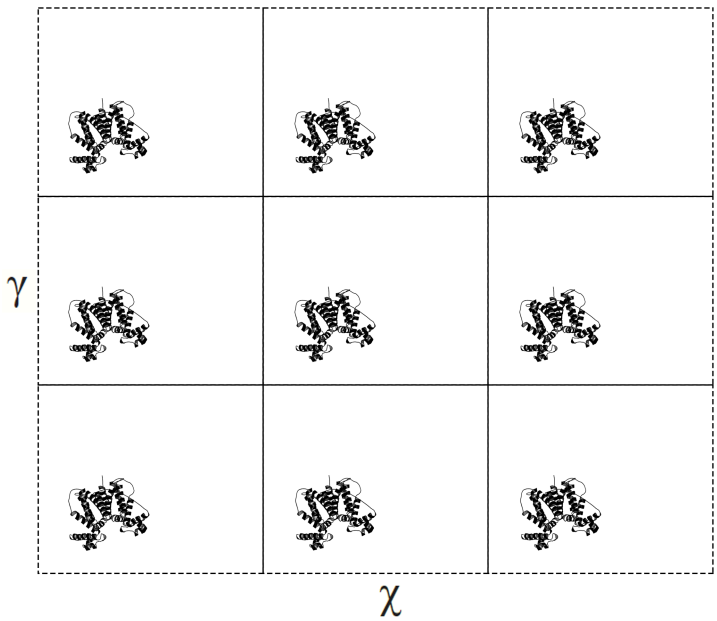

Figure 1.3.1: *Periodic boundary conditions in two dimensions [5].*

1.3.12 Geometric boundary conditions

Geometric boundary conditions refer to the constraints and restraints on designated elements of a system. Constraints are applied to keep specified properties fixed throughout a simulation. Restraints differ from constraints in that they allow for some motion by assigning a harmonic restraining force. Constraints are often used to fix bond length and angles between two atoms. A number of different algorithms can be used to introduce constraints: SHAKE, LINCS RATTLE are commonly used constraints [5] [70].

Restraints are often used in the preparation of a simulation. As the subject is being heated the properties of the system need to be restrained as velocities are assigned in random directions to start the dynamics. These velocities may lead two bonded atoms to want to move in opposite directions straining the bond, or some pair of atoms to move towards one another and clash. To avoid the large forces that would come with uncorrelated initial velocities, velocities are assigned at low temperature so if the directions are uncorrelated, the velocities are small enough that atoms hardly move. As the dynamics continues the force fields initiate and become correlated. Fundamentally, restraints are used to avoid the described uncorrelated movements which result in warping of molecules before the beginning of an unrestrained simulation.

1.3.13 Thermodynamic boundary conditions

Thermodynamic boundary conditions refers to the thermodynamic ensemble in which the simulation is conducted. In MD, the NVE (constant number of particles, constant volume, constant energy) ensemble is the natural ensemble, but NVT (constant temperature) and NPT (same but pressure instead of volume) correspond better to experimental or physiological conditions. These

ensembles are achieved by using thermostats and barostats to maintain the temperature and pressure, respectively.

Canonical ensemble:
Canonical ensemble (NVT), indicates that the amount of substance N, volume V, and temperature T are conserved for the system. In a system under NVT conditions the temperature can be calculated when N_f equals the number of degrees of freedom and k_b is Boltzmann constant in the equation below (Eq. 1.3.21):

$$r(\Delta t) = \Sigma_{i=1}^{N} \frac{m_i v_i^2(\Delta t)}{k_b N_f} \qquad \text{(Eq. 1.3.21)}$$

where Δt is the change of temperature. The thermostat alters the energy of a system to maintain an average temperature [71]. There have been a number of different thermostats developed however here the velocity-rescaling thermostat was used [72].

Velocity-rescaling thermostat:
Thermostats are applied to regulate a system's temperature over the course of a simulation. Modifications to the temperature are typically introduced through altering the velocities of atoms. Velocities are introduced to approximate the kinetic energy and temperature of a heated system. The velocity-rescaler thermostat employs a stochastic term to ensure a canonical ensemble and has similar weak coupling with first order kinetics to an external heat bath with a reference temperature [72].

Isothermal-isobaric ensemble:
Isothermal-isobaric ensemble (NPT) applies a constant pressure to a simulation system. Number of particles N, pressure P, and temperature T are kept constant. Barostats perform a similar function to thermostats, but regulate a system's pressure [73]. The Berendsen barostat was utilised for these simulations. The Berendsen barostat rescales the coordinates and box every time step according to a first-order kinetic relaxation.

1.3.14 Free Energy

Ultimately it is the free energy that determines the likelihood of any given conformation or arrangement of molecules to occur [69]. For a dynamic system kept at a constant particle number, volume and temperature, the probability of finding the system in some state i is given by (Eq. 1.3.22):

$$P_i = \frac{exp^{-\beta E_i}}{Z} \qquad \text{(Eq. 1.3.22)}$$

where $\beta = \frac{1}{k_B T}$, E_i is the potential energy of state i, and $Z = \Sigma_i exp^{-\beta E_i}$ is known as the partition function. From this, the average internal energy of the system, U, can be calculated as (Eq. 1.3.23):

$$U = \Sigma_i P_i E_i \qquad \text{(Eq. 1.3.23)}$$

and the entropy for the system, S, can be determined as (Eq. 1.3.24):

$$S = k_B \Sigma_i P_i log P_i \qquad \text{(Eq. 1.3.24)}$$

and the Helmholtz free energy, A, is defined below (Eq. 1.3.25).

$$A = U - TS \qquad \text{(Eq. 1.3.25)}$$

By substituting Eq. 1.3.22 to Eq. 1.3.24 into Eq. 1.3.25 it can be shown that the Gelmhotlz free energy is equivalent to Eq. 1.3.26

$$A = -kTlogZ \qquad \text{(Eq. 1.3.26)}$$

As such, if the partition function for a system can be determined, the free energy of that system can be calculated. MD provides a means to generate samples of system configurations which, given sufficient sampling, can be used to directly calculate the free energy.

1.4 Objectives

The structure and function of KstR has been researched thoroughly. However, how the protein transitions between *apo*, ligand-bound and DNA-bound forms is unknown. By employing MD techniques the details of the mechanism of action can be revealed through investigating the questions listed below.

1. Do ligand-bound KstR, *apo* KstR, and DNA-bound KstR have distinctive conformational dynamics?
2. Are the conformational dynamics of KstR affected by the removal of ligand and DNA?
3. Are the conformational dynamics of KstR affected by the size of ligand or DNA bound to it?
4. Which regions of KstR are critical for conformational rearrangement?

Chapter 2

Methodology

2.1 MD simulation procedure

All MD simulations were prepared, performed and analysed using GROMACS 5.1.2 MD package, Pymol 1.5, Chimera 1.12, MODELLER 9.18, and CONAN [74] [75] [76]. The protein, DNA, and ligand parameters were determined by CHARMM36, CHARMM27, and the online CGENFF database respectively [59] [77] [78]. CHARMM36 could not be used for the DNA simulations as the DNA parameters had not been updated since version 27.

For all simulations TIP3P was selected as the water model, leap-frog as the time integrator, and LINCS as the constraint algorithm [66] [79]. Verlet was the cutoff scheme employed with a short range cut off of 1.2 nm and the long-range electrostatics treated using the PME method [80] [81]. The temperature was maintained using the velocity-rescaling thermostat, while pressure was maintained using the Berendsen barostat [71] [81].

The MD process can be broken into five stages: 1) Coordinates, 2) Parameters, 3) Solvation, 4) Heating and equilibration, and 5) Simulation production.

2.1.1 Coordinates

Initialisation of an MD simulation requires atomic coordinates in Cartesian (x, y, z) format and atomic velocities. As folding of large peptides such as proteins is not generally feasible given current computational power, an MD simulation of a protein requires provision of initial atomic coordinates. These coordinates must first be energy minimised as the FF used during structure determination is likely to favour different structures to the ones used for the simulations. This has the added benefit of removing any steric clashes.

All initial coordinates aside from 3OChA-CoA, which was built by Thu Ho in the Shaun Lott lab, used x-ray structures from the PDB (3MNL, 5CW8, and 5UA2) to locate the initial atomic coordinates for each system. Ordered waters or agents which were not protein, ligand of interest, or DNA were removed from the PDB structure.

The PDB database KstR structures had residues which were not modelled, typically between the $\alpha 4$ - $\alpha 5$ flexible linker domain and several residues at the N and C-terminal regions of each chain. These residues were omitted from the PDB structure because there was insufficient electron density to accurately resolve their positions. Where residues in a chain were left unmodelled in the PDB structure MODELLER was used to predict their placement in the protein chain. MODELLER is a computational software package which determines the position of missing residues based on chemical knowledge (FF) and, if requested other restraints such as secondary structure or based on a template structure to determine the position of missing residues; along a chain as shown in Figure 2.1.1. MODELLER offers several different positions for the missing residues; along the chain, the most energetically favourable coordinates was selected in every case [76].

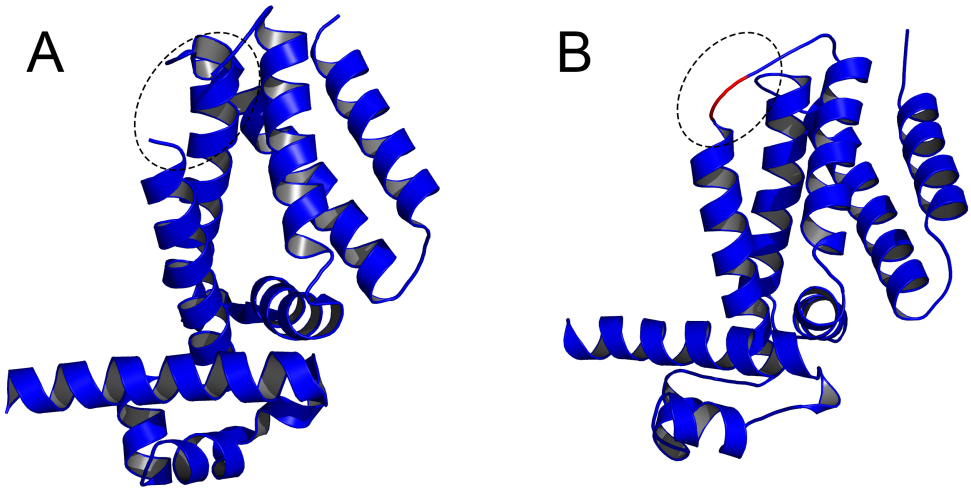

Figure 2.1.1: *Cartoon comparing A) KstR with broken chain to B) KstR with loop region (red) determined by MODELLER circled.*

Different KstR crystals were expressed from different operon sequences. Crystal structure 3MNL was expressed from a different sequence than crystal structures 5CW8, and 5UA2 [4](*Ho et al. manuscript in preparation*). These structures vary in terms of the number of amino acids and the sequence, which would complicate comparison of the simulations therefore the same sequence used for all simulations. Figure 2.1.2 shows the 185 residues used for every simulation, from residue 11 (N-Ser11-Ala12-Glu13...) to 195 (...Iso193-Gly194-Asp195-C). This sequence was selected as this was the largest experimentally validated chain.

```
11                  SEAQRE RRKRILDATM AIASKGGYEA VQMRAVADRA

51   DVAVGTLYRY FPSKVHLLVS ALGREFSRID AKTDRSAVAG ATPFQRLNFM

101  VGKLNRAMQR NPLLTEAMTR AYVFADASAA SEVDQVEKLI DSMFARAMAN

151  GEPTEDQYHI ARVISDVWLS NLLAWLTRRA SATDVSKRLD LAVRLLIGD
```

Figure 2.1.2: *AA sequence used for KstR simulations.*

The 26 base pair (bp) DNA sequence from the DNA-bound KstR crystal structure (PDB: 5UA2) was maintained. For the ligand-bound KstR structure (PDB: 5CW8) the Acetyl-CoA tail was not modelled as part of 3OChA-CoA due to a insufficient electron density. The complete ligand structure was acquired from Thu Ho who produced the ligand-bound KstR x-ray crystal structure [4]. The 3OChA ligand from the crystal structure (PDB: 5CW8) was used to orientate 3OChA-CoA by superimposing the two ligands steroid ring atoms in Chimera. Missing hydrogens were built using Chimera and Pymol [74].

Chimera and Pymol were used to carefully examine the structures for any poorly assigned or missing atoms, mistakes in bond assignment, residue sequence and length, DNA shape and sequence, steric clashing, and misplaced hydrogens.

2.1.2 Parameters

Table 2.1.1 shows the parameters for each KstR system.

Parameters for ligands, 3OChA and 3OChA-CoA, were not available in the CHARMM27 or CHARMM36 FFs. Ligand parameters were generated using CGENFF, a server that determines the parameters of a submitted PDB structure compatible with CHARMM force field by comparison to already parameterised molecules [59] [77]. Water was modelled using TIP3P [66].

Table 2.1.1: *Composition of each KstR MD system. C* in the Force Fields column abbreviates CHARMM*

MD system breakdown							
KstR System	Force fields	Protein atoms	Ligand atoms	DNA atoms	Water atoms	Ion atoms	Size (nm^3)
Full ligand-bound (FLB)	C*36, CGENFF	5,824	292 (3OChA-CoA)	0	162,288	4 CL	11.969
Short ligand-bound (SLB)	C*36, CGENFF	5,824	150 (3OChA)	0	95,466	10 CL	10.102
From ligand-bound (FrFLB)	C*36	5,824	0	0	95,625	10 CL	10.103
Apo	C*36	5822	0	0	116,5541	11 CL	10.742
From DNA-bound (FrLDB)	C*36	5,825	0	0	82,866	11 CL	9.645
Shortened DNA-bound (SLDB)	C*27	5,824	0	1,144 (18 bp)	170,304	24 NA	12.186
Long DNA-bound (LoDB)	C*27	5,824	0	1,524 (26 bp)	177,549	36 NA	12.343
Q38A	C*27	5,804	0	1,524 (26 bp)	130,734	36 NA	11.343
R40A	C*27	5,790	0	1,524 (26 bp)	130,737	38 NA	11.343
Y54A	C*27	5,796	0	1,524 (26 bp)	130,758	36 NA	11.343
Q38A/R40A	C*27	5,770	0	1,524 (26 bp)	130,734	38 NA	11.343
Q38A/Y54A	C*27	5,776	0	1,524 (26 bp)	130,752	36 NA	11.343
R40A/Y54A	C*27	5,762	0	1,524 (26 bp)	130,752	38 NA	11.343
Q38A/R40A/Y54A	C*27	5,742	0	1,524 (26 bp)	130,755	38 NA	11.343

2.1.3 Solvation

The initial coordinates of the protein and ligand and DNA (where appropriate) were energy minimised using the steepest descent minimisation algorithm for 0.1 ns until the force converged to within 1000 kJ/mol/nm.

The box was defined having a minimum 1 nm between the molecules of interest and the edge of the box (to ensure periodic copies of the protein do not interact). The box was then solvated by adding multiple translated copies of a pre-equilibrated box of water. Any molecules that clash with the solute were removed. The system was then energy-minimised again with the same settings, to remove any steric clashes. Ions were then placed within the solvent as to neutralise the system as described in Table 2.1.1. An example of a final system is presented in Figure 2.1.3.

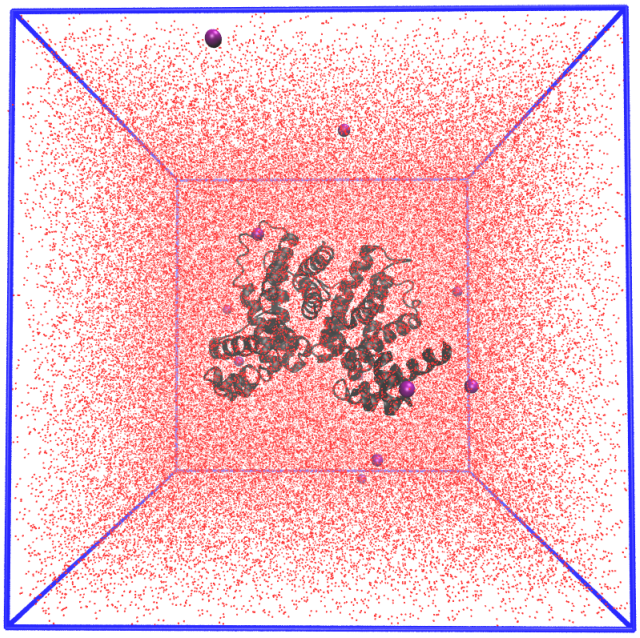

Figure 2.1.3: *Cartoon showing solvated neutral box environment of apo KstR. KstR in centre (black) with Cl^- (purple spheres), and water molecules filling the box (red). The cube represents the boundaries of the box (blue).*

2.1.4 Heating and equilibration

The system was heated from 60 K to 300 K over 3 ns in the NVT ensemble. The temperature was maintained using the velocity-rescale algorithm with τ_T of 0.1 ns^{-1} [81]. The system was then equilibrated in the NPT ensemble for 20 ns, with the pressure maintained at 1 bar using the Berendsen barostat and a τ_P of 0.5 ns^{-1} and an isothermal compressibility of 4.5 x 10^{-5}.

2.1.5 Simulation production

After heating each system was simulated at 20 ns intervals. The configuration from the NPT file was introduced at each interval. The mutation systems were simulated for 100 ns while all other systems were simulated for 200 ns. For all systems the coordinates, velocities, and energies were saved to file at 20 ps intervals.

2.2 Analysis

A variety of different computational tools from different packages were employed for the analysis of the simulations. All measurements are of KstR alone, where DNA or ligand is bound the DNA or ligand contributions have not been included in the measurement.

2.2.1 Radius of gyration

The radius of gyration (Rg) was calculated using the GROMACS gmx gyrate program, gmx gyrate. This calculates the root mean square distance of a molecules regions from its centre of mass. The Rg is calculated through the equation:

$$Rg = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(r_i - r_{mean})^2} \qquad \text{(Eq. 2.2.1)}$$

where N is the number of atoms, r_{mean} is the centre of mass of the molecule and r_i is the coordinates of atom i.

The Rg of the LBD and DBD of KstR were calculated separately by calculating the Rg for the atoms in α4 - α9 (LBD) and α1 - α3 (DBD).

2.2.2 Solvent accessible surface area

The solvent accessible surface area (SASA) was calculated using the GROMACS gmx sasa program. The SASA is approximated by rolling a 0.14 nm radius probe over successive 2D slices of the protein structures, where 0.14 nm is the radius of a water molecule if it were spherical and the slices are made at 0.24 nm increments [82] (Figure 2.2.1).

The volume was also calculated by passing the "-tv" flag to gmx sasa. Volume is another metric which can be used to determine the size of a molecule. The volume is calculated in the same way as the SASA however across the inside of the protein VDW shell and is thereby an approximation of volume [5].

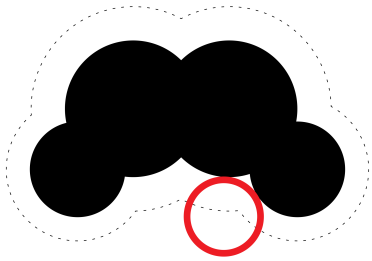

Figure 2.2.1: *Cartoon representation of how volume and SASA are determined. The red circle conveys the probe which calculates the SASA of a 2D slice of the molecule by following the black dashed lines. The areas of the slices are then integrated to give the SASA of the entire molecule in 3D.*

2.2.3 Energy

The enthalpic contributions or potential energy between groups of atoms was calculated using the GROMACS gmx energy program, gmx energy. The Coulombic and VDW contributions were summed to give the enthalpy between selected groups of atoms.

2.2.4 Cluster

KstR conformations were grouped together using the GROMACS gmx cluster program, gmx cluster. Different KstR trajectories were concatenated to see if conformations from independent simulations would group into the same clusters. The method of cluster determination was GROMOS and a root mean square deviation (RMSD) cut-off of 0.2 nm was used to determine if two structures were similar or different [83]. Each cluster procedure produced a unique number of clusters, the criteria for the number of clusters to display varied depending on the systems involved.

2.2.5 Root mean square fluctuation

The root mean square fluctuation (RMSF) for specified atoms over the course of the simulation was calculated using the GROMACS gmx rmsf program, gmx rmsf. To limit the effects of side chain contributions to the RMSF it was only calculated for the backbone (Cα, C and N) atoms of the protein using the equation below (Eq. 2.2.2).

$$RMSF = \frac{1}{N}\sum_{i=1}^{N}(x_i(t) - x_i(0))^2 \qquad \text{(Eq. 2.2.2)}$$

Where N is the number of particles to be averaged, $x_i(0) = x_0$ is the reference position of each particle (i), $x_i(t)$ is the position of each particle (i) at time t.

2.2.6 Define secondary structure protein

The secondary structure assignment of residues across the simulation is called within GROMACS using gmx do dssp. The Define Secondary Structure of Protein (DSSP) algorithm assigns residues a secondary structure determined by the hydrogen bonding relationship between residues and torsion about the Cα atoms of the protein backbone [84] [85].

2.2.7 Residue distance

VMD offers a time correlated distance analysis by producing vectors between selected atom pairs. Distances were measured between the Cα atom of residues: Ala-17, Gln-38, Arg-40, Arg-45, Val-50, Tyr-54, Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 as the point from which distances were measured, as shown in Figure 2.2.2 and 2.2.3. These residues were selected as they were highlighted as important residues interactions with DNA or ligand in previous KstR studies [4](*Ho et al. manuscript in preparation*).

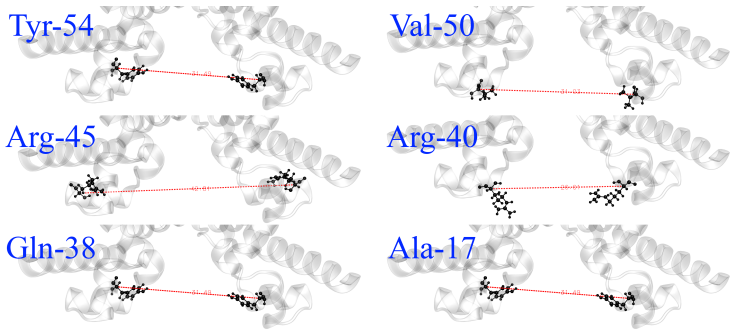

Figure 2.2.2: *Annotated cartoon representation of KstR DBD demonstrating how the distances are measured for selected residues.*

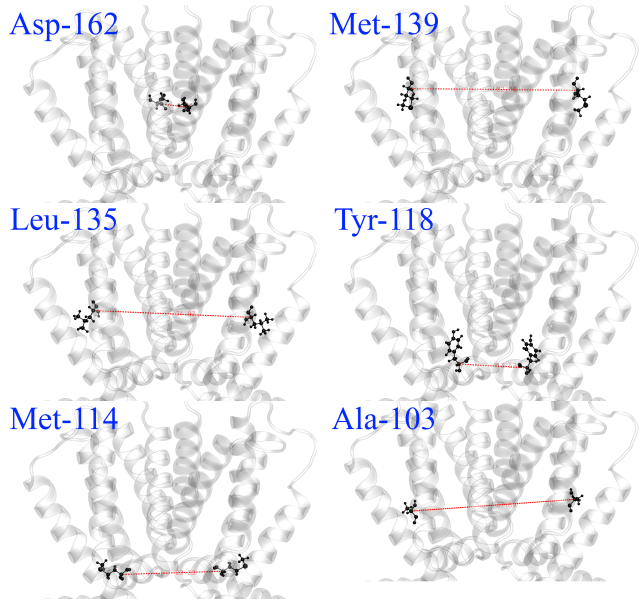

Figure 2.2.3: *Annotated cartoon representation of KstR LBD demonstrating how the distances are measured for selected residues.*

2.2.8 Dihedral angles

VMD was used to calculate the dihedral angles within residues of KstR over the course of the simulation. VMD demands four points to produce a dihedral angle, two points for each plane. The atoms selected to calculate the dihedral angle for Tyr-54 and Trp-164 is shown in Figure 2.2.4. The dihedral angle was calculated using the formula below (Eq. 2.2.3):

$$\theta = \cos^{-1}\left(\frac{AE + BF + CG}{\sqrt{A^2 + B^2 + C^2}\sqrt{E^2 + F^2 G^2}}\right) \qquad \text{(Eq. 2.2.3)}$$

where A, B, C are the dimensions of the first plane, while D, E and F are the dimensions of the second plane.

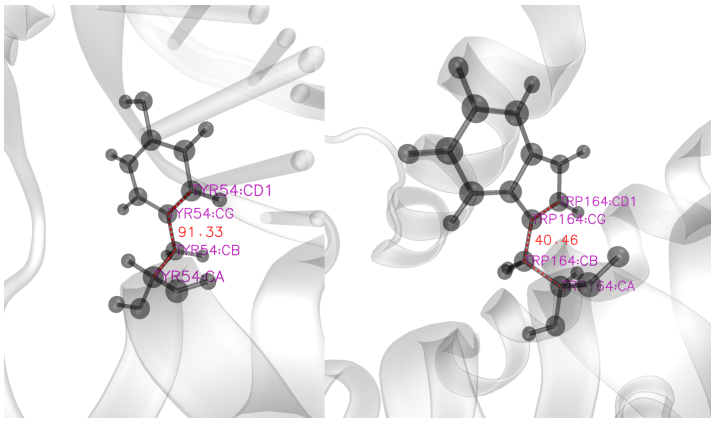

Figure 2.2.4: *Annotated cartoon representation of demonstrating how the dihedral angle between helical domains is measured.*

2.2.9 Contact map

CONAN is a tool which provides information about how inter-atomic contacts evolve over time. The Average MDMAT contact map averaged over 200 ns, displaying contacts between residues within a cut-off of 1.6 nm.

2.2.10 Ligand binding cavity volume

The trj cavity software was used to calculate the ligand binding cavity volume. This program performs a grid-based cavity search across a molecule to identify cavities [86]. Atoms are organised into a 3D grid with their corresponding VDW radius. Once atoms are positioned the cavity search of the atom-mapped grid is pursued. The process of cavity identification is incremental, once a voxel has been assigned as part of a cavity the algorithm searches for potential cavity members to define the cavity, proceeding in a iterative fashion, terminating when no more members in a certain direction are found. Once a cavity is defined its volume is measured over the course of the simulation. For KstR the search region was restricted to the LBD, after which the only cavity identified was the ligand binding cavity. What trj cavity measures to quantify the volume of the ligand binding cavity is shown in Figure 2.2.5.

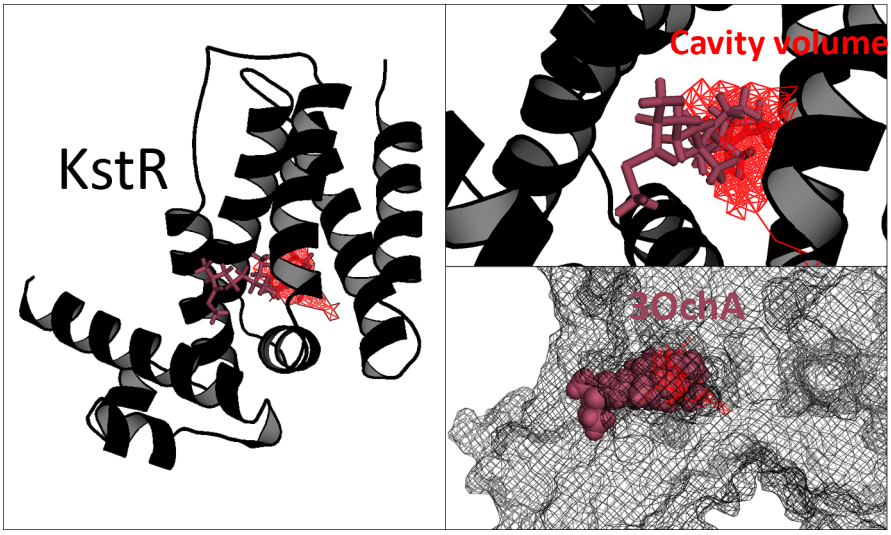

Figure 2.2.5: *Annotated cartoon representation of KstR (black) and ligand (maroon) with cavity volume determined by trj cavity (red mesh).*

2.2.11 Statistics

Average ($\bar{x}$), standard deviation (σ), standard error (σ_x) and range were produced for Rg, SASA, volume, distance, and dihedral measurements using a local script. In each instance the first 10 ns of the simulation were not included in the calculation as they are part of the equilibration.

Statistics for secondary structure composition were also calculated using a local script and did not take the first 10 ns into account. The calculations for energy statistics were produced as part of the energy program.

Chapter 3

Results

3.1 DNA or ligand binding drive KstR to take on distinctive conformational ensembles

To compare the conformational dynamics between ligand-bound, *apo*, and DNA-bound KstR, three systems were prepared. The Full Ligand bound (FLB) system is comprised of 3OChA-CoA in the ligand binding cavity of KstR and represents ligand-bound KstR. The Long DNA bound (LoDB) system is comprised of 26 bp DNA bound to KstR (as seen in crystal structure PDB: 5UA2) and represents DNA-bound KstR. *Apo* KstR is KstR with neither ligand or DNA bound (as seen in crystal structure PDB: 3MNL) to it.

The FLB, *apo*, and LoDB KstR were then solvated as described in Section 2.1.3, and the full atomic composition of each system is detailed in Table 2.2.1. After solvation, all three systems were equilibrated as described in Section 2.1.4. Finally, the systems were simulated for a total of 200 ns each as described in Section 2.1.5. Once the simulations had been completed, various analysis were conducted as described in Section 2.2.

By comparing the FLB, *apo*, and LoDB KstR simulations the aim is to determine how conformational dynamics of KstR is altered when bound to ligand or DNA, resolving the first objective (Section 1.4).

3.1.1 Ligand or DNA binding affect the shape of KstR

The Radius of gyration (Rg) of FLB, *apo*, and LoDB KstR were calculated as described in Section 2.2.1. By comparing the Rg of FLB, *apo*, and LoDB KstR how the size and shape of KstR is altered by ligand or DNA binding is shown.

The FLB KstR produced the largest average,and LoDB KstR had the smallest average Rg, and *apo* KstR had a Rg between the two (Table 3.1.1). The standard deviation of Rg in *apo* KstR was the highest among the three systems while for LoDB it was the smallest. The standard error for each section also indicates that the differences are significant. This reveals that KstR

is more compact when bound to DNA and larger when bound to ligand, and that the size and shape of *apo* KstR changed more than those of FLB and LoDB KstR.

While *apo* KstR samples *R*g values within the range of LoDB and FLB KstR *R*g, the ranges of FLB and LoDB KstR *R*g values never overlap. Furthermore, from 90 ns the *apo* KstR *R*g gradually decreases towards values like those of LoDB KstR *R*g (Figure 3.1.1).

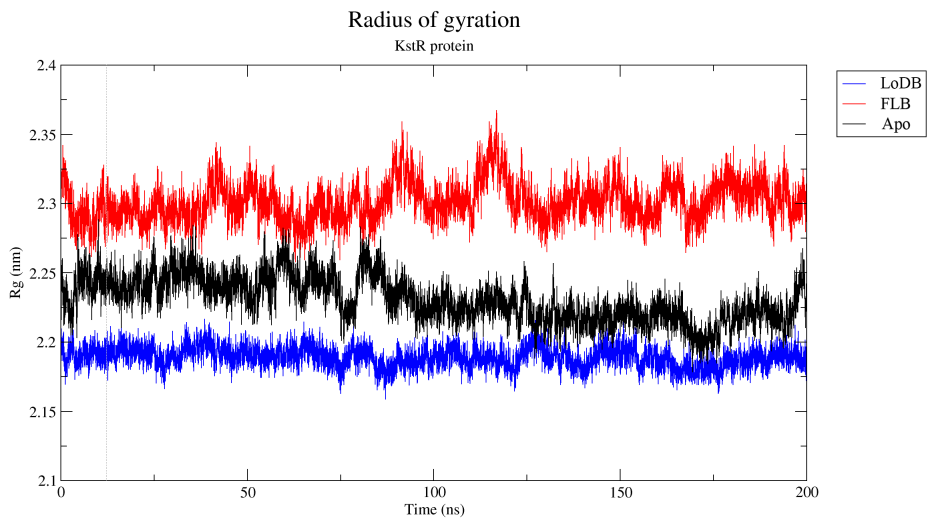

Figure 3.1.1: *Rg of FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.1.1: *Average, standard deviation, standard error, and range of FLB, apo, and LoDB KstR Rg over time (10 - 200 ns).*

Statistics of KstR R*R*				
	$\bar{x}(nm)$	$\sigma(nm)$	$\sigma_{\bar{x}}$ (nm)	Range (max - min)
FLB	2.301	0.014	0.000	2.257 - 2.367 (0.110)
Apo	2.230	0.016	0.002	2.168 - 2.288 (0.120)
LoDB	2.189	0.008	0.000	2.158 - 2.216 (0.058)

The solvent accessible surface area (SASA) of FLB, *apo*, and LoDB KstR were calculated as described in Section 2.2.2. By comparing the SASA of FLB, *apo*, and LoDB KstR how the size and shape of KstR is altered by ligand or DNA binding is highlighted.

FLB KstR exhibited the largest average SASA, LoDB KstR produced the smallest average SASA, and *apo* KstR had a SASA between the two (Table 3.1.2). There is a large difference of 16 nm^2 between FLB and LoDB KstR. The SASA standard error also indicates that these differences are substantial.

The the FLB and LoDB KstR sample different SASA ranges, while *apo* values overlap those of FLB (Figure 3.1.2). Additionally, *apo* KstR sampled SASA values within LoDB SASA range. This reveals that *apo* KstR can exhibit FLB and LoDB SASA, although the *apo* SASA range does not completely overlap with those of FLB and LoDB KstR.

Figure 3.1.2 shows that around 122 ns and 125 ns the FLB KstR SASA falls to about 181 nm^2 which is within the range of LoDB KstR. However, this is among the lowest FLB KstR SASA samples and is never seen again during the simulation. Together, this shows that the size and shape of KstR is largest and fluctuates the most when bound to ligand, while when bound to DNA it is the smallest and changes least.

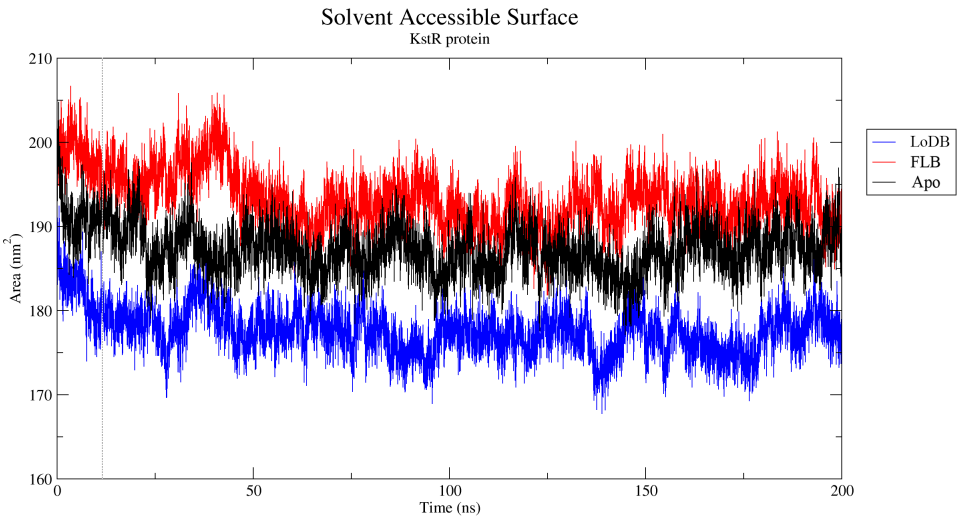

Figure 3.1.2: *SASA of FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.1.2: *Average, standard deviation, standard error, and range of FLB, apo, and LoDB KstR SASA over time (10 - 200 ns).*

Statistics of KstR SASA				
	$\bar{x}(nm^2)$	$\sigma(nm^2)$	$\sigma_{\bar{x}}$ (nm^2)	Range (max - min)
FLB	193.05	3.18	0.03	181.87 - 205.87 (24.00)
Apo	187.08	2.70	0.03	175.22 - 197.71 (22.49)
LoDB	177.39	2.58	0.03	167.76 - 187.85 (20.09)

The volume of FLB, *apo*, and LoDB KstR were calculated as detailed in Section 2.2.2. By comparing the volume of FLB, *apo*, and LoDB KstR how the size of KstR is altered by ligand or DNA binding is shown.

The average KstR volumes of the three systems are very similar, lying within just 1.5nm^3 of one another (Table 3.1.3). FLB KstR maintains the

highest average volume, followed by *apo* and then LoDB KstR as the smallest (Table 3.1.3). While the differences are slight the standard errors indicate they are significant. The range of volumes sampled by each of the three systems KstR volume ranges heavily overlap. This data reveals that there is a slight difference in size among the three systems.

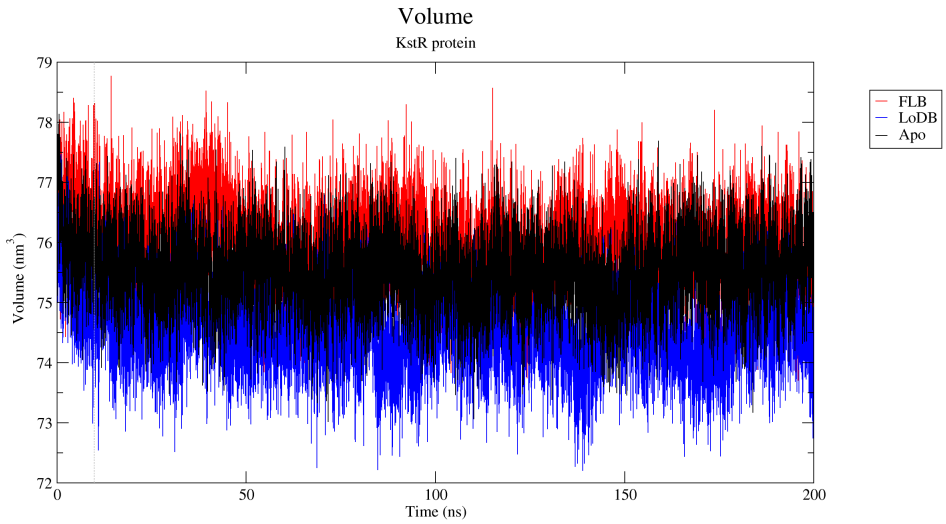

Figure 3.1.3: *Volume of FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.1.3: *Average, standard deviation and standard error, and range of FLB, apo, and LoDB KstR volume over time (10 - 200 ns).*

Statistics of KstR volume				
	$\bar{x}(nm^3)$	$\sigma(nm^3)$	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
FLB	76.004	0.679	0.007	73.497 - 78.769 (5.272)
Apo	75.488	0.632	0.007	72.986 - 77.809 (4.823)
LoDB	74.490	0.629	0.007	72.209 - 77.334 (5.125)

The ligand binding cavity volume of FLB, *apo* KstR, and LoDB was calculated according to Section 2.2.10. By comparing the ligand binding cavity volume of FLB, *apo*, and LoDB KstR h ligand or DNA binding has an impact on the size of the ligand binding cavity is revealed.

The FLB KstR possessed the largest average ligand binding cavity, followed by LoDB, and then the smallest ligand binding cavity average volume was that of *apo* KstR (Table 3.1.4). The ligand binding cavity average volume standard deviation for the *apo* KstR is the highest. The standard error suggests that the differences between the systems ligand binding cavity volumes is significant. This indicates that ligand-bound KstR causes the ligand binding cavity to take on the largest and and most dynamic sizes. This is due to the presence of ligand in the ligand binding cavity itself while the other two systems have

no ligand occupying the cavity.

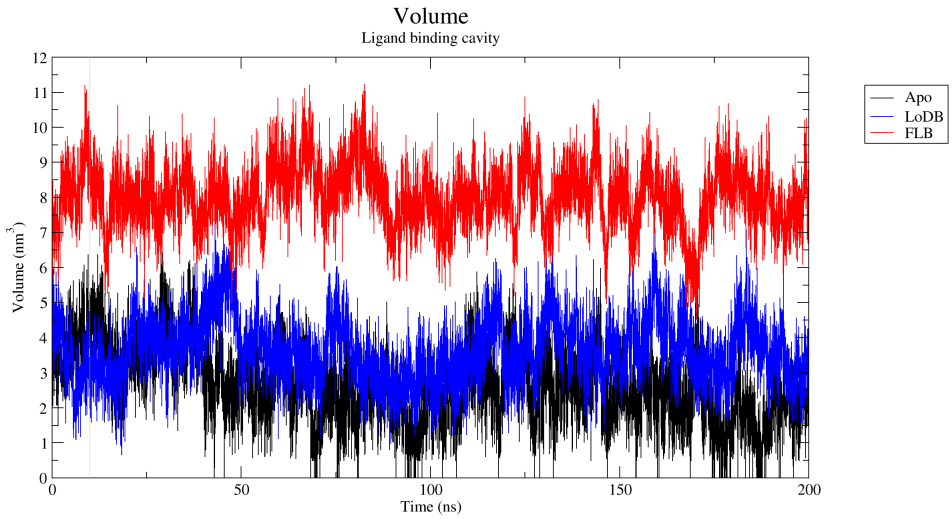

Figure 3.1.4: *Ligand binding cavity volume of FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.1.4: *Table providing average, standard deviation and standard error, and range of FLB, apo, and LoDB of KstR ligand binding cavity volume over time (10 - 200 ns).*

Statistics of KstR ligand binding cavity volume				
	$\bar{x}$ (nm^3)	σ (nm^3)	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
FLB	7.99	0.09	0.01	4.53 - 11.22 (6.69)
Apo	2.62	0.10	0.01	0.00 - 6.56 (6.56)
LoDB	3.65	0.09	0.01	0.85 - 7.74 (6.89)

The *R*g, SASA and volume information demonstrates that ligand or DNA binding drive KstR to take on distinct shapes and sizes. The *R*g and SASA measurements demonstrated that FLB KstR produced the largest structure which changed the most while LoDB KstR produced the smallest structure which changed the least. Furthermore, the volume measurements demonstrated that the volume of KstR was scarcely affected by DNA and ligand binding. Together this demonstrates that ligand and DNA impact KstR shape and dynamics in different ways.

3.1.2 DNA or ligand binding affect the conformational pool of KstR

The conformational ensembles of FLB, *apo*, and LoDB KstR were produced as described in Section 2.2.4. By comparing the conformational clustering between the systems, how ligand or DNA binding affects the nature and rate

of conformational sampling of KstR is revealed with greater detail than the measurements used in Section 3.1.1.

The graphical representation of conformational clustering information is ambiguous. Unfortunately for all conformation clusters their appearance, even if brief, is shorter than the width of the band on the figure indicates. This is because if the width of each band were truly representative of the occupancy time it represents, it would not be visible. Additionally, the number of clusters for each conformational pool varied intensely and an adequate ubiquitous criteria for cluster visualisation was difficult to assign.

The FLB KstR samples 23 different conformational clusters while *apo* KstR samples 13 as shown in Figure 3.1.5. FLB KstR preferentially samples cluster 3 (golden brown), and 4 (magenta), while *apo* KstR predominantly resides within cluster 1 (purple) and 2 (green). FLB KstR fluctuates between a larger number of clusters, persisting within each cluster for shorter periods of time than FLB KstR. FLB and *apo* KstR do not share any conformational clusters. This indicates that ligand binding drives KstR to fluctuate more within a bigger and different pool of conformations.

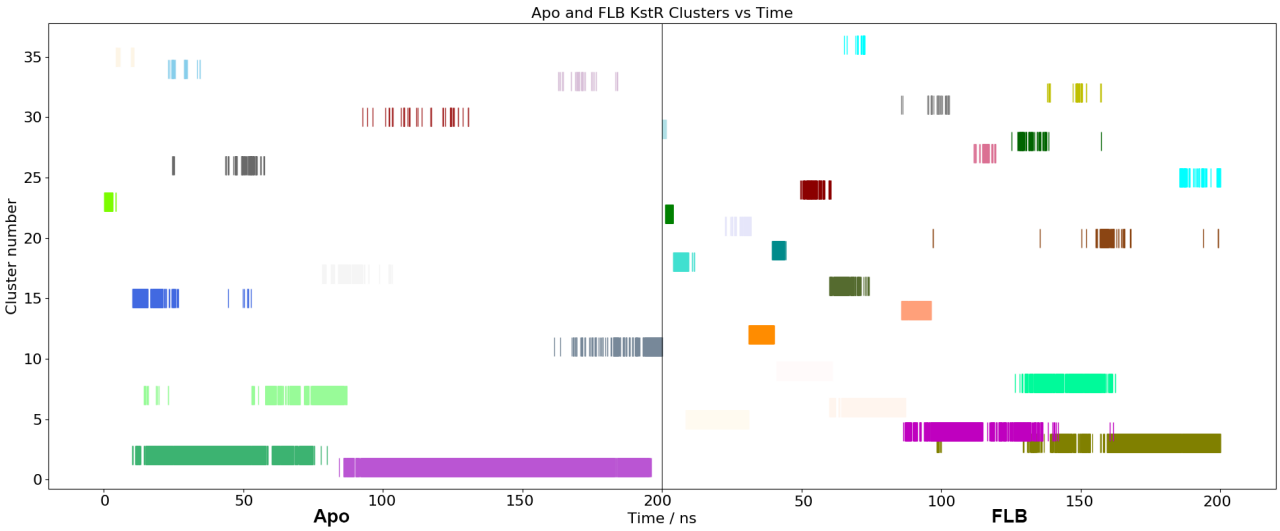

Figure 3.1.5: *36 conformation cluster groups of apo (left), and FLB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with 30 or more conformations in their group with a RMSD cut-off of 0.2 nm.*

Figure 3.1.6 shows that *apo* KstR samples 13 clusters while the LoDB KstR samples 4 over the course of the simulation. *Apo* KstR preferentially samples cluster 2 (lime green) and 3 (black) and fluctuates to sample the 11 other conformational clusters briefly. LoDB KstR predominantly samples cluster 1 (grey), briefly sampling its 3 other clusters. *Apo* and LoDB KstR do not share any conformational clusters. This shows that DNA binding to KstR causes KstR to sample a unique and smaller conformational pool and reduces the conformational fluctuations of KstR.

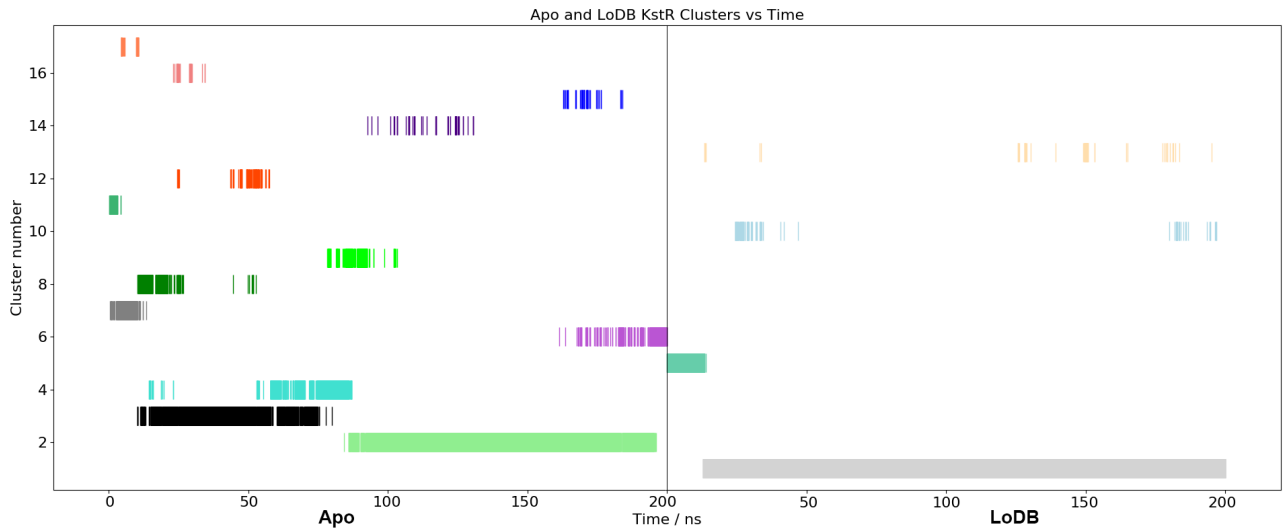

Figure 3.1.6: *17 conformation cluster groups of apo (left), and LoDB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with 30 or more conformations in their group with a RMSD cut-off of 0.2 nm.*

LoDB KstR samples 4 clusters while the FLB KstR samples 23 over the course of the simulation, as illustrated in Figure 3.1.7. LoDB KstR persists within cluster 1 (dark green) for 180 ns briefly and infrequently sampling the 3 other conformational groups. FLB KstR predominantly samples clusters 2 (grey) and 3 (black) while sporadically sampling its 21 other clusters. This shows that compared to ligand-bound KstR, DNA binding reduces the conformational fluctuations and conformational breath of KstR. Together this indicates that DNA binding causes KstR to less freely sample other conformational clusters and exhibit a smaller and unique conformational pool compared to ligand-bound KstR.

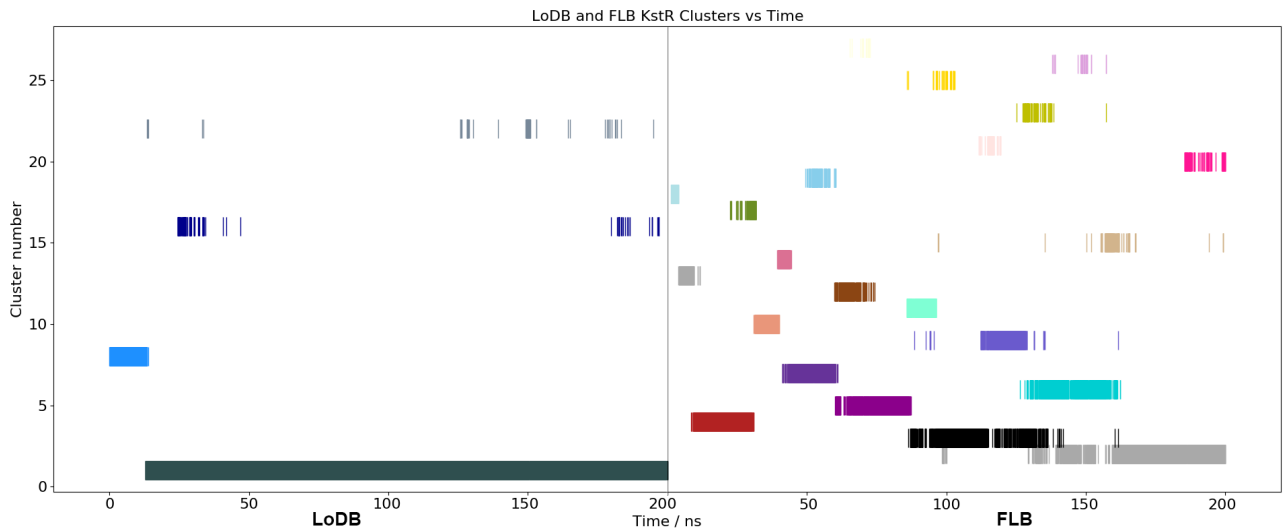

Figure 3.1.7: *Graph plotting 27 conformation cluster groups of LoDB (left), and FLB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with 30 or more conformations in their group with a RMSD cut-off of 0.2 nm.*

The different rates of transition between conformational clusters among FLB, *apo*, and LoDB KstR show that each system produces distinctive dynamics. None of the systems sample the same conformational cluster, this shows that binding with DNA or ligand skews the conformational population of KstR into distinct pools and that *apo* KstR does not sample the complete range of conformations accessible to KstR, in that it does not sample the DNA-bound or ligand-bound conformational clusters. LoDB KstR produced the fewest conformational clusters while FLB KstR produces the most, indicating that DNA binding restrains the conformation of KstR.

3.1.3 DNA or ligand binding affect the flexibility of KstR

The root mean square fluctuation (RMSF) was calculated for FLB, *apo*, and LoDB KstR as detailed in Section 2.2.5. The RMSF reports KstR flexibility at a per-residue level (Figure 3.1.8). Comparison of the RMSF between FLB, *apo* and LoDB KstR demonstrates how KstR flexibility is affected by DNA or ligand binding while also displaying the nature of the conformational changes that underlie the differences in Rg, SASA, volume, and conformational clustering.

Figure 3.1.8 indicates that for residues between 10 and 20 (α1), LoDB KstR has the smallest RMSF while the *apo* and FLB KstR have a larger and similar RMSF. From residue 21 to 30 (α1) LoDB possesses the smallest RMSF however FLB and *apo* KstR then diverge, with *apo* KstR increasing. The RMSF of residues 31 to 34 (α1) of LoDB KstR increases to meet that of FLB while the *apo* still exhibits the highest RMSF. From residue 38 to 50 (α2 to α3), the LoDB RMSF decreases while FLB increases to converge with that of *apo* KstR. The RMSF values of all three systems slowly decrease along the backbone until residue 70 (α4). Together the data indicates that α1 is restrained by the presence of DNA. This also shows that the presence of DNA and ligand decrease the flexibility of residues near the α1 - α2 linker. Finally, these results show that DNA has a restraining effect on residues throughout α2 to α3. Overall, the flexibility of residues throughout the DBD is impacted by DNA while aside from α1 the differences in flexibility between ligand-bound and *apo* KstR are few.

Residues 71 to 90 of the KstR backbone in all three simulations exhibit a substantial peak RMSF as shown in Figure 3.1.8. These residues are embedded in the linker between α4 and α5, the longest non-helical region in the KstR structure. The RMSF of all three systems peaks for residues 75 to 85 (α4 - α5 linker domain) with RMSF values of 0.37 nm at residue 81 for LoDB, 0.45 nm at residue 80 for FLB, and 0.5 nm at residue 83 for *apo* KstR. The RMSF values for all three systems decrease to around 0.1 nm at residue 90 (α5). From residue 92 to 96 (α5) all systems have similar RMSF values. For subsequent residues, the RMSF of *apo* KstR is the highest until residue 110 (α5 - α6 linker). After residue 105 LoDB exhibits the smallest RMSF until residue 118 (α6). From residue 118 to 125 (α6 - α7 linker) FLB exhibits the

highest RMSF while the LoDB KstR continues to present the lowest RMSF. This shows that the $\alpha 4$ and $\alpha 5$ linker domain is intrinsically flexible, and that DNA and ligand binding decreased the flexibility of the $\alpha 4$ and $\alpha 5$ linker domain in addition to $\alpha 5$, $\alpha 6$, and $\alpha 7$, although DNA more so.

For residues 126 to 140 ($\alpha 7$) the *apo* and FLB RMSF decreases, eventually meeting the RMSF of LoDB KstR. From residue 140 the RMSF of all three systems increases until residue 146 ($\alpha 7$) with FLB exhibiting the highest RMSF. From residue 147 to 160 ($\alpha 8$) all three simulations exhibit the same RMSF. Throughout residue 161 to the C-terminus ($\alpha 8$ and $\alpha 9$) the residues of *apo* KstR have a larger RMSF, while the FLB and LoDB KstR have the same RMSF, although increase across residues 170 to 180 ($\alpha 8$ and $\alpha 9$) but then decrease from residues 181 to 190 ($\alpha 9$). For residues 191 to the C-terminus, the RMSF for all three systems increase exponentially. These results demonstrate that ligand-binding enhances the flexibility of the $\alpha 7$ - $\alpha 8$ linker, and that DNA binding reduces the flexibility of $\alpha 8$, $\alpha 9$. Finally, as is commonly observed for most proteins, the C-terminus of all three systems exhibits large scale fluctuations [87]. DNA binding decreased the flexibility of the LBD while ligand binding restrained regions about the ligand binding cavity and $\alpha 9$ but enhanced the flexibility of the $\alpha 7$ - $\alpha 8$ linker.

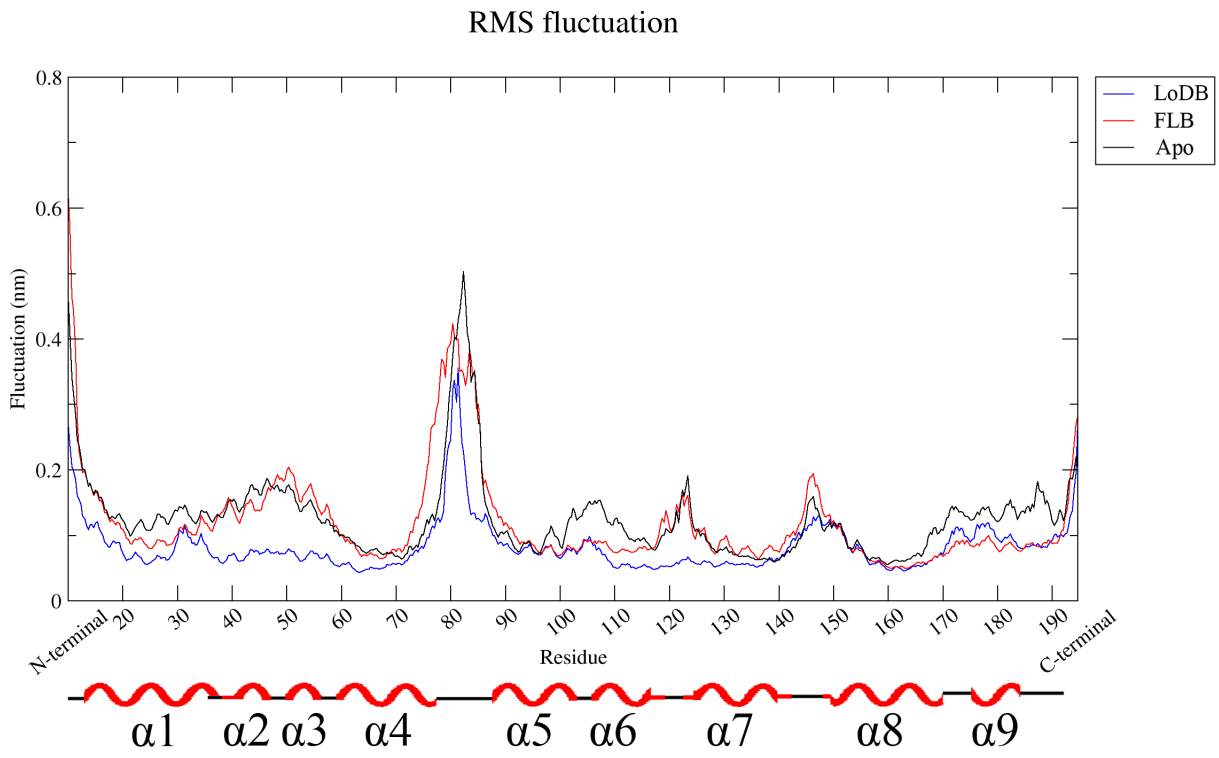

Figure 3.1.8: *Backbone RMSF of KstR chain for FLB (red), apo (black), and LoDB (blue) KstR. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

DNA and ligand have unique effects on the flexibility of KstR. Throughout

KstR, DNA binding reduced the flexibility of the backbone. Ligand binding restrained a lesser portion of the LBD than DNA binding. This demonstrates that ligand binding does not impact the flexibility as greatly as DNA binding.

3.1.4 DNA or ligand binding impact how the DBD and LBD of KstR are arranged

The distances between residue pairs for FLB, *apo*, and LoDB KstR were calculated as described in Section 2.2.7. By comparing distances between residues in FLB, *apo*, and LoDB DBDs how the arrangement of the KstR DBD is altered by DNA or ligand binding is shown. The locations of the DBD residue pairs measured are shown in Figure 2.2.2.

Table 3.1.5 shows that the LoDB KstR exhibited the smallest distances, smallest average distance, and narrowest range of distances. For all residue pairs other than Tyr-54, *apo* KstR sampled distances within the FLB KstR range but by 80 ns diverged to sample distances closer to LoDB KstR. For all residue pairs but Arg-40 and Tyr-54, *apo* KstR sampled residue pair distances within both FLB and LoDB KstR ranges. For all residue pairs *apo* KstR had the largest average distance standard deviation. The FLB KstR exhibited the largest distances and largest average distance for all residue pairs. For all residue pairs LoDB and FLB KstR never overlapped in distances sampled (Figure 3.1.9).

Together this demonstrates that DNA binding causes the distances between residues at the DBD to decrease, while in most cases ligand binding resulted in the distances between residues at the DBD to increase. *Apo* KstR exhibited the most dynamic distances between residues at the DBD and also could sample distances within both the FLB and LoDB KstR ranges, however FLB KstR and LoDB never sampled within the same range of distances for any pair of residues.

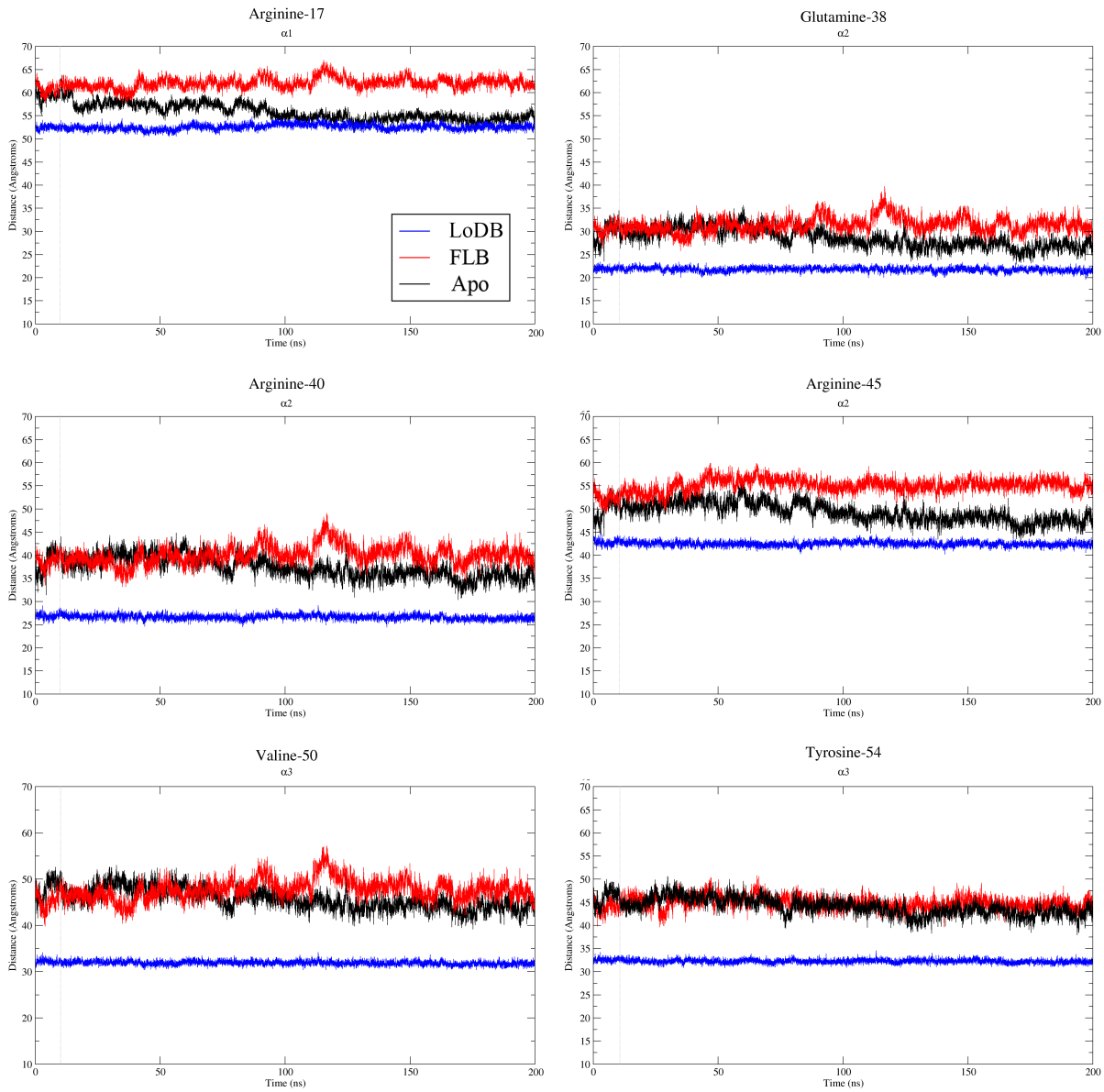

Figure 3.1.9: *Distance between Arg-17, Gln-38, and Arg-40, Arg-45, Val-50, and Tyr-54 of each DBD of FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.2 for the DBD residue position reference.*

Table 3.1.5: *Average, standard deviation and standard error, and range of distance between Arg-17, Gln-38, and Arg-40, Arg-45, Val-50, and Tyr-54 for FLB, apo, and LoDB KstR over time (10 - 200 ns).*

Statistics of Arg-17 distance				
	$\bar{x}$(Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	62.08	1.16	0.01	57.52 - 67.02 (9.50)
Apo	55.78	1.59	0.02	51.58 - 63.79 (12.21)
LoDB	52.58	0.61	0.01	50.47 - 55.19 (4.72)
Statistics of Gln-38 distance				
FLB	31.45	1.64	0.02	25.84 - 39.70 (13.86)
Apo	28.54	2.05	0.02	22.90 - 35.59 (12.69)
LoDB	21.78	0.47	0.01	19.61 - 23.39 (3.78)
Statistics of Arg-40 distance				
FLB	40.09	2.05	0.02	33.19 - 49.07 (15.88)
Apo	37.36	2.28	0.02	30.41 - 44.83 (14.42)
LoDB	26.63	0.53	0.05	24.45 - 29.13 (4.68)
Statistics of Arg-45 distance				
FLB	55.17	1.33	0.01	49.61 - 59.91 (10.30)
Apo	49.29	2.03	0.02	43.30 - 55.73 (12.43)
LoDB	42.49	0.51	0.01	40.25 - 44.53 (4.28)
Statistics of Val-50 distance				
FLB	47.81	2.20	0.02	40.37 - 57.14 (16.77)
Apo	45.68	2.21	0.02	39.19 - 52.97 (13.78)
LoDB	31.89	0.46	0.01	30.39 - 34.19 (3.80)
Statistics of Tyr-54 distance				
FLB	45.47	1.39	0.01	39.63 - 50.75 (11.12)
Apo	43.96	1.83	0.01	37.94 - 50.59 (12.65)
LoDB	32.25	0.42	0.04	30.79 - 34.53 (3.74)

By comparing distances between residues in FLB, *apo*, and LoDB LBDs how the arrangement of the KstR LBD is altered by DNA or ligand binding is shown. See Figure 2.2.3 for the LBD residue position reference

For all residue pairs but Asp-162, LoDB KstR exhibited the smallest average distance for each residue pair (Table 3.1.6). For all residue pairs but Tyr-118 and Asp-162, FLB KstR exhibited the largest average distance. For all residue pairs but Ala-103 and Met-109, LoDB KstR exhibited the smallest standard deviation. For all residue pairs but Tyr-118, the three systems ranges overlap (Figure 3.1.10).

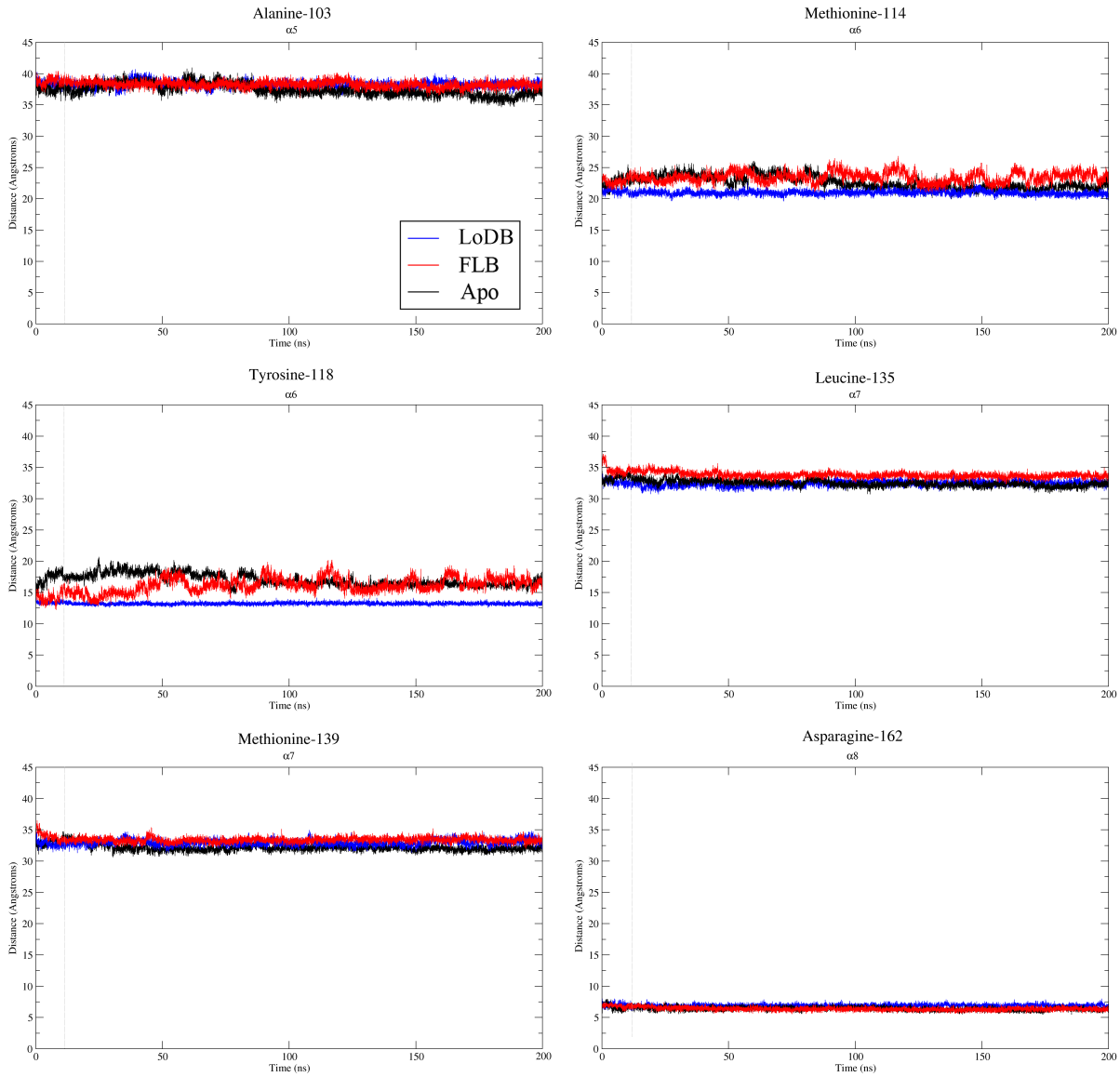

Figure 3.1.10: *Distance between Ala-103, Met-114, and Tyr-118, Leu-135, Met-139, Asp-162 of each LBD of FLB (red), apo (black), and FLB (blue) protein volume over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.3 for the residue position reference.*

Table 3.1.6: *Average, standard deviation and standard error, and range of distance between Ala-103, Met-114, and Tyr-118, Leu-135, Met-139, Asp-162 for FLB, apo, and LoDB KstR over time (10 - 200 ns).*

Statistics of Ala-103 distance				
	$\bar{x}$(Å)	σ(Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	38.285	0.519	0.005	36.099 - 40.189 (4.090)
Apo	37.424	0.929	0.009	34.678 - 40.908 (6.229)
LoDB	38.234	0.553	0.006	36.369 - 40.655 (4.285)
Statistics of Met-114 distance				
FLB	23.45	0.85	0.01	20.98 - 26.77 (5.79)
Apo	22.57	1.06	0.01	19.98 - 26.01 (6.03)
LoDB	20.94	0.35	0.00	19.63 - 22.55 (2.92)
Statistics of Tyr-118 distance				
FLB	16.22	1.11	0.01	12.88 - 20.19 (7.31)
Apo	17.01	0.95	0.01	14.77 - 20.66 (5.89)
LoDB	13.22	0.19	0.00	12.59 - 14.16 (1.57)
Statistics of Leu-135 distance				
FLB	33.711	0.518	0.005	32.240 - 36.032 (3.791)
Apo	32.439	0.471	0.005	30.734 - 34.924 (4.190)
LoDB	32.403	0.430	0.004	30.725 - 33.934 (3.208)
Statistics of Met-139 distance				
FLB	33.387	0.389	0.004	31.976 - 35.364 (3.388)
Apo	32.136	0.420	0.004	30.624 - 34.732 (4.107)
LoDB	32.998	0.471	0.005	31.323 - 34.936 (3.612)
Statistics of Asp-162 distance				
FLB	6.367	0.231	0.002	5.457 - 7.397 (1.939)
Apo	6.400	0.285	0.003	5.489 - 7.595 (2.105)
LoDB	6.902	0.217	0.002	6.014 - 7.963 (1.948)

DNA binding induces the smallest distances between residues at the KstR LBD which changed the least, suggesting the contracting and restraining effects of DNA. The ligand-bound KstR exhibited the largest average distance for most residue pair distances. There were no trends consistent among all residue pairs at the LBD and there are fewer notable differences between the three systems than for the DBD.

3.1.5 DNA or ligand binding to KstR perturbs the structure of KstR

The inter-residue contact maps for FLB, *apo*, and LoDB KstR were generated according to Section 2.2.9. Differences among Figures 3.1.11, 3.1.12, and 3.1.13 demonstrate how ligand or DNA binding affects the inter-residue structure of KstR and for which regions.

The LoDB KstR induces the shortest distances between residue 34 and 54 (α1 with α2) as well as between residues 38 to 50 (α1 with α2) and 70 (α4).

DNA binding caused shorter distances between residues 66 to 76 ($\alpha4$) and 106 to 118 ($\alpha6$). The distances between residues 22 to 30 ($\alpha1$) and 102 to 110 ($\alpha5$), and between residues 14 to 30 ($\alpha1$) and 118 to 138 ($\alpha6$) are shorter for LoDB KstR. The distances between residues 82 to 90 ($\alpha5$) and 138 to 146 ($\alpha7$) were farthest for the FLB KstR and closest for LoDB KstR. This demonstrates that DNA binding brings together residues within the DBD and LBD, and between the DBD and LBD.

The FLB KstR exhibited the shortest distances between residues 46 to 70 ($\alpha2$ - $\alpha4$) and 66 to 70 ($\alpha4$). Ligand binding also induced KstR to take on shorter distances for residues between 86 to 90 and 90 to 94 ($\alpha4$ and $\alpha5$). This shows that ligand binding brought together regions around the $\alpha4$ - $\alpha5$ linker domain.

Both the LoDB and FLB KstR exhibit shorter distances between residues 70 to 90 ($\alpha5$) and 90 to 110 ($\alpha6$) compared to *apo* KstR. The distances between residues 82 to 90 ($\alpha5$) and 150 to 158 ($\alpha8$), 94 to 102 ($\alpha5$) and 118 ($\alpha6$, in addition 90 to 102 ($\alpha5$) and 170 to 182 ($\alpha9$) of *apo* KstR are the shortest. This reveals that ligand and DNA binding caused distances within the LBD to become longer.

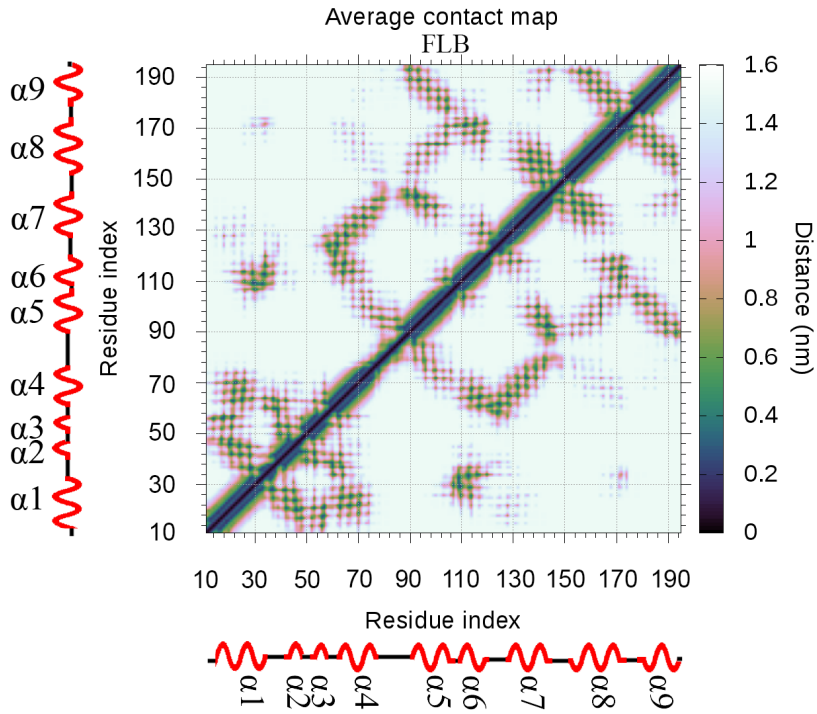

Figure 3.1.11: *Average inter-residue distance of intramolecular interactions of FLB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

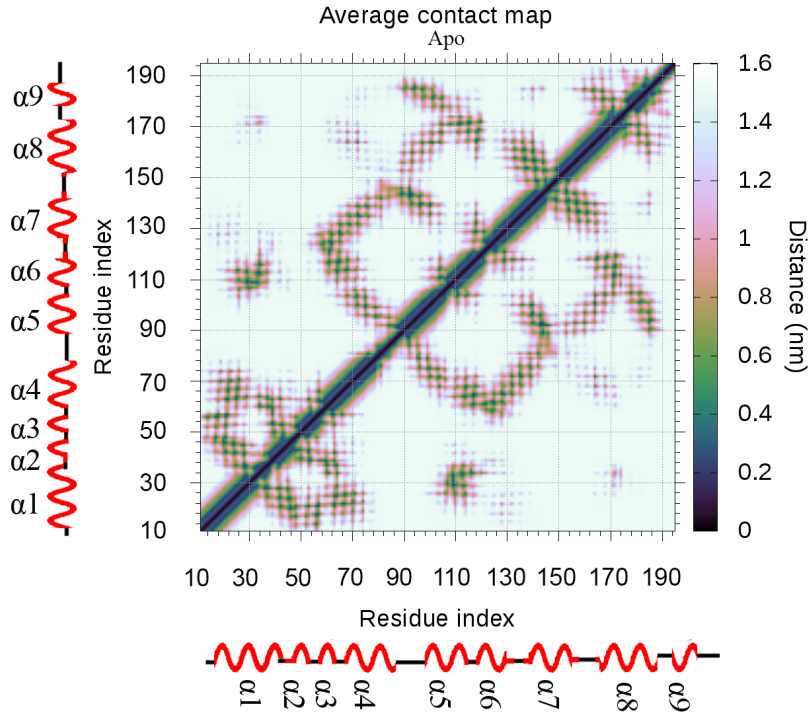

Figure 3.1.12: *Average inter-residue distance of intramolecular interactions of apo KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

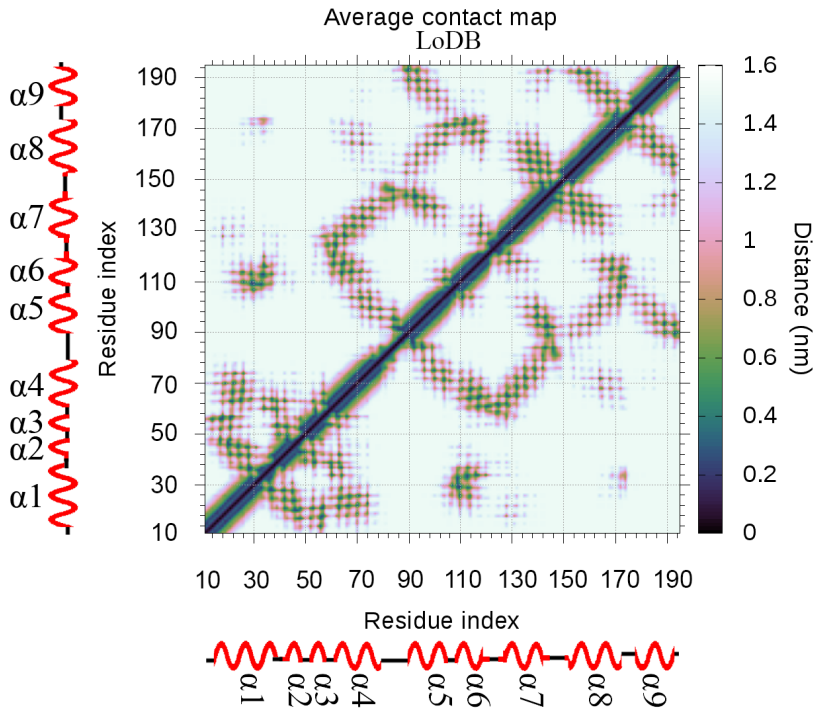

Figure 3.1.13: *Average inter-residue distance of intramolecular interactions of LoDB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

DNA binding altered contacts within the DBD and LBD, and between the LBD and DBD, while ligand binding only affected the organisation within the DBD and LBD.

3.1.6 DNA or ligand binding affect the secondary structure of KstR

The secondary structure time series for FLB, *apo*, and LoDB KstR were calculated as described in Section 2.2.6. A comparison of the FLB, *apo*, and LoDB KstR secondary identifies if regions of KstR are affected by ligand or DNA binding. Figures 3.1.14, 3.1.15, and 3.1.16 illustrate the differences between the secondary structure of each system.

The FLB KstR exhibited a 3_{10}-helix instead of an α-helix at the $\alpha1$ helix. The $\alpha4$ - $\alpha5$ linker domain is composed of the fewest residues for LoDB KstR while composed of the most for FLB KstR. The LoDB KstR had the largest $\alpha6$ - $\alpha7$ linker. Together this indicates that ligand extends the $\alpha4$ and $\alpha5$ helices.

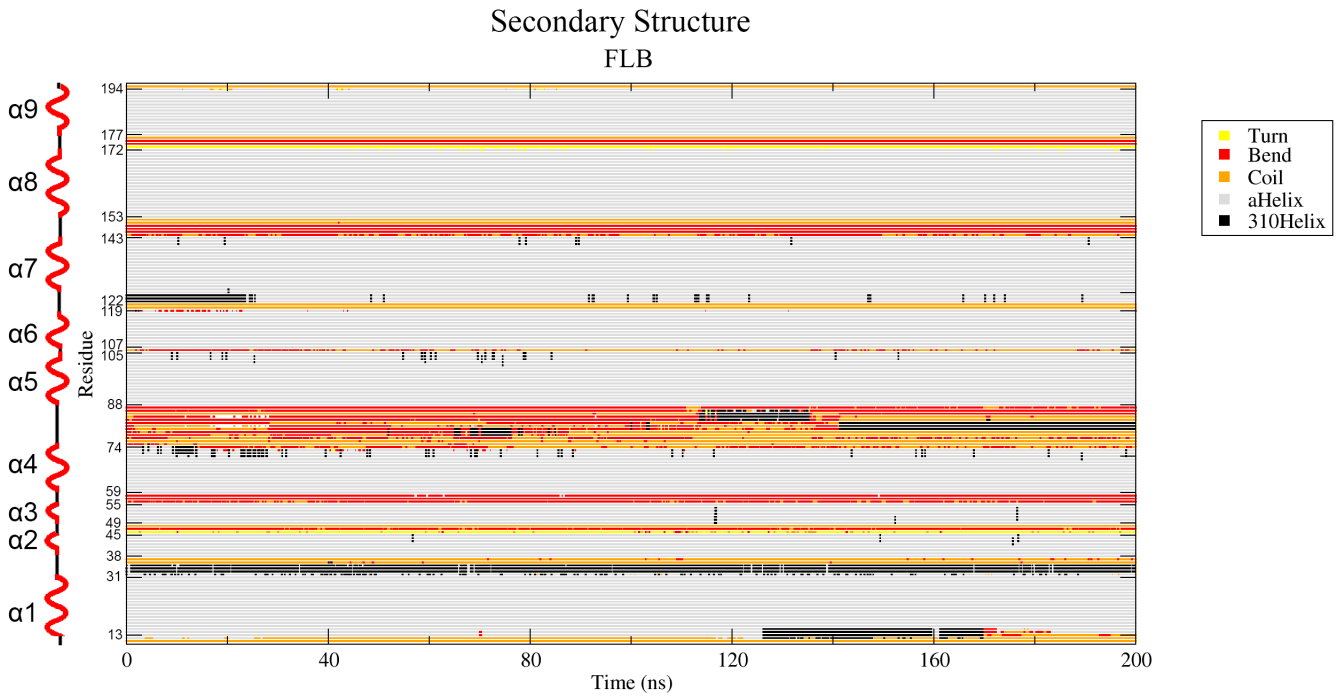

Figure 3.1.14: *Secondary structure assignment of FLB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

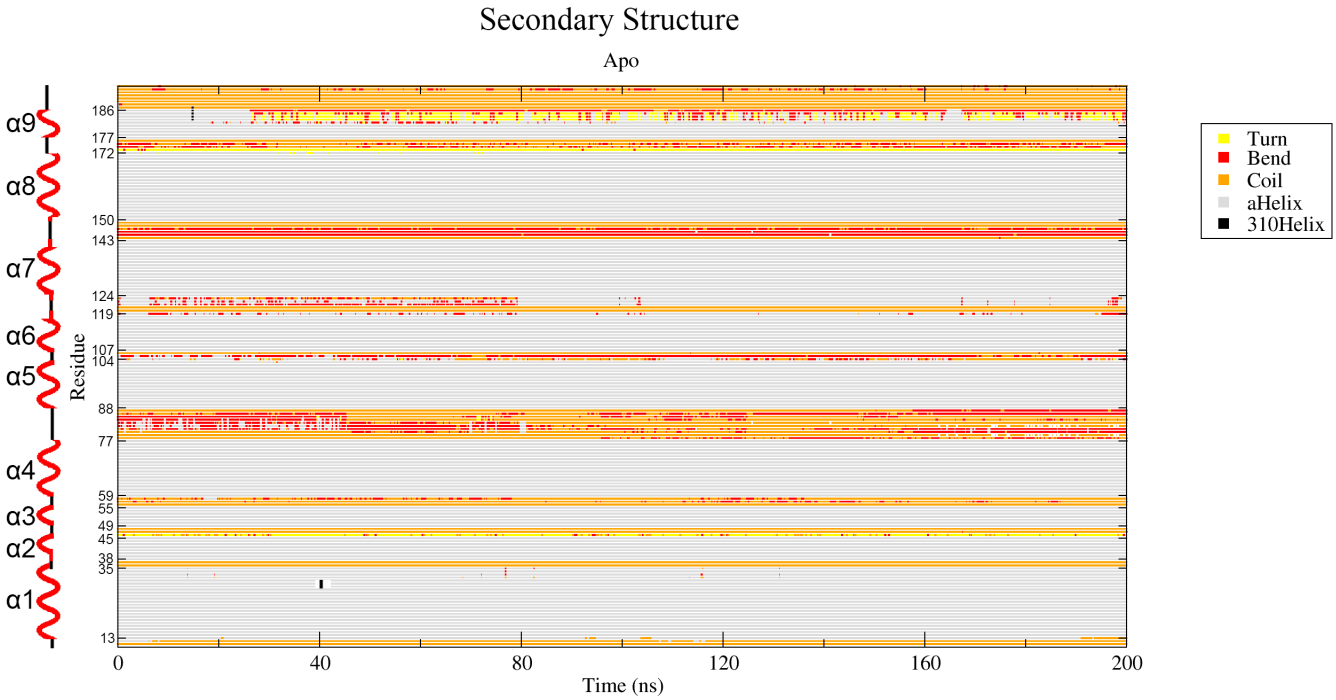

Figure 3.1.15: *Secondary structure assignment of apo KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

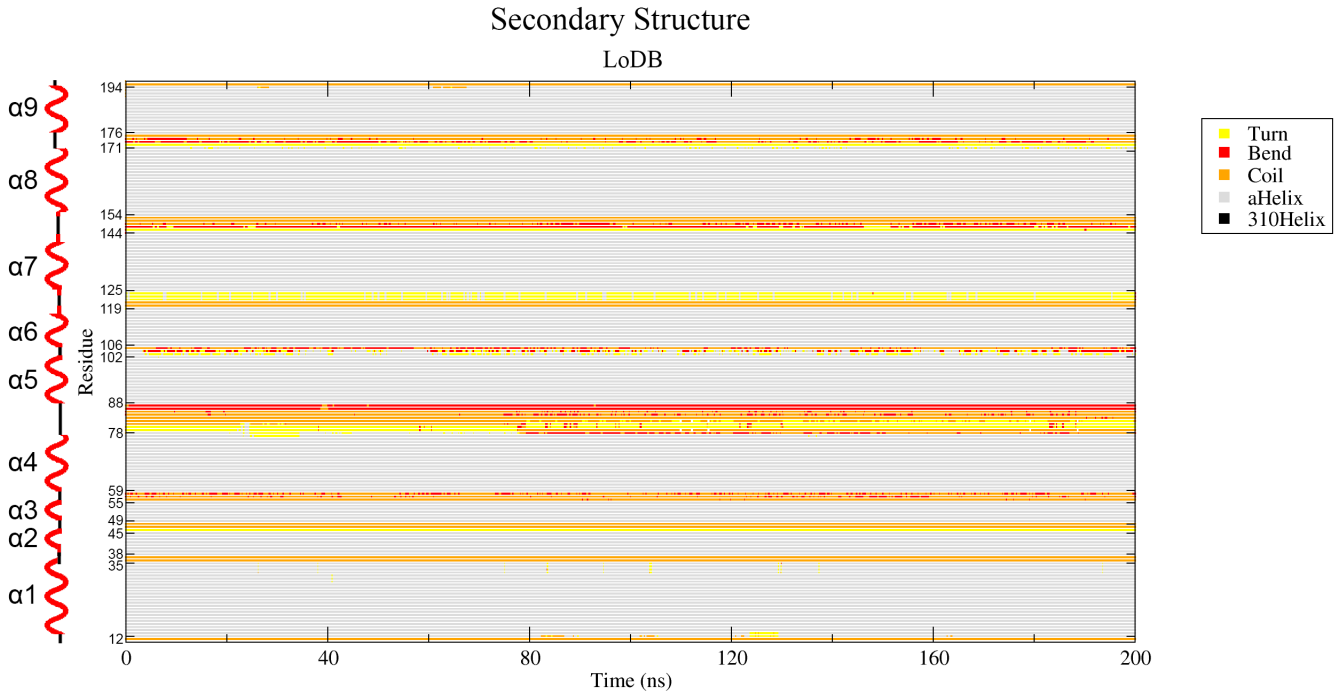

Figure 3.1.16: *Secondary structure assignment of LoDB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

The secondary structure composition of KstR is altered depending on how DNA or ligand binds to KstR. The *apo* KstR exhibits the smallest population of helical secondary structure over the simulation of 71.7% exceeded by FLB KstR at 75.6%, and then LoDB KstR with highest proportion of 77.9% (Table 3.1.7). This demonstrates that DNA or ligand binding induces the formation of helices, DNA more so.

Coil, turn, and bend secondary structures typically occupy regions between helices. The *apo* KstR showed the highest coil, turn, and bend composition of 28.3%, followed by FLB KstR at 23.8%, and then LoDB KstR at 21.5%. This indicates that ligand or DNA binding to KstR disfavours loop regions instead restraining them as helices.

Table 3.1.7: *Average secondary structure assignment for FLB, apo, and LoDB KstR over time (10 - 200 ns).*

Secondary structure distribution of FLB, *apo*, and LoDB KstR					
	α-helices	Coil	Turn	Bend	3_{10}-helices
FLB	74.3%	11.9%	5.6%	6.3%	1.3%
Apo	70.5%	14.3%	6.1%	7.9%	1.2%
LoDB	76.6%	8.7%	5.3%	7.5%	1.4%

3.1.7 DNA or ligand binding affect Tyr-54 and Trp-164 flipping differently

The Tyr-54 and Trp-164 dihedral angles for FLB, *apo*, and LoDB KstR were calculated according to Section 2.2.8. Tyr-54 is located at the DBD and has close contacts with DNA, while Trp-164 is located within the ligand binding cavity and has close contacts with the ligand. By comparing FLB, *apo*, and FLB KstR Tyr-54 and Trp-164 dihedral angles how DNA or ligand binding affects the conformation of the residues is shown.

There are four possible states that Tyr-54 exhibited, *apo* and FLB KstR sampled three of them stochastically over the course of the simulation, while LoDB KstR is fixed in one angle (Figure 3.1.17). Furthermore, the LoDB angle is exclusive to it and lies between two possible angles sampled by FLB and *apo* KstR. The standard deviation of Tyr-54 dihedral angle was smallest for LoDB KstR (Table 3.1.8). This shows that when DNA was bound to KstR Tyr-54 favoured a significantly larger dihedral angle and did not fluctuate as much.

The FLB KstR Trp-164 had a substantially different dihedral angle compared to *apo* and LoDB KstR which had a similar Trp-164 dihedral angle (Table 3.1.8). This indicates that the presence of ligand induces Trp-164 to take on a unique flipped position (Figure 3.1.17).

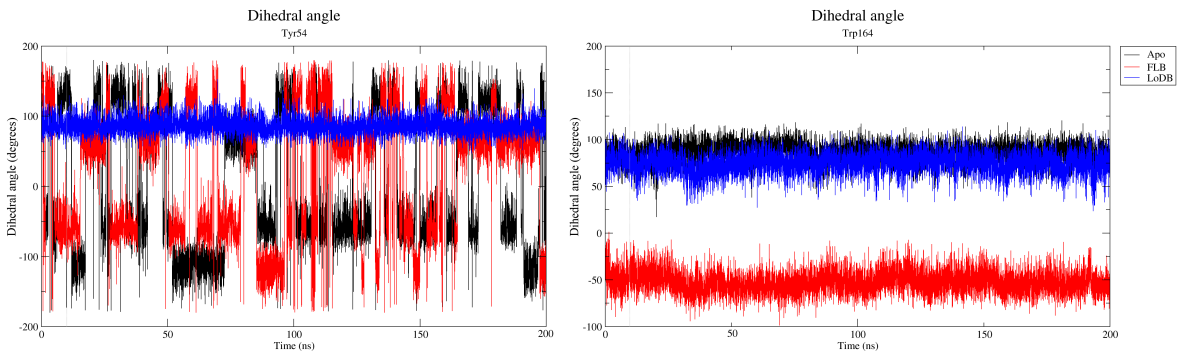

Figure 3.1.17: *Dihedral angle across Tyr-54 (left) and Trp-164 (right) for FLB (red), apo (black), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.4 for the residue dihedral reference.*

Table 3.1.8: *Average, standard deviation and standard error, and range of FLB, apo, and LoDB KstR of Tyr-54 and Trp-164 dihedral angle over time (10 - 200 ns).*

Statistics of Tyr-54 dihedral angle				
	$\bar{x}$ (°)	σ (°)	$\sigma_{\bar{x}}$ (°)	Range (max - min)
FLB	23.5	87.2	0.9	-179.9 - 179.4 (358.9)
Apo	6.9	95.1	0.9	-178.8 - 179.9 (358.7)
LoDB	86.4	11.9	0.1	45.6 - 152.4 (106.8)
Statistics of Trp-164 dihedral angle				
FLB	-52.9	11.6	0.1	-98.7 - -7.1 (91.6)
Apo	86.1	10.5	0.1	17.6 - 120.3 (102.7)
LoDB	74.2	11.3	0.1	23.7 - 113.8 (90.1)

This shows that Tyr-54 takes on a distinct and strict organisation when KstR is bound to DNA. Additionally, that when KstR is bound to ligand Trp-164 takes on a rigid conformation.

3.1.8 DNA or ligand contributions produce different enthalpic contributions with KstR

The enthalpy calculations for FLB and LoDB KstR are detailed in Section 2.2.3. By measuring the enthalpy between KstR and DNA or ligand, the contribution that binding to either molecule makes to the total enthalpy of KstR can be estimated.

The DNA:KstR has a more favourable interaction than the DNA:KstR interaction (Table 3.1.9). The DNA:DBD enthalpy energy is immensely favourable, however the DNA:LBD enthalpy energy is unfavourable. The ligand:LBD and ligand:DBD enthalpy energies are favourable, ligand:LBD more so. Together this shows that DNA binding is heavily skewed towards the DBD where it

is favourable, but has unfavourable interactions with the LBD. The enthalpy between KstR and ligand is more evenly dispersed and favourable for both the LBD and DBD, although more intensely favourable for the LBD.

Table 3.1.9: *Average and error estimate of the contribution of the KstR-ligand or KstR-DNA enthalpy to the overall enthalpy between KstR and ligand or DNA for FLB or LoDB KstR respectively over time (0 - 200 ns).*

Enthalpy between KstR and ligand or DNA		
	Average (kJ/mol)	Error estimate
FLB	-1390	24
LoDB	-1980	31
Enthalpy between KstR DBD and ligand or DNA		
FLB	-400	50
LoDB	-1987	34
Enthalpy between KstR LBD and ligand or DNA		
FLB	-990	30
LoDB	7	1

3.2 Ligand removal affects KstR conformational dynamics

To examine if ligand removal changes the conformational dynamics of KstR a ligand removed KstR system was prepared. The From Full Ligand bound (FrFLB) system was produced by removing 3OChA-CoA from the FLB KstR structure prepared earlier (Section 3.1).

The FrFLB KstR was then solvated as described in Section 2.2.3, and the full atomic composition of the system is detailed in Table 2.2.1. After solvation, the system was equilibrated as described in Section 2.1.4. Finally, the system was simulated for a total of 200 ns as described in Section 2.1.5. Once the simulations had been completed various analysis were conducted as described in Section 2.2.

By comparing the FLB, FrFLB, and *apo* KstR simulations the aim is to determine how the conformational dynamics of KstR changes upon ligand removal, partially resolving the second objective (Section 1.4)

3.2.1 KstR shape becomes smaller after ligand removal

The Rg of FrFLB KstR was calculated as described in Section 2.2.1. Comparison of the Rg of FLB, FrFLB, and *apo* KstR demonstrates how ligand removal changes the shape and size of KstR.

FLB KstR has the greatest average, followed by FrFLB and then the *apo* KstR (Table 3.2.1). The standard error indicates that the differences are significant. Figure 3.2.1 shows that within the first 10 ns the FrFLB KstR Rg decreased from 2.3 nm to 2.25 nm, similar to the Rg of *apo* KstR. Until 90 ns the Rg of FrFLB and *apo* KstR both sample Rg around 2.25 nm, after which *apo* KstR Rg decreases to about 2.2 nm. This shows that the removal of ligand caused KstR to take on smaller shapes like that of *apo* KstR. However, FrFLB KstR displayed a larger shapes than *apo* KstR during the latter part of the simulation.

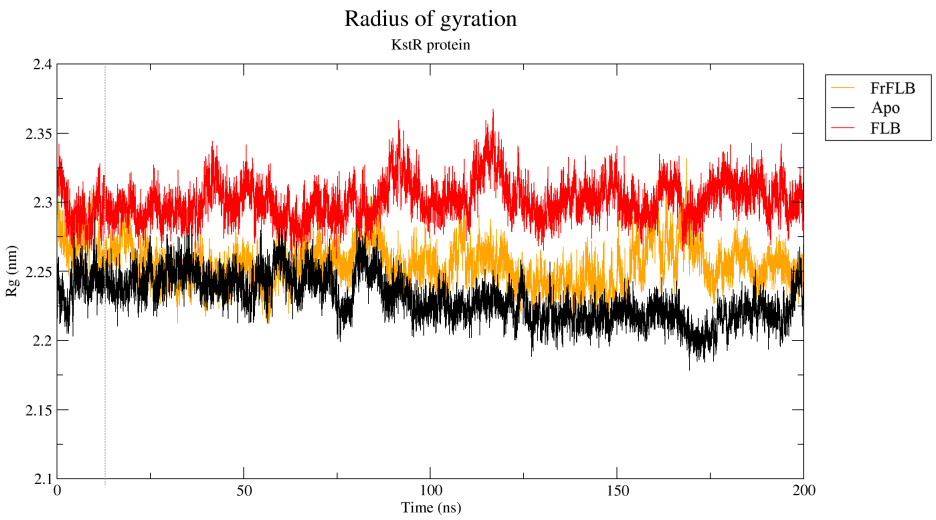

Figure 3.2.1: *KstR Rg of FLB (red), FrFLB (orange), and apo (black) over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.2.1: *Average, standard deviation, standard error, and range of FLB, FrFLB, and apo KstR Rg over time (10 - 200 ns).*

Statistics of KstR Rg				
	$\bar{x}(nm)$	$\sigma(nm)$	$\sigma_{\bar{x}}$ (nm)	Range (max - min)
FLB	2.301	0.014	0.000	2.257 - 2.367 (0.109)
FrFLB	2.252	0.014	0.000	2.211 - 2.307 (0.096)
Apo	2.230	0.016	0.002	2.168 - 2.288 (0.119)

The SASA of FrFLB KstR was calculated as described in Section 2.2.2. Comparison of FLB, FrFLB, and *apo* KstR SASA demonstrates how ligand removal changes the shape and size of KstR.

The greatest average KstR SASA is produced by FLB, followed by that of *apo* KstR, and the smallest is the FrFLB KstR average SASA (Table 3.2.2). The SASA range of the three systems overlap but FLB KstR samples the largest values. The average SASA standard deviation of FLB was the largest, followed by FrFLB and then *apo* KstR. The standard error indicates that these differences are significant. Within the first 10 ns the FLB KstR SASA decreased to about 195 nm^2, and *apo* and FrFLB KstR decreases to about 188 nm^2 (Figure 3.2.2). The *apo* and FrFLB KstR sample SASA beneath the range of FLB KstR around 180 nm^2. Together this indicates that when ligand is removed the KstR SASA becomes smaller and fluctuates less.

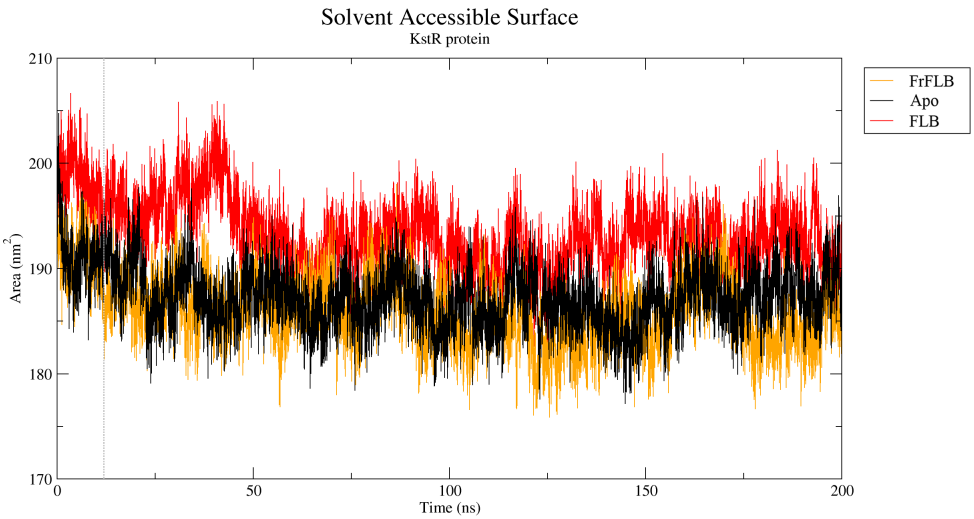

Figure 3.2.2: *SASA of FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.2.2: *Average, standard deviation and standard error, and range of FLB, FrFLB, and apo KstR SASA over time (10 - 200 ns).*

Statistics of KstR SASA				
	$\bar{x}(nm^2)$	$\sigma(nm^2)$	$\sigma_{\bar{x}}$ (nm^2)	Range (max - min)
FLB	193.05	3.18	0.03	181.87 - 205.87 (24.00)
FrFLB	185.79	3.01	0.03	175.88 - 198.00 (22.12)
Apo	187.08	2.69	0.03	175.22 - 197.71 (22.49)

The volume of FrFLB KstR was calculated as described in Section 2.2.2. Comparison of the KstR volume of FLB, FrFLB, and *apo* KstR demonstrates how ligand removal changes the size of KstR.

There is a slight difference of about 0.6 nm^3 between the largest volume for FLB KstR and the lowest for FrFLB KstR (Table 3.2.3). The standard error values indicate that differences in volume are significant. All three systems volumes heavily overlap in volume range (Figure 3.2.3). The volume of KstR decreases upon ligand removal, however the volume difference between the different systems is small.

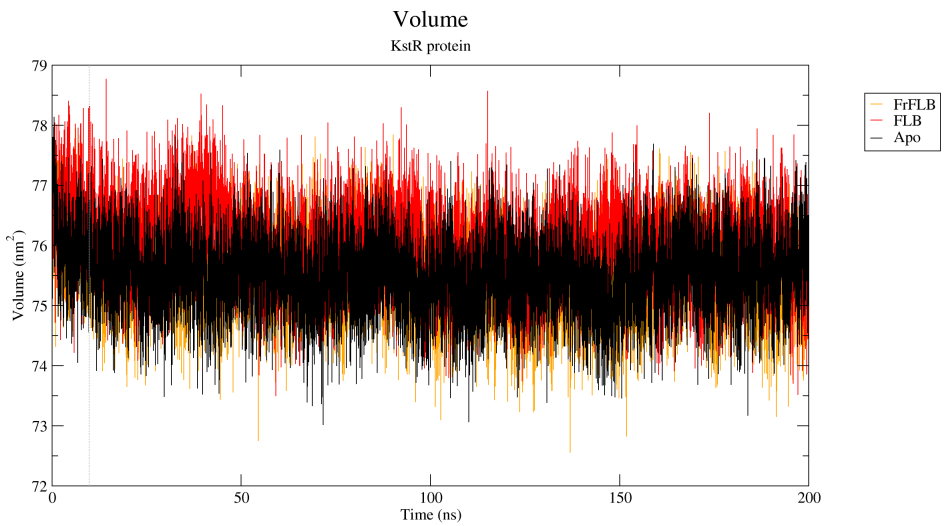

Figure 3.2.3: *Volume of FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.2.3: *Average, standard deviation and standard error, and range of FLB, FrFLB, and apo KstR volume over time (10 - 200 ns).*

Statistics of KstR volume				
	$\bar{x}(nm^3)$	$\sigma(nm^3)$	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
FLB	76.004	0.679	0.007	73.497 - 78.769 (5.272)
FrFLB	75.446	0.653	0.007	72.558 - 77.837 (5.279)
Apo	75.488	0.632	0.007	72.986 - 77.809 (4.823)

The ligand binding cavity volume of FrFLB KstR was calculated as described in Section 2.2.10. Comparison of of FLB, FrFLB, and *apo* KstR the KstR ligand binding cavity volume demonstrates how ligand removal changes the ligand binding cavity size.

The ligand binding cavity is largest for FLB KstR, followed by FrFLB KstR, and the smallest for *apo* KstR (Table 3.2.4). The standard deviation of the ligand binding cavity average volume was largest for *apo* KstR, which was near that of FrFLB, while the ligand binding cavity standard deviation was substantially smaller for FLB KstR. The standard error demonstrates that these differences are significant The smallest ligand binding cavity volume was produced by *apo* KstR at 0.00 nm^3 and the largest for FLB at 11.22 nm^3. Figure 3.2.4 also shows that FrFLB KstR ligand binding cavity volume overlapped with that of *apo* KstR more frequently than FLB(Figure 3.2.4). Together this shows that upon ligand removal the volume of the ligand binding cavity became smaller and changed more frequently. However, the cavity never closed completely like that of *apo* KstR on the timescale of the simulation (Section 4.0.1).

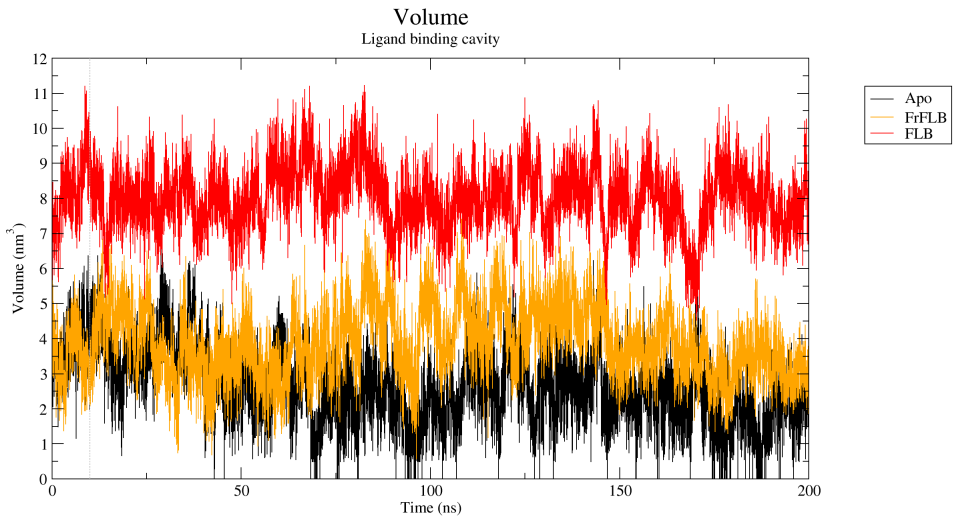

Figure 3.2.4: *Ligand binding cavity volume of FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.2.4: *Table providing average, standard deviation and standard error, and range of FLB, FrFLB, and apo of KstR average ligand binding cavity volume over time (10 - 200 ns).*

Statistics of KstR ligand binding cavity volume				
	$\bar{x}$ (nm^3)	σ (nm^3)	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
FLB	7.99	0.89	0.01	4.53 - 11.22 (6.69)
FrFLB	3.89	1.01	0.01	0.58 - 7.63 (7.05)
Apo	2.62	1.04	0.01	0.00 - 6.56 (6.56)

Section 3.2.1 has shown that the removal of ligand causes the shape of KstR to become smaller and change more, taking on *apo* KstR-like shapes. The ligand binding cavity volume of KstR changed upon ligand removal although did not shrink to *apo* KstR volumes over the timescale of the simulation (Section 4.0.1). This shows that when ligand is removed the architecture of the ligand binding cavity does not fully collapse. Finally, the volume of KstR was not affected by ligand removal.

3.2.2 Ligand removal alters the conformational pool of KstR

The conformational ensembles of FLB, FrFLB, and *apo* KstR were produced as described in Section 2.2.4. By comparing the conformational clusters of FrFLB, FLB, and *apo* KstR how the conformational ensemble changes as a result of ligand removal is shown. This provides more detailed information on the conformation and rate of transition between conformations of KstR than Section 3.2.1.

FrFLB KstR sampled 26 conformational clusters and preferentially resided in cluster 2 (khaki) and 4 (light blue), occasionally sampling the 24 other clusters (Figure 3.2.5). FLB KstR sampled 20 conformational clusters predominantly residing in cluster 1 (dark green) and cluster 3 (cyan), briefly sampling the 18 other cluster within the conformational pool. This demonstrates that the removal of ligand induces KstR to sample a slightly larger pool of conformational clusters more frequently than ligand-bound KstR. Furthermore, neither system shares the same conformational cluster indicating that ligand removal results in KstR migrating into unique conformational pool.

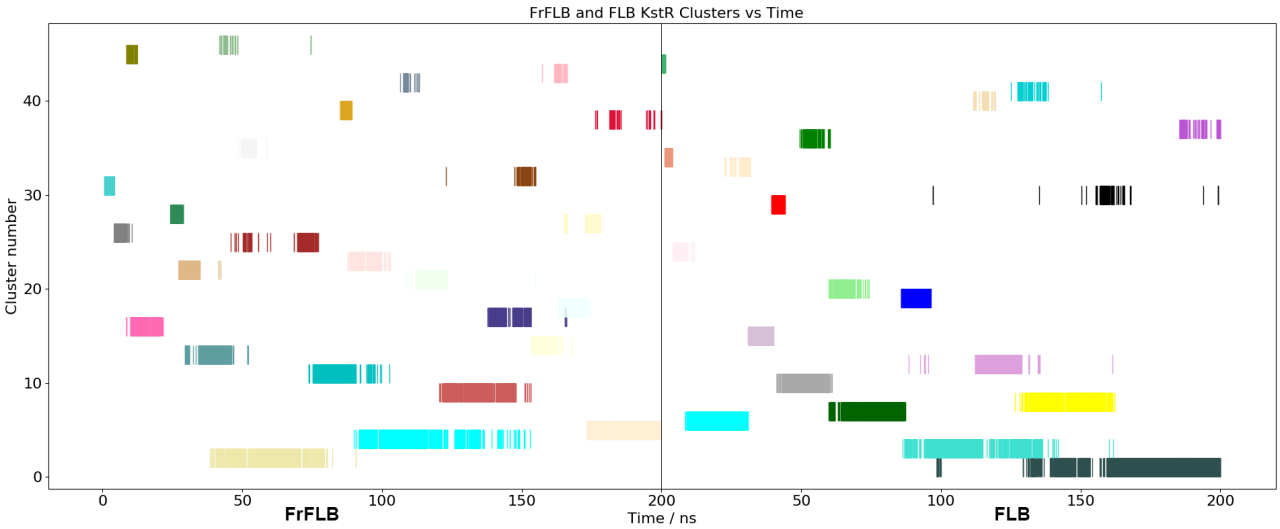

Figure 3.2.5: *46 conformation cluster groups of FrFLB (left), and FLB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with 50 or more conformations in their group with a RMSD cut-off of 0.2 nm.*

Figure 3.2.6 shows that *apo* KstR produces 9 conformational clusters, and predominantly resides within clusters 1 (navy) and 2 (green), while briefly sampling among 7 other conformational clusters. FrFLB KstR presented 26 conformational clusters, it sampled conformational clusters 3 (moss green) and 4 (egg white) most consistently and briefly fluctuated among 24 other conformational clusters. This shows that the removal of ligand caused KstR to fluctuate between more conformational clusters than *apo* KstR.

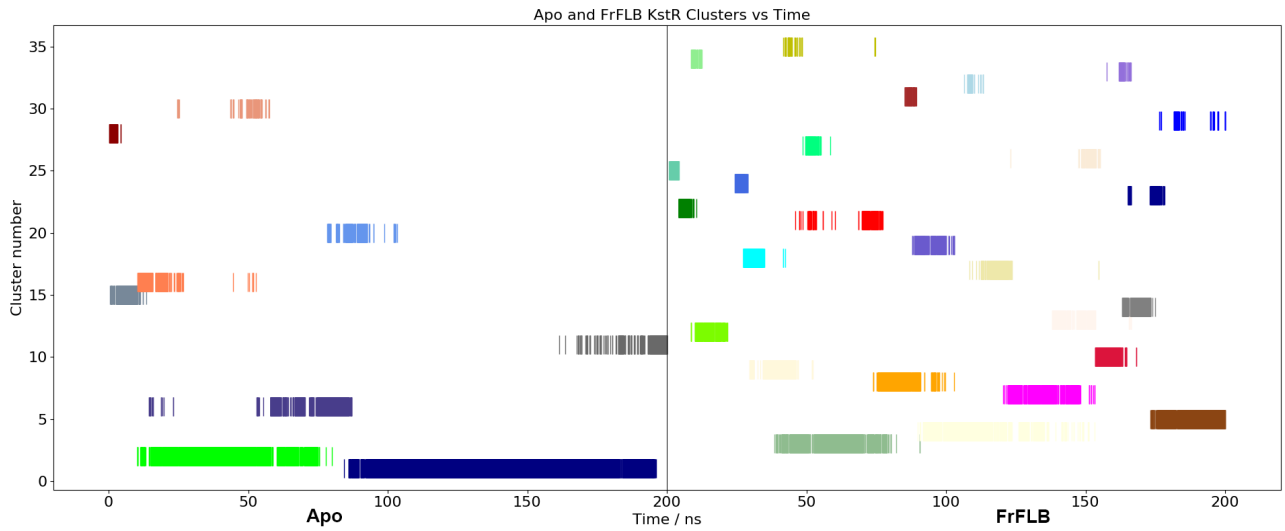

Figure 3.2.6: *35 conformation cluster groups of apo (left), and FrFLB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with 50 or more conformations in their group with a RMSD cut-off of 0.2 nm.*

Overall the removal of ligand causes KstR to sample a conformational pool and behaviour different from ligand-bound and *apo* KstR.

3.2.3 Ligand removal affects the flexibility of KstR

The RMSF for FrFLB was calculated as detailed in Section 2.2.5. The RMSF is a per-residue measurement of flexibility (Figure 3.2.7). Comparison of the RMSF between FLB, FrFLB and *apo* KstR highlights which regions of KstR are perturbed by ligand removal and to what extent. This shows the mechanism of conformational changes which drives the differences seen in Sections 3.2.1 and 3.2.2.

From residue 20 to 30 (α1) the *apo* and FrFLB KstR increases, diverging from FLB KstR. The FrFLB RMSF for residues 30 to 36 (α1 - α2 linker) spikes to sample the largest RMSF. At residue 36 the RMSF of the three systems converge and increase until residue 40. From residue 40 to 48 (α2) the *apo* RMSF is the largest while FrFLB and FLB KstR demonstrate a smaller and similar RMSF. FLB KstR exhibits the largest RMSF between residues 48 to 60 (α3) while *apo* and FrFLB KstR have a similar RMSF. This shows that α1 - α2 linker of KstR becomes more flexible as a result of ligand removal, however α2 itself is unchanged by ligand removal while α3 becomes less flexible upon ligand removal.

The three systems take on their largest RMSF between the α4 - α5 linker domain. The FLB and FrFLB KstR have the same RMSF which dramatically increases from residue 75 - 80 while *apo* KstR RMSF increases from residue 78. FLB and *apo* RMSF peak at about 0.45 nm at residue 82, whereas FrFLB KstR has a larger peak of 0.62 nm which occurs at residue 83. Following these peaks the RMSF of the three systems decreases and converges by residue 90.

From residue 90 to 97 (α5) the RMSF is similar for all three systems. From residue 98 to 105 (α5) the RMSF of *apo* and FrFLB KstR increases while FLB RMSF is lowest. For residues 105 to 115 (α6) the FrFLB and *apo* RMSF decreases. This reveals that the removal of ligand does not impact the length of the α4 - α5 linker domain but increases its flexibility, and does not impact the flexibility of the α5 N-terminus, but impacts the flexibility of the α5 - α6 linker.

The RMSF of residues between 140 and 150 (α7 - α8 linker), demonstrates how ligand affects the flexibility of the helices. The *apo* and FrFLB KstR initially have a similar RMSF and both increase by residue 146, while FLB shows the same trend however by residue 146 has the largest RMSF, although all three systems converge at residue 150. From residue 150 to 165 the RMSF of the three systems is similar. From residue 165 (α8) the RMSF of *apo* KstR increases to until residue 190 (α9 C-terminus), while the RMSF of FrFLB and FLB KstR are similar and smaller. This shows that ligand removal has no impact on any other region of the LBD aside from the α7 - α8 linker where KstR exhibits reduced flexibility following ligand removal.

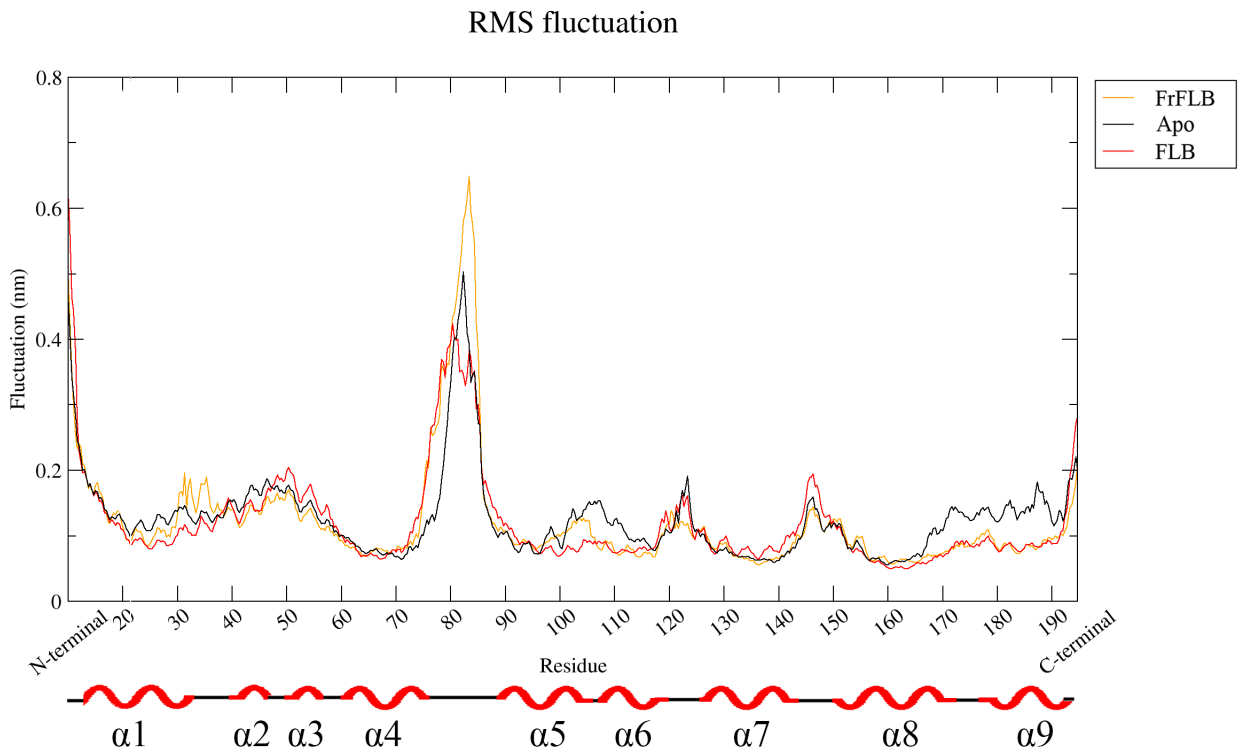

Figure 3.2.7: *Backbone RMSF of KstR chain for FLB (red), FrFLB (orange), and apo (black) KstR. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

Helices within the DBD and about the ligand binding cavity of KstR became more flexible after ligand removal. Ligand removal reduced the flexibility of residues closer to the C-terminus of KstR .

3.2.4 Ligand removal affects how the DBD and LBD of KstR are arranged

The distances between residue pairs for FrFLB KstR were calculated as described in Section 2.2.7. By comparing distances between residues in FLB, FrFLB, and *apo* DBDs how the arrangement of the KstR DBD is altered by ligand removal is shown. The locations of the DBD residue pairs measured are shown in Figure 2.2.2.

Table 3.2.5 shows that the average distance between residue pairs was the largest for FLB KstR, and lowest for *apo* KstR. For all residue pairs either *apo* KstR or FrFLB KstR had the largest standard deviation. For all residue pairs aside from Tyr-54, *apo* KstR sampled the smallest distances. For all residue pairs aside from Arg-17 and Tyr-54, FLB KstR samples the largest average distance.

Together this shows that the removal of ligand alters the conformation of KstR so that the distance between the DBD of each monomer decreased to become more like *apo* KstR. Furthermore, the removal of ligand from KstR caused the distances between residues at the DBD to fluctuate most.

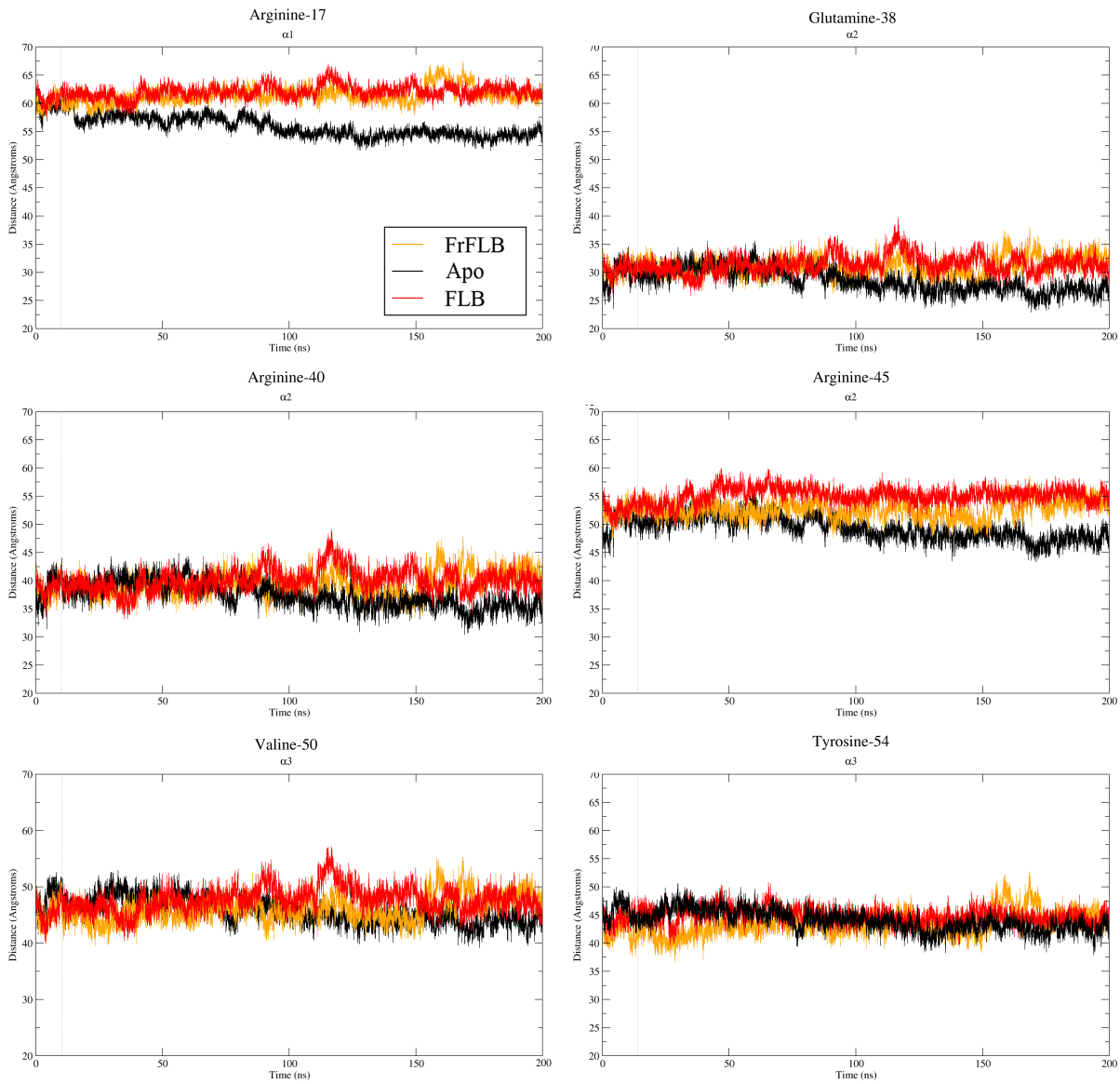

Figure 3.2.8: *Distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 of each DBD of FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.2 for the residue position reference.*

Table 3.2.5: *Average, standard deviation and standard error, and range of distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 for FLB, FrFLB, and apo KstR over time (10 - 200 ns).*

Statistics of Arg-17 distance				
	$\bar{x}$ (Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	62.08	1.16	0.01	57.52 - 67.02 (9.50)
FrFLB	61.45	1.37	0.01	57.44 - 67.37 (9.93)
Apo	55.78	1.59	0.02	51.58 - 63.79 (12.21)
Statistics of Gln-38 distance				
FLB	31.45	1.64	0.02	25.84 - 39.70 (13.86)
FrFLB	30.89	1.56	0.02	26.03 - 37.41 (11.38)
Apo	28.54	2.05	0.02	22.90 - 35.59 (12.69)
Statistics of Arg-40 distance				
FLB	40.09	2.05	0.02	33.19 - 49.07 (15.88)
FrFLB	39.25	1.94	0.02	33.44 - 47.21 (13.77)
Apo	37.36	2.28	0.02	30.41 - 44.83 (14.42)
Statistics of Arg-45 distance				
FLB	55.17	1.33	0.01	49.62 - 59.91 (10.29)
FrFLB	52.62	1.54	0.02	47.65 - 58.55 (16.90)
Apo	49.29	2.03	0.02	43.30 - 55.73 (12.43)
Statistics of Val-50 distance				
FLB	47.81	2.20	0.02	40.37 - 57.14 (16.77)
FrFLB	46.19	2.25	0.02	39.59 - 55.29 (15.70)
Apo	45.68	2.21	0.02	39.19 - 52.97 (13.78)
Statistics of Tyr-54 distance				
FLB	45.47	1.39	0.01	39.63 - 50.75 (11.12)
FrFLB	43.15	1.92	0.02	36.82 - 52.16 (15.34)
Apo	43.96	1.83	0.02	37.94 - 50.59 (12.65)

By comparing distances between residues in FLB, FrFLB, and *apo* LBDs how the arrangement of the KstR LBD is altered by ligand removal is shown. The locations of the LBD residue pairs measured are shown in Figure 2.2.3.

Among the residue pairs at the LBD no distinctive trends specific to a system arise (Figure 3.2.9). For residue pairs Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 the removal of ligand has no discernible impact on the arrangement of the KstR LBD.

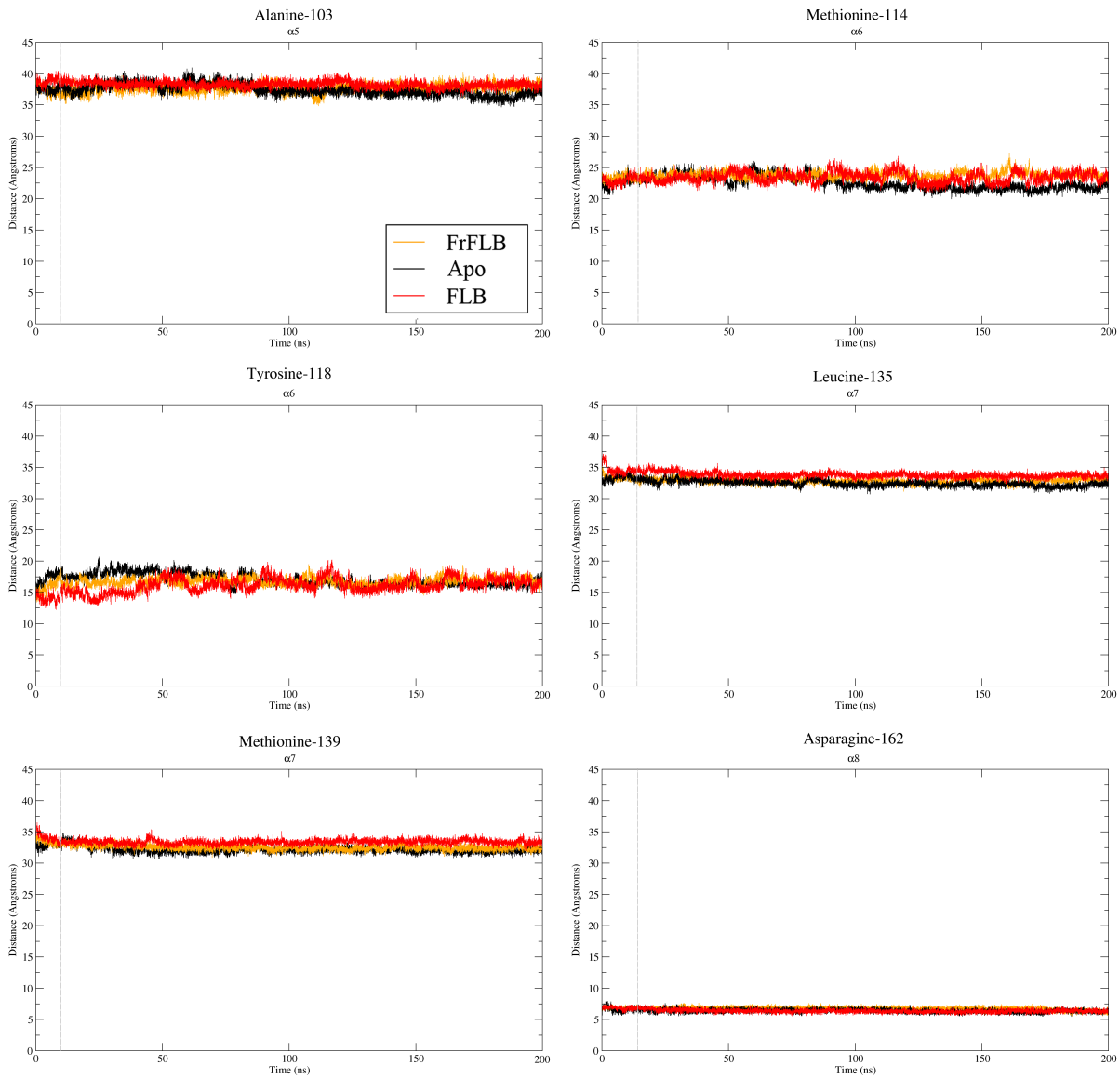

Figure 3.2.9: *Distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 of each LBD of FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.3 for the residue position reference.*

Table 3.2.6: *Average, standard deviation and standard error, and range of distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 for FLB, FrFLB, and apo KstR over time (10 - 200 ns).*

Statistics of Ala-103 distance				
	$\bar{x}$ (Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	38.285	0.519	0.005	36.099 - 40.189 (4.090)
FrFLB	37.645	0.658	0.007	34.826 - 40.082 (5.256)
Apo	37.424	0.929	0.009	34.678 - 40.908 (6.229)
Statistics of Met-114 distance				
FLB	23.45	0.85	0.01	20.98 - 26.77 (5.79)
FrFLB	23.78	0.53	0.01	21.94 - 27.25 (5.31)
Apo	22.57	1.06	0.01	19.98 - 26.01 (6.03)
Statistics of Tyr-118 distance				
FLB	16.22	1.11	0.01	12.88 - 20.19 (7.31)
FrFLB	16.78	0.60	0.01	14.01 - 19.44 (5.43)
Apo	17.01	0.95	0.01	14.77 - 20.66 (5.89)
Statistics of Leu-135 distance				
FLB	33.711	0.518	0.005	32.240 - 36.032 (3.791)
FrFLB	32.821	0.411	0.004	31.152 - 35.391 (4.238)
Apo	32.439	0.471	0.005	30.734 - 34.924 (4.190)
Statistics of Met-139 distance				
FLB	33.387	0.389	0.004	31.976 - 35.364 (3.388)
FrFLB	32.500	0.387	0.004	30.905 - 34.087 (3.181)
Apo	32.136	0.420	0.004	30.624 - 34.732 (4.107)
Statistics of Asp-162 distance				
FLB	6.367	0.231	0.002	5.457 - 7.397 (1.939)
FrFLB	6.806	0.230	0.002	5.739 - 7.647 (1.907)
Apo	6.400	0.285	0.003	5.489 - 7.595 (2.105)

This shows that upon ligand removal the DBD takes on more *apo* KstR like characteristics and that the residue pairs at the LBD do not manifest these differences. However, it FrFLB never fully returns to *apo* KstR like behaviour indicating a longer simulation may have been appropriate (Section 4.0.1).

3.2.5 Ligand removal perturbs the structure of KstR

The inter-residue contact maps for FrFLB were generated as detailed in Section 2.2.9. By comparing the intramolecular contacts of FLB, FrFLB, and *apo* KstR the affects of ligand removal on KstR structure are demonstrated when comparing Figure 3.2.10 to Figure 3.2.11 and Figure 3.2.12.

FrFLB and *apo* KstR have shorter distances between residues 70 to 74 (α4) and residues 94 to 110 (α5) but longer distances between residues 78 to 86 (α4 and α5 linker domain) and itself than FLB KstR. The distances between residues 62 to 74 (α4) and residue 162 to residue 170 (α8) are shortest for *apo* and FrFLB KstR. Residue 82 (α4 and α5 linker) and residues 138 to 146

(α7) are shortest for *apo* KstR while longest for FLB KstR, however FrFLB KstR displays the shortest distances between these residues. This indicates that ligand removal induces shorter distances between α4 and α5, α4 and α8, and the α4 - α5 linker and α7.

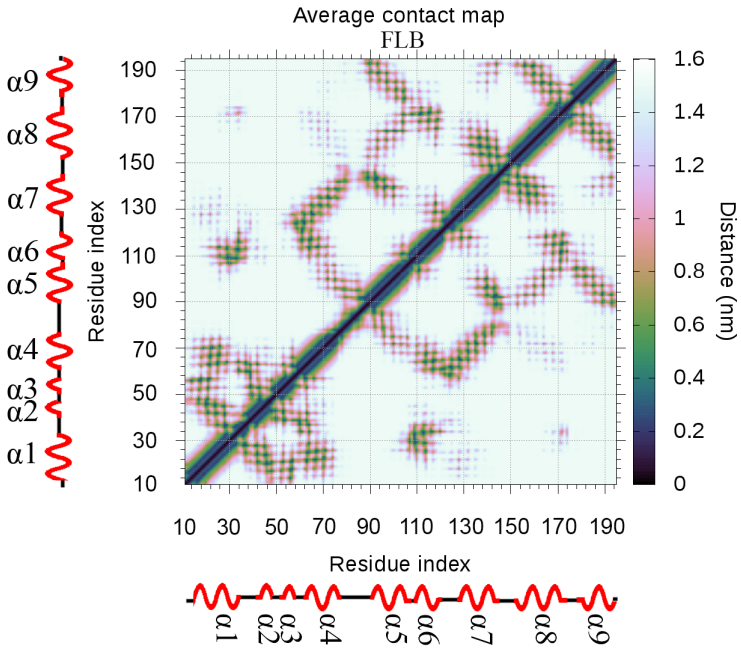

Figure 3.2.10: *Average inter-residue distance of intramolecular interactions of FLB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

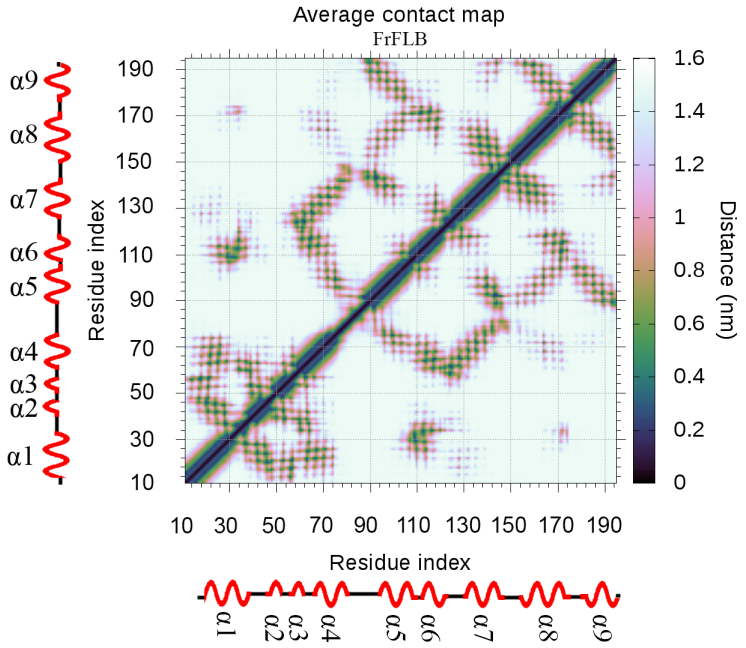

Figure 3.2.11: *Average inter-residue distance of intramolecular interactions of FrFLB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

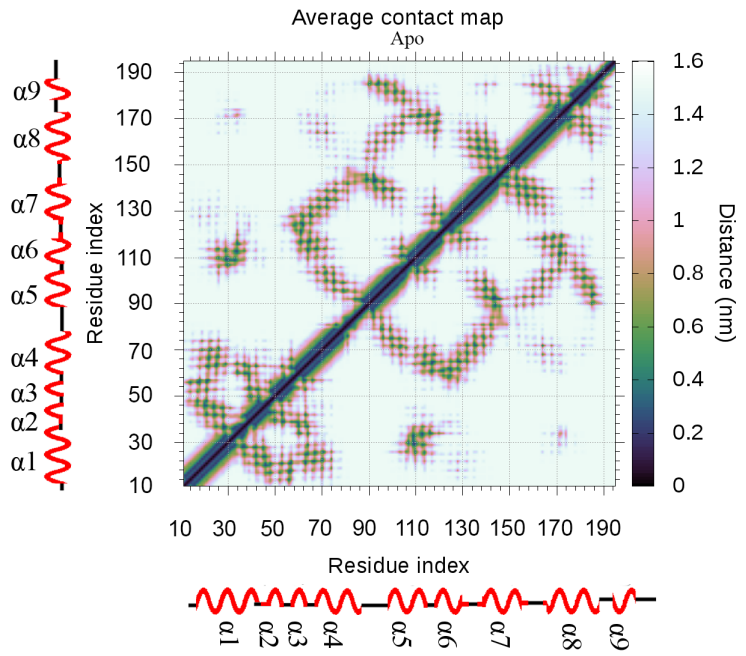

Figure 3.2.12: *Average inter-residue distance of intramolecular interactions of apo KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

Ligand removal resulted in closer distances among the LBDs of KstR.

3.2.6 Ligand removal impacts the secondary structure of KstR

The secondary structure time series for FrFLB KstR was calculated as described in Section 2.2.6. By comparing the secondary structure of FLB, FrFLB, and *apo* KstR how ligand removal affects the secondary structure of KstR of KstR is revealed. The differences in secondary structure as a consequence of ligand removal are highlighted when Figures 3.2.13, 3.2.14 and 3.2.15 are compared.

The linker between $\alpha1$ and $\alpha2$ of FrFLB KstR presents a unique helices produced of 3_{10}-helices. This is not seen in the secondary structure of *apo* or FLB KstR, indicating that the removal of ligand induced residues 31 and 32 to shift from coil to helical conformation. Furthermore, an additional 3_{10} segment composed of residue 33 to 35 was produced. The secondary structure assignment throughout the DBD and LBD is otherwise identical to that of FLB KstR. The removal of ligand unsettled the structure of KstR, causing $\alpha1$ and $\alpha2$ to shorten and causing the formation of a helix within the $\alpha1$ - $\alpha2$ linker.

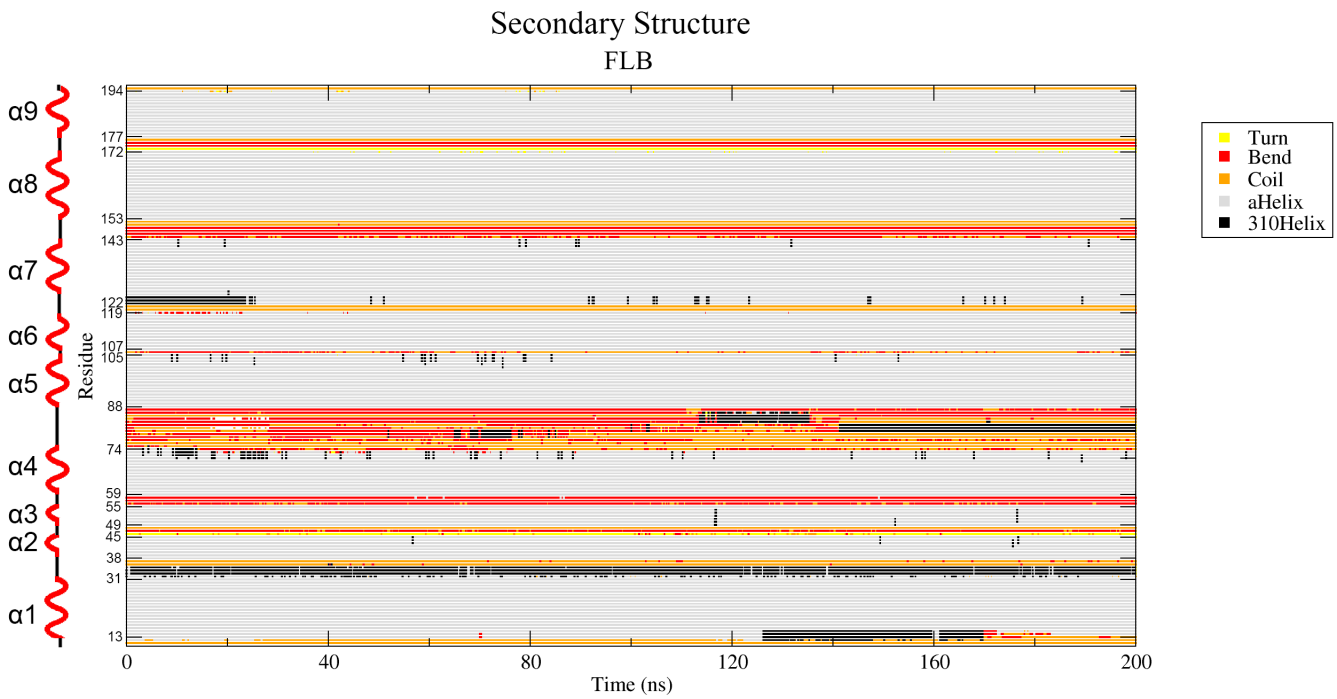

Figure 3.2.13: *Secondary structure assignment of FLB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

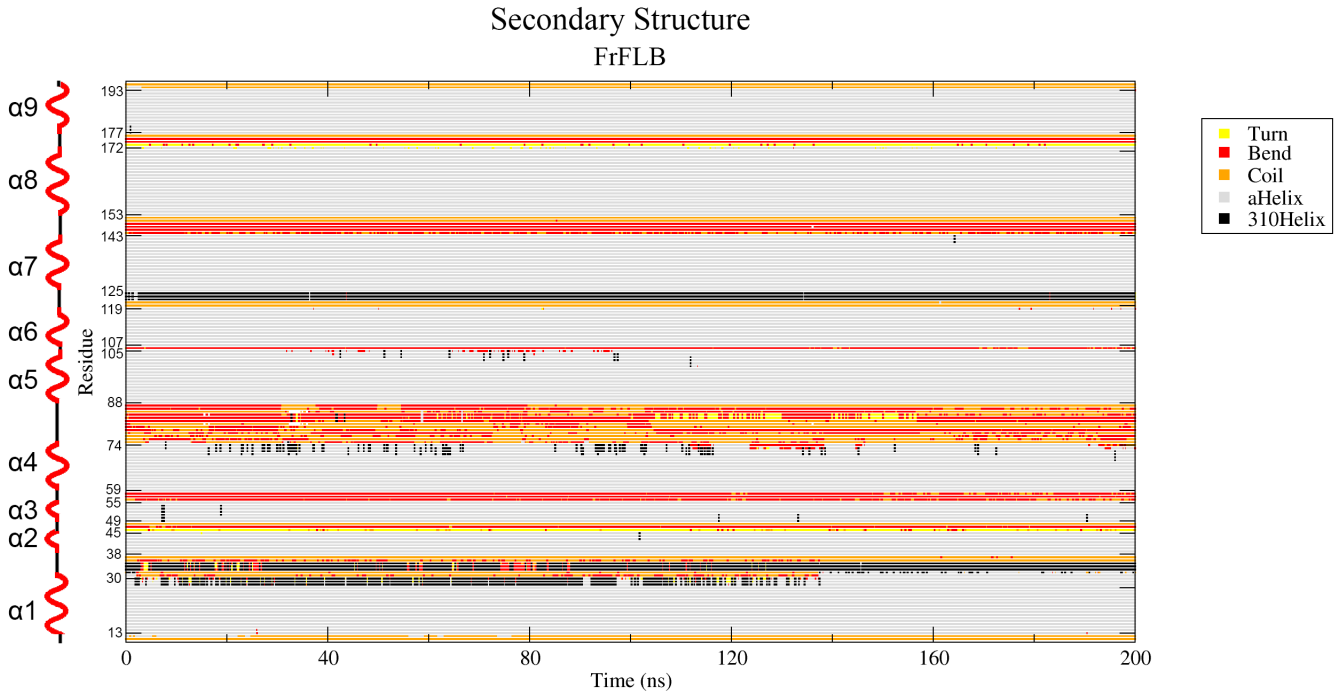

Figure 3.2.14: *Secondary structure assignment of FrFLB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

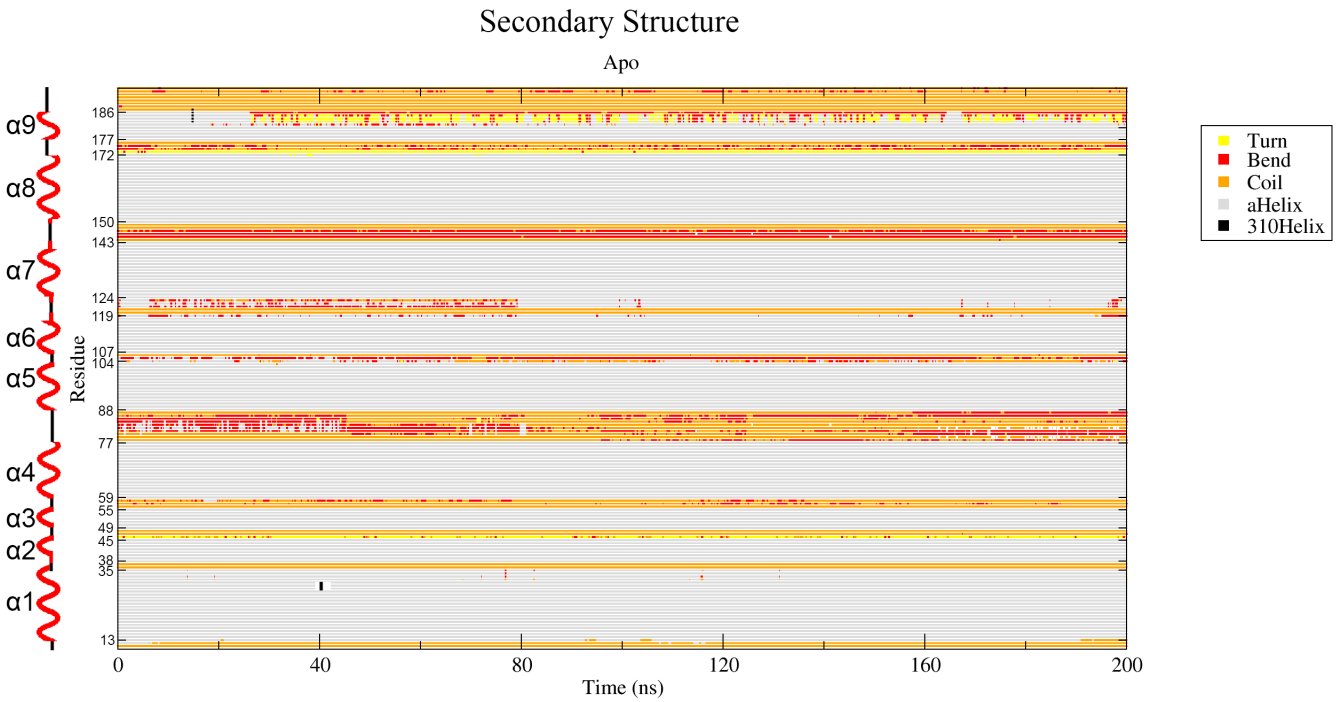

Figure 3.2.15: *Secondary structure assignment of apo KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

KstR is dominated by helical regions, comprising approximately 73.4% of the FrFLB structure (Table 3.2.7). The helical composition of FrFLB is 2% smaller than that of FLB and 2% higher than that of *apo* KstR. This reveals that the removal of ligand induced the transition of helical regions into loop regions. Cumulatively this shows that when ligand is lost the secondary structure of FrFLB partially transitioned towards *apo* KstR secondary structure distribution on the timescale simulated (Section 4.0.1).

Table 3.2.7: *Average secondary structure assignment for FLB, FrFLB, and apo KstR over time (10 - 200 ns).*

Secondary structure distribution of FLB, FrFLB, and *apo* KstR					
	α-helices	Coil	Turn	Bend	3_{10}-helices
FLB	74.3%	11.9%	5.6%	6.3%	1.3%
FrFLB	72.2%	14.3%	5.6%	6.2%	1.2%
Apo	70.5%	14.3%	6.1%	7.9%	1.2%

3.2.7 Ligand removal affects Trp-164 flipping

The Trp-164 dihedral angle for FrFLB KstR was calculated as detailed in Section 2.2.8. Trp-164 is located within the ligand binding cavity and has close contacts with the ligand. By comparing the dihedral angle dynamics of FLB, FrFLB, and *apo* KstR how ligand removal impacts the formation of the residue flipping can be determined, as seen in Figure 3.2.16.

FLB displays a distinct dihedral angle for Trp-164, while FrFLB and *apo* KstR display similar Trp-164 behaviour. Additionally, the FrFLB KstR Trp-164 dihedral angle initially exhibits values within the range of dihedral angles for Trp-164 produced by FLB KstR. However, within the first 1 ns FrFLB Trp-164 flips to take on forms like that *apo* Trp-164. This indicates that ligand removal caused Trp-164 to flip back and adopt *apo* KstR-like Trp-164 conformation.

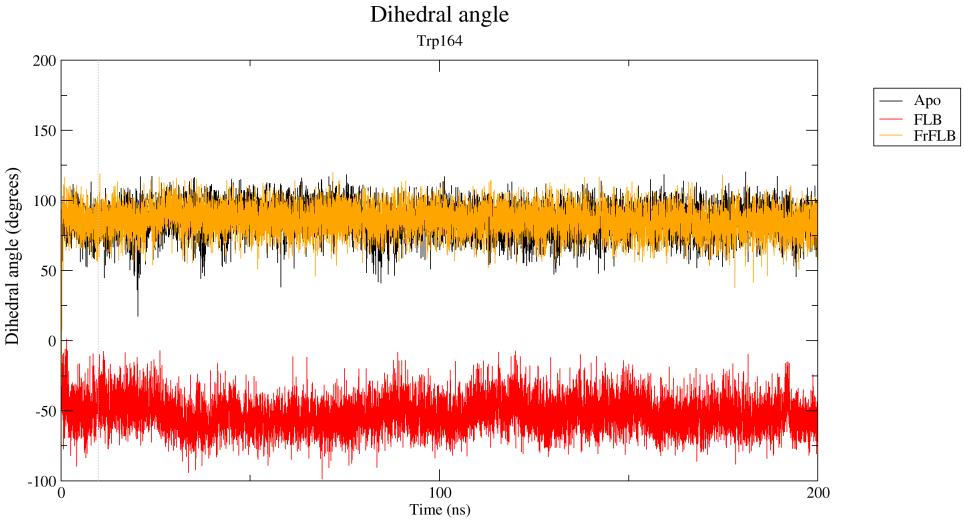

Figure 3.2.16: *Dihedral angle across Trp-164 for FLB (red), FrFLB (orange), and apo (black) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.4 for the residue dihedral reference.*

Table 3.2.8: *Average, standard deviation and standard error, and range of FLB, FrFLB, and apo KstR Trp-164 dihedral angle over time (10 - 200 ns).*

Statistics of Trp-164 dihedral angle				
	$\bar{x}$ (°)	σ (°)	$\sigma_{\bar{x}}$ (°)	Range (max - min)
FLB	-52.9	11.6	0.1	-98.7 - -7.1 (91.6)
FrFLB	85.9	9.7	0.1	-29.7 - 120.0 (149.7)
Apo	86.1	10.5	0.1	17.6 - 120.3 (102.7)

3.3 DNA removal affects KstR conformational dynamics

To examine if DNA removal changes the conformational dynamics of KstR a DNA removed KstR system was prepared and is described in Table 2.2.1. The From Long DNA bound (FrLDB) system was produced by removing DNA from LoDB KstR structure prepared earlier (Section 3.1).

The FrLDB KstR was then solvated as described in Section 2.2.3, and the full atomic composition is detailed in Table 2.2.1. After solvation, the system was equilibrated as described in Section 2.1.4. Finally, the system was simulated for a total of 200 ns as described in Section 2.1.5. Once the simulation had been completed various analysis were conducted as described in Section 2.2.

By comparing the *apo*, FrLDB, and LoDB KstR simulations the aim was to determine how the conformational dynamics of KstR changes upon DNA removal, partially resolving the second objective (Section 1.4).

3.3.1 KstR shape expands after DNA removal

The Rg of FrLDB KstR were calculated as explained in Section 2.2.1. By comparing the Rg of *apo*, FrLDB, and LoDB how the size and shape of KstR is changed by DNA removal is shown.

Apo and FrLDB KstR exhibit significantly larger average Rg than that of LoDB KstR (Table 3.3.1). Furthermore, *apo* KstR presents a larger average Rg standard deviation than FrLDB, but the Rg standard deviation of FrLDB is larger than that of LoDB KstR. The standard error indicates that these differences are significant. Figure 3.3.1 shows that within the first 2 ns the FrLDB KstR Rg increases from within the LoDB KstR Rg range to that of *apo* KstR. This shows that the removal of DNA causes KstR to transition rapidly to a larger and more dynamic shape.

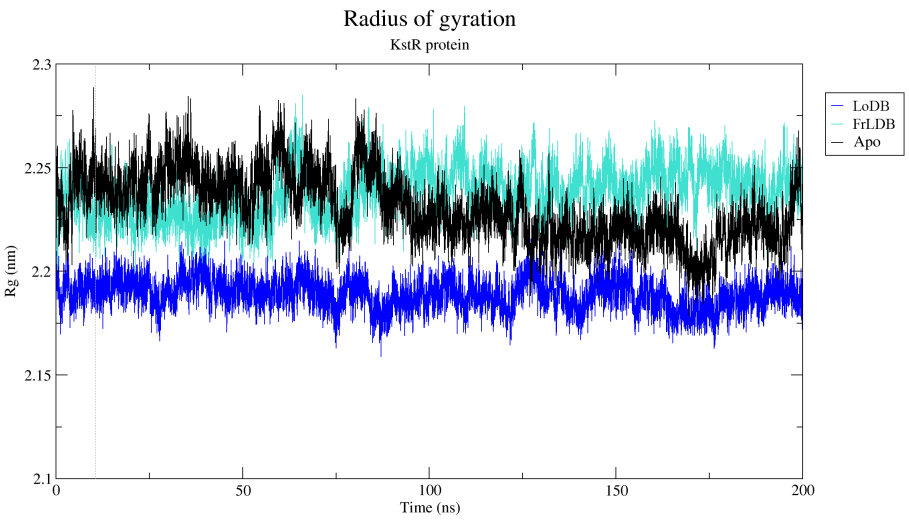

Figure 3.3.1: *KstR Rg of apo (black), FrLDB (cyan), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.3.1: *Average, standard deviation, standard error, and range of apo, FrLDB, and LoDB KstR Rg over time (10 - 200 ns).*

Statistics of KstR Rg				
	$\bar{x}(nm)$	$\sigma(nm)$	$\sigma_{\bar{x}}$ (nm)	Range (max - min)
Apo	2.230	0.016	0.002	2.168 - 2.288 (0.119)
FrLDB	2.237	0.012	0.000	2.191 - 2.285 (0.094)
LoDB	2.189	0.008	0.000	2.158 - 2.216 (0.058)

The SASA of FrLDB KstR was calculated as described in Section 2.2.2. By comparing the SASA of *apo*, FrLDB, and LoDB KstR how the size and shape of KstR is impacted by DNA removal is highlighted.

Apo and FrLDB KstR presented a significantly larger average SASA than that of LoDB (Table 3.3.2). FrLDB KstR had a much larger SASA standard deviation than that of *apo* KstR and LoDB. The standard error also indicates that these differences are significant. At the beginning of the simulations, the SASA of FrLDB and LoDB KstR are both about 188 nm^2, but by 10 ns the LoDB SASA decreases to about 180 nm^2 while FrLDB KstR increases to above 190 nm^2 (Figure 3.2.2). Over the course of the simulation *apo*, FrLDB, and LoDB SASA sample slightly overlapping SASA ranges, with *apo* KstR and FrLDB sampling SASA within a similar range of 175 - 198 nm^2, whereas the LoDB SASA ranges from 167 - 187 nm^2. This shows that LoDB KstR has different SASA to *apo* and FrLDB KstR. The removal of DNA causes the SASA of KstR to rapidly increase and fluctuate more, like that of *apo* KstR. Once DNA-free it is excluded from sampling the range of SASA values nearby to LoDB.

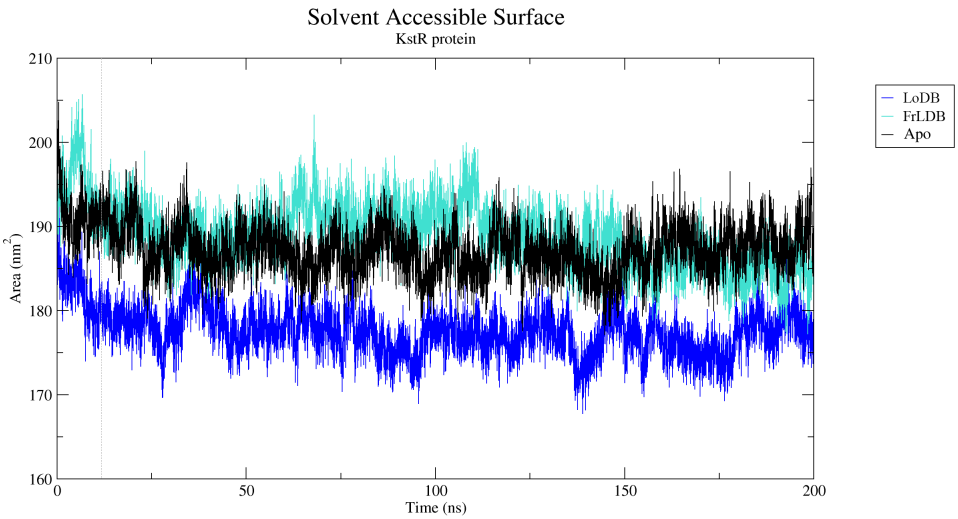

Figure 3.3.2: *KstR SASA of apo (black), FrLDB (cyan), and LoDB (blue) systems over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.3.2: *Average, standard deviation and standard error, and range of apo, FrLDB, and LoDB KstR SASA over time (10 - 200 ns).*

Statistics of KstR SASA				
	$\bar{x}(nm^2)$	$\sigma(nm^2)$	$\sigma_{\bar{x}}$ (nm^2)	Range (max - min)
Apo	187.08	2.69	0.03	175.22 - 197.71 (22.49)
FrLDB	185.79	3.01	0.03	175.88 - 198.00 (22.12)
LoDB	177.39	2.58	0.03	167.76 - 187.85 (20.09)

The volume of FrLDB KstR was calculated as explained in Section 2.2.2. By comparing the volume of *apo*, FrLDB, and LoDB KstR how the size of KstR is impacted by DNA removal is highlighted.

The FrLDB KstR had the largest average volume while LoDB KstR had the lowest (Table 3.3.3). While the difference between the two averages is slight the standard error indicates it is significant. Figure 3.3.3 shows that the volume of all three systems heavily overlap, however LoDB presents the smallest average in keeping with trends of the *R*g and SASA values. This demonstrates that DNA removal has a minor affect on the volume of KstR.

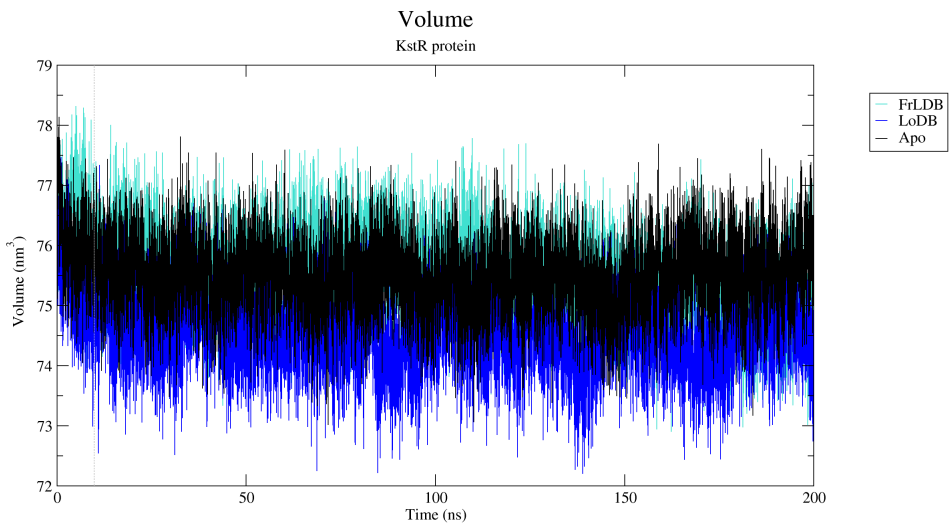

Figure 3.3.3: *Volume of apo (black), FrLDB (cyan), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.3.3: *Average, standard deviation and standard error, and range of apo, FrLDB, and LoDB KstR volume over time (10 - 200 ns).*

Statistics of KstR volume				
	$\bar{x}(nm^3)$	$\sigma(nm^3)$	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
Apo	75.488	0.632	0.007	72.986 - 77.809 (4.823)
FrLDB	75.529	0.693	0.007	72.904 - 78.002 (5.098)
LoDB	74.490	0.629	0.007	72.209 - 77.334 (5.125)

The *R*g and SASA information indicate that upon DNA removal the overall shape of KstR became larger. Overall, the data indicate that KstR quickly becomes larger and fluctuates more upon the removal of DNA.

3.3.2 DNA removal alters the conformational pool of KstR

The conformational ensembles of *apo*, FrLDB and LoDB KstR were calculated as detailed in Section 2.2.4. By clustering conformations between LoDB, FrLDB, and *apo* KstR systems, how DNA removal impacts KstR conformational dynamics is highlighted. This analysis presents information at a more detailed level than the global indicators used in Section 3.3.1, and the rate of transition between different conformational ensembles is revealed.

Figure 3.3.4 demonstrates that FrLDB samples a much larger number of different conformation ensembles than LoDB and each ensemble has a shorter lifetime. LoDB samples 4 clusters, past the 5 ns it predominantly resides within cluster 1 (ocean-green). The FrLDB KstR fluctuates more among 28 clusters,

none of which it samples consistently over the simulation timescale. Neither LoDB or FrLDB KstR sample the same cluster. This shows that LoDB KstR shares no conformations with FrLDB KstR, and the removal of DNA causes KstR to change conformations more and sample a larger population of conformations.

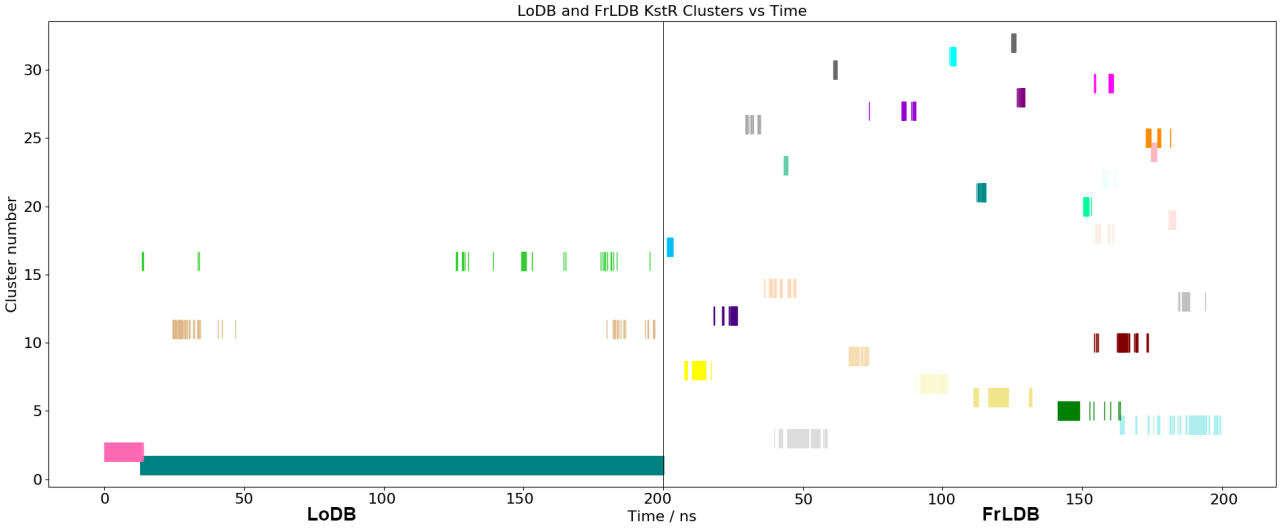

Figure 3.3.4: *32 conformation cluster groups of LoDB (left), and FrLDB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with more than 50 conformations in their group with a RMSD cut-off of 0.2 nm.*

FrLDB KstR samples a more conformational ensembles than *apo* KstR as shown in Figure 3.3.5. *Apo* KstR predominantly samples clusters 1 (orange) and 2 (grey) and infrequently samples the 9 other conformations. FrLDB KstR dynamically fluctuates between its 28 conformational clusters. None of the FrLDB KstR clusters are sampled by *apo* KstR and vice versa. This indicates that once DNA is removed KstR does not demonstrate similar conformational pools to *apo* KstR on this timescale and fluctuates more dynamically. This is unexpected but is discussed in Section 4.0.1.

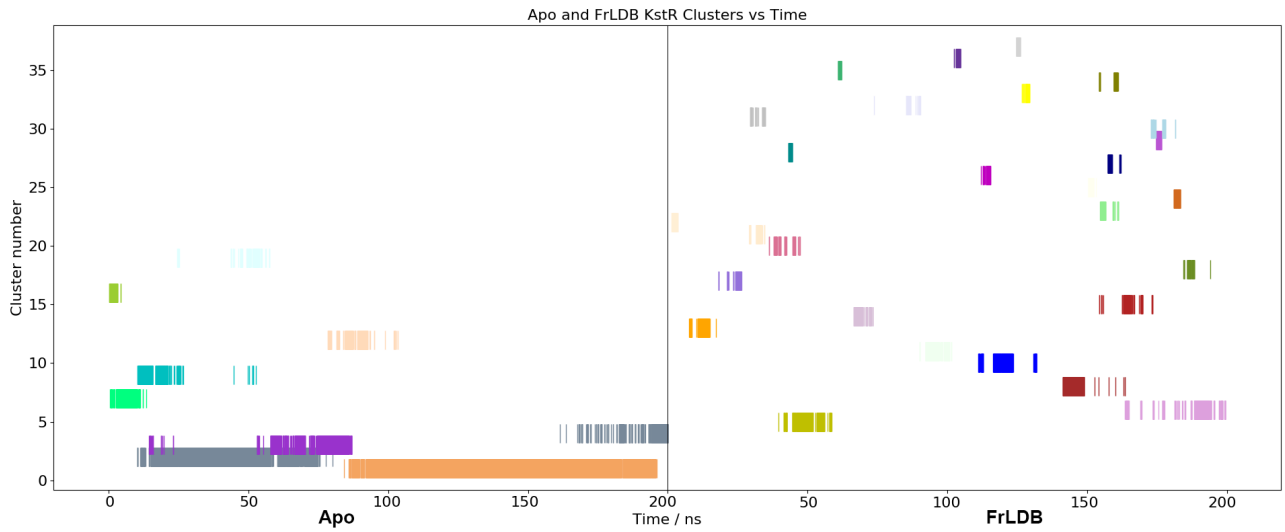

Figure 3.3.5: *37 conformation cluster groups of apo (left), and FrLDB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with more than 50 conformations in their group with a RMSD cut-off of 0.2 nm.*

Both conformational clusters demonstrate that DNA removal causes KstR to take on a larger conformational ensemble. The rapid transitions of FrLDB KstR between conformations compared to the stability of *apo* KstR indicates that KstR is unsettled by the removal of DNA. Similar to Section 3.2.2, given a longer time scale perhaps the conformational clusters of DNA removed KstR and *apo* KstR might begin to overlap (Section 4.0.1).

3.3.3 DNA removal affects flexibility of KstR

The RMSF for FrLDB KstR was calculated as detailed in Section 2.2.5. The RMSF reports at a per-residue level on KstR flexibility (Figure 3.3.6). Comparison of the RMSF between *apo*, FrLDB and LoDB KstR reveals how the flexibility of KstR is perturbed by DNA removal, and hints at the nature of the conformational changes that underlie the differences in Rg, SASA, volume and conformational clustering.

Figure 3.3.6 shows that from residue 10 to residue 30 (α1) there are differences between the RMSF of *apo*, FrLDB and LoDB KstR. *Apo* KstR throughout α1 has a larger RMSF values compared to those of LoDB, while FrLDB KstR exhibits intermediate RMSF values. Throughout residue 30 to 70 (α2 - α4) the FrLDB RMSF is similar to that of *apo* KstR while LoDB KstR exhibits a lower RMSF. The RMSF of the three systems converges at around residue 70. These results indicate that the flexibility of the DBD is directly influenced by DNA binding, with its flexibility increasing markedly upon removal of DNA.

The *apo* RMSF through residues 70 to 90 (α4 - α5 linker domain) shows novel features compared to that of FrLDB and LoDB KstR. The FrLDB and LoDB RMSF peak at about 0.36 nm at residue 80 and 82 respectively, while *apo* KstR peaks at 0.5 nm at residue 82. This indicates that the removal of

DNA does not impact the flexibility of the $\alpha4$ - $\alpha5$ linker domain although it extends the domain into $\alpha5$, akin to *apo* KstR.

The $\alpha5$ has similar flexibility in all three systems, *apo* KstR has the largest RMSF for residues 95 to 130 ($\alpha5$ - $\alpha7$ N-terminus). This indicates that DNA removal causes KstR to become more flexible but it does not become as flexible as *apo* KstR at $\alpha5$, $\alpha6$, and $\alpha7$.

From residue 140 ($\alpha7$) to 160 ($\alpha8$ C-terminus) the RMSF of *apo*, FrLDB and LoDB KstR are indiscernible. From residue 160 to 190 ($\alpha8$ - $\alpha9$) *apo* KstR has a larger RMSF than FrLDB and LoDB KstR. This shows that DNA removal has no impact on the flexibility of $\alpha7$, $\alpha8$, through $\alpha8$ to $\alpha9$, nor the flexibility nearer to the C-terminus of KstR.

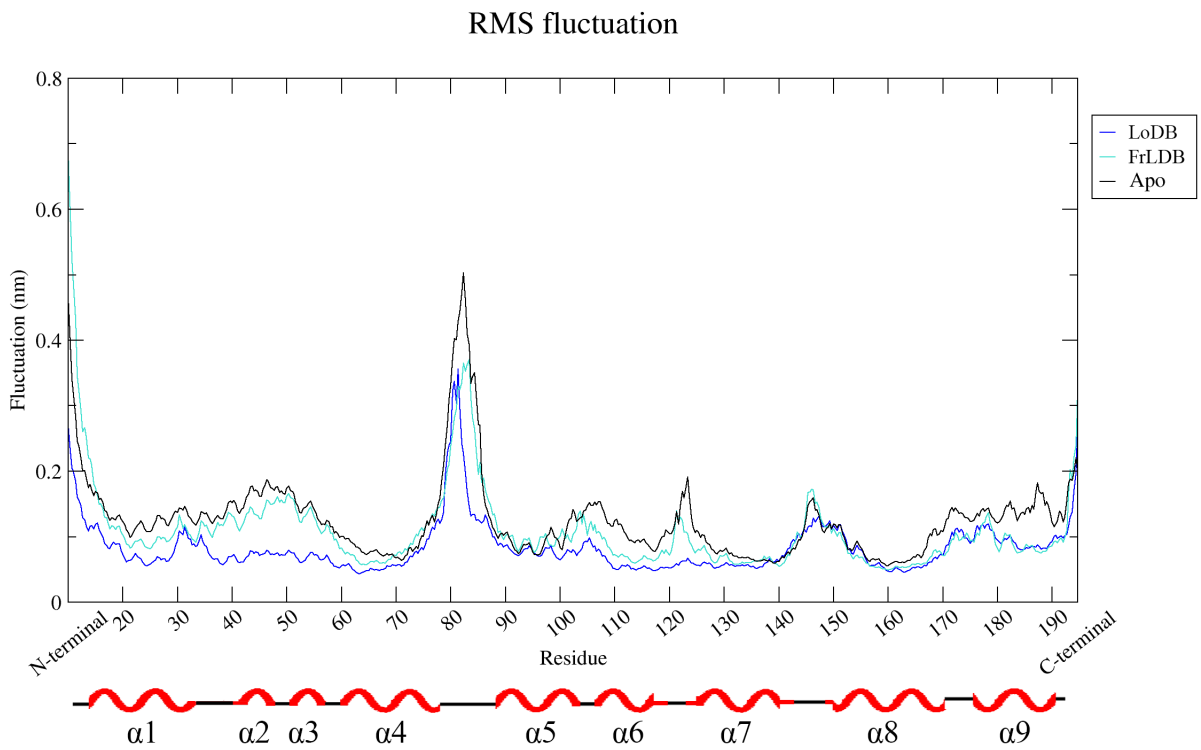

Figure 3.3.6: *Backbone RMSF of KstR chain for apo (black), FrLDB (cyan), and LoDB (blue) KstR. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

DNA removal caused the entire DBD of KstR to become more flexible. However DNA removal did not impact the LBD aside from through $\alpha5$ - $\alpha7$. This shows that the effects of DNA removal on KstR flexibility are not propagated throughout the KstR structure.

3.3.4 DNA removal affects how the DBD and LBD of KstR are arranged

The distances between residues pairs for FrLDB KstR were calculated as described in Section 2.2.7. By comparing distances between residues in *apo*, FrLDB, and LoDB DBDs how the arrangement of the KstR DBD is altered by DNA removal is shown. The locations of the DBD residue pairs measured are shown in Figure 2.2.2.

Table 3.3.4 shows that for all residue pairs aside from Tyr-54, the FrLDB KstR distances are similar to the LoDB distances at the start of the simulation, but increase to become similar to the *apo* KstR distances during the course of the simulation. For all residue pairs *apo* KstR has the largest average distance, followed by FrLDB KstR, and then LoDB KstR is the smallest. For all residue pairs *apo* KstR has the largest standard deviation followed by FrLDB KstR, and then LoDB KstR has the smallest. For all residue pairs other than Tyr-54, the range of FrLDB distances lies within the ranges of *apo* and LoDB KstR distances, while the *apo* and LoDB KstR distance ranges do not always overlap (Figure 3.3.7).

Overall, this data indicates that upon the removal of DNA, KstR alters its conformation such that the distance between the DBD of each monomer increases to become more like *apo* KstR.

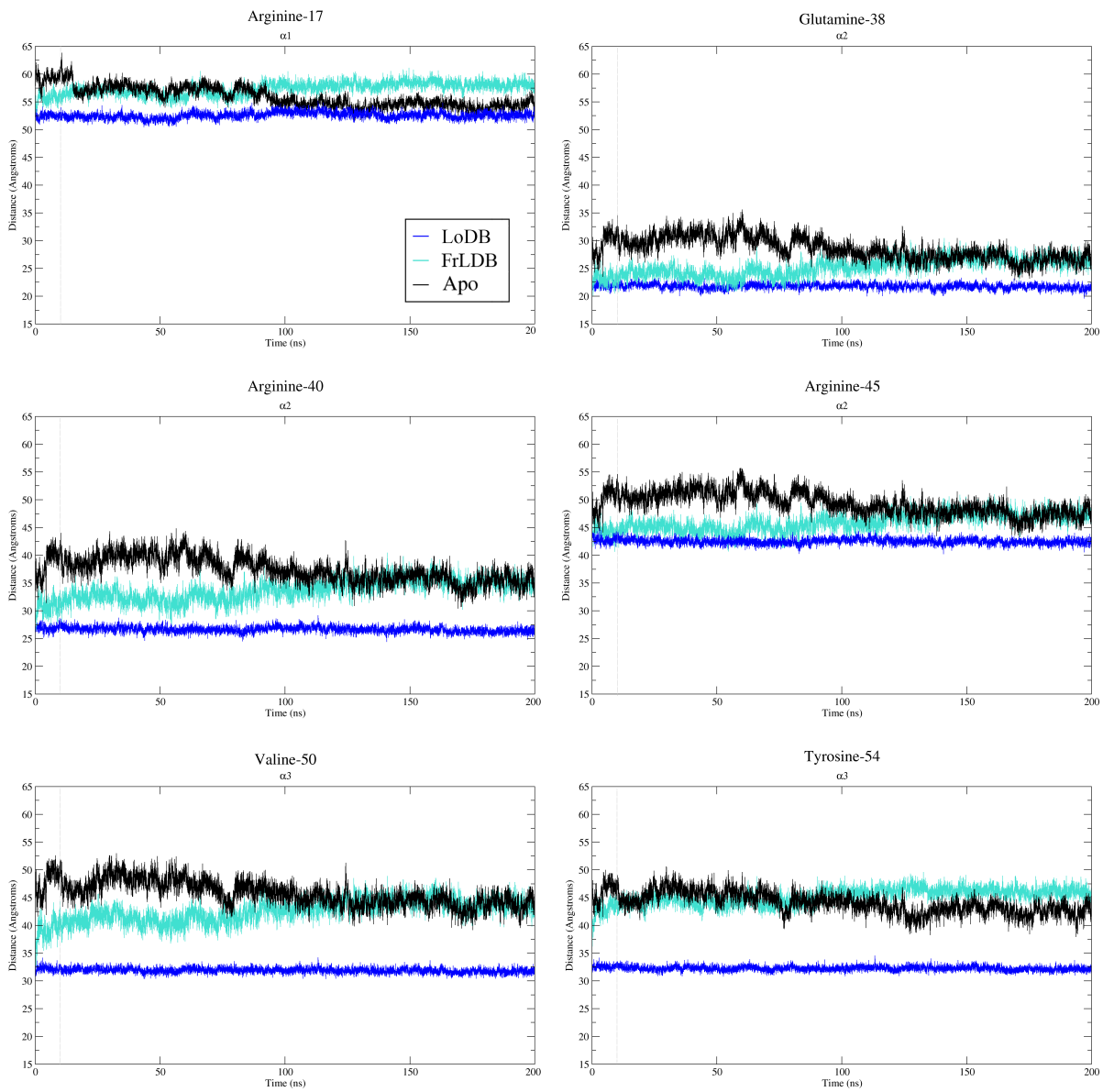

Figure 3.3.7: *Distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 of each chain of apo (black), FrLDB (cyan), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.2 for the residue position reference.*

Table 3.3.4: *Average, standard deviation and standard error, and range of distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 for apo, FrLDB, and LoDB KstR over time (10 - 200 ns).*

Statistics of Arg-17 distance				
	$\bar{x}$ (Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
Apo	55.78	1.59	0.02	51.58 - 63.79 (12.21)
FrLDB	57.43	1.09	0.01	51.44 - 61.08 (9.64)
LoDB	52.58	0.61	0.01	50.47 - 55.19 (4.72)
Statistics of Gln-38 distance				
Apo	28.54	2.05	0.02	22.90 - 35.59 (12.69)
FrLDB	25.30	1.52	0.02	20.75 - 30.45 (9.70)
LoDB	21.78	0.47	0.01	19.61 - 23.39 (3.78)
Statistics of Arg-40 distance				
Apo	37.36	2.28	0.02	30.41 - 44.82 (14.41)
FrLDB	33.79	1.81	0.02	27.17 - 40.43 (13.26)
LoDB	26.63	0.53	0.05	24.45 - 29.13 (4.68)
Statistics of Arg-45 distance				
Apo	49.29	2.03	0.02	43.30 - 55.73 (12.43)
FrLDB	46.02	1.55	0.02	40.69 - 50.95 (10.26)
LoDB	42.49	0.51	0.01	40.25 - 44.53 (4.28)
Statistics of Val-50 distance				
Apo	45.68	2.21	0.02	39.19 - 52.97 (13.78)
FrLDB	42.73	1.84	0.02	32.41 - 48.86 (16.45)
LoDB	31.89	0.46	0.01	30.39 - 34.19 (3.80)
Statistics of Tyr-54 distance				
Apo	43.96	1.83	0.02	37.94 - 50.59 (12.65)
FrLDB	45.31	1.13	0.01	36.61 - 49.55 (12.94)
LoDB	32.25	0.42	0.04	30.79 - 34.53 (3.74)

By comparing distances between residues in *apo*, FrLDB, and LoDB LBDs how the arrangement of the KstR LBD is altered by DNA removal is shown. The locations of the LBD residue pairs measured are shown in Figure 2.2.3.

There are no consistent trends among the residue pairs at the LBD. While the distance between Tyr-118 residues shows similar larger distances for FrLDB and *apo* compared to LoDB, as seen for the DBD distances, the opposite is seen for Ala-103, with the LoDB and *apo* KstR exhibiting similar large distances (Table 3.3.5). This demonstrates that the removal of DNA has little impact on the organisation of the LBD (Figure 3.3.8).

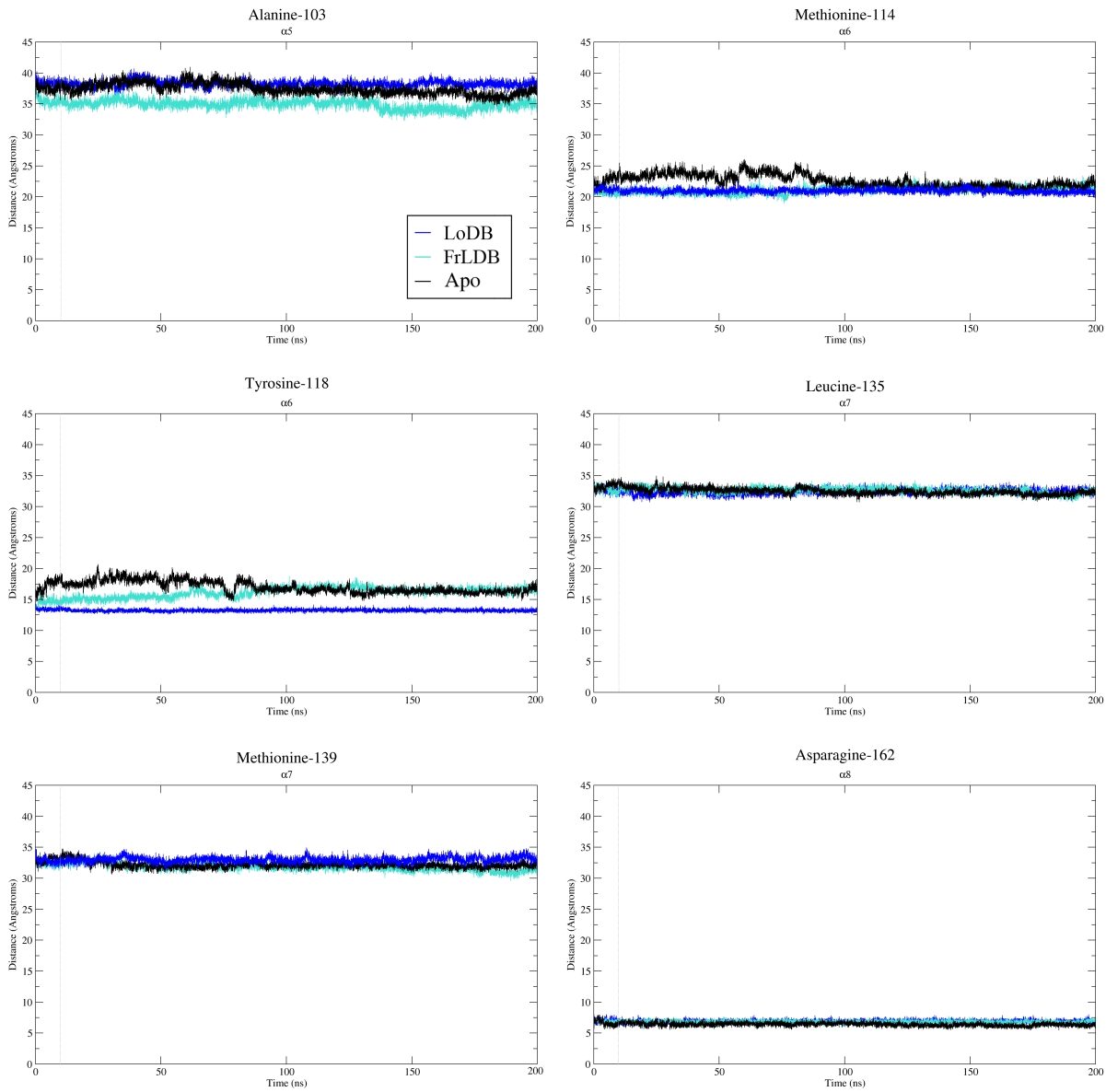

Figure 3.3.8: *Distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 of each chain of apo (black), FrLDB (cyan), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.3 for the residue position reference.*

Table 3.3.5: *Average, standard deviation and standard error, and range of distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 for apo, FrLDB, and LoDB KstR over time (10 - 200 ns).*

Statistics of Ala-103 distance				
	$\bar{x}$(Å)	σ(Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
Apo	37.424	0.929	0.009	34.678 - 40.908 (6.229)
FrLDB	34.888	0.729	0.008	32.326 - 37.249 (4.923)
LoDB	38.234	0.553	0.006	36.369 - 40.655 (4.285)
Statistics of Met-114 distance				
Apo	22.57	1.06	0.01	19.98 - 26.01 (6.03)
FrLDB	21.11	0.56	0.01	18.91 - 23.37 (4.46)
LoDB	20.94	0.35	0.00	19.63 - 22.55 (2.92)
Statistics of Tyr-118 distance				
Apo	17.011	0.950	0.009	14.765 - 20.661 (8.895)
FrLDB	16.188	0.715	0.007	13.276 - 18.792 (5.515)
LoDB	13.224	0.188	0.002	12.597 - 14.161 (1.563)
Statistics of Leu-135 distance				
Apo	32.439	0.471	0.005	30.734 - 34.924 (4.190)
FrLDB	32.629	0.475	0.005	30.684 - 34.187 (3.503)
LoDB	32.403	0.430	0.004	30.725 - 33.934 (3.208)
Statistics of Met-139 distance				
Apo	32.136	0.420	0.004	30.624 - 34.732 (4.107)
FrLDB	31.753	0.532	0.006	29.621 - 34.0838 (4.462)
LoDB	32.998	0.471	0.005	31.323 - 34.936 (3.612)
Statistics of Asp-162 distance				
Apo	6.400	0.285	0.003	5.489 - 7.595 (2.105)
FrLDB	6.856	0.196	0.002	5.612 - 7.731 (2.119)
LoDB	6.902	0.217	0.002	6.014 - 7.963 (1.948)

The removal of DNA caused the distance between pairs of residues at the DBD to become larger. Furthermore, the removal of DNA caused distances between the DBD of each monomer to increase towards the *apo* state distances and to fluctuate more. Removal of DNA does not impact distances between the LBD consistently, in keeping with the RMSF data (Section 3.3.3).

3.3.5 DNA removal perturbs the structure of KstR

The inter-residue contact maps for FrLDB KstR were produced as detailed in Section 2.2.9. Comparison of the inter-residue distance maps for *apo*, FrLDB and LoDB KstR identifies how the inter-residue structure of KstR is affected by DNA removal. Figure 3.3.9, 3.3.10 and 3.3.11 demonstrate this and give insight into the networks of inter-residue contacts and how KstR might be affected by DNA removal.

Between residues 34 to 46 (α2) and 58 to 70 (α4) FrLDB KstR displays longer distances than LoDB KstR, similar to *apo* KstR distances. The dis-

tances between 50 to 58 (α3) and 110 to 130 (α6 - α7) are shorter for FrLDB and *apo* KstR than LoDB KstR. FrLDB KstR possesses longer distances between residues 30 to 38 (α1 - α2) and 170 to 174 (α9) compared to LoDB KstR, although its distances in this region is similar to *apo* KstR. There are more contacts between residues 110 to 118 (α6) and 170 to 190 (α9) for *apo* KstR than for FrLDB and LoDB KstR. This demonstrates that upon removal of DNA the distances between α2 and α4, α3 and α6 - α7, α1 - α2 and α9 are longer, while having no effect on distances between α6 and α9.

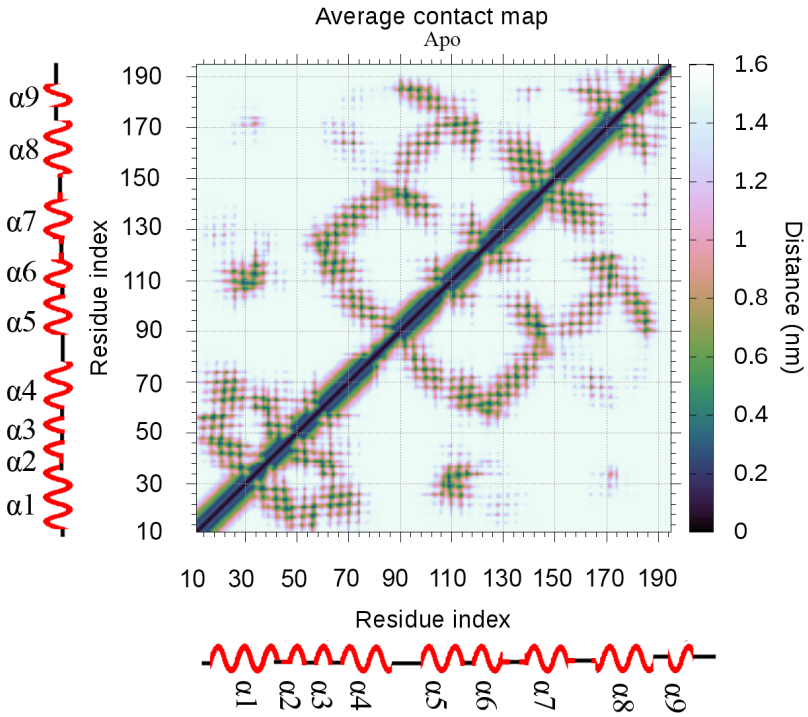

Figure 3.3.9: *Average inter-residue distance of intramolecular interactions of apo KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

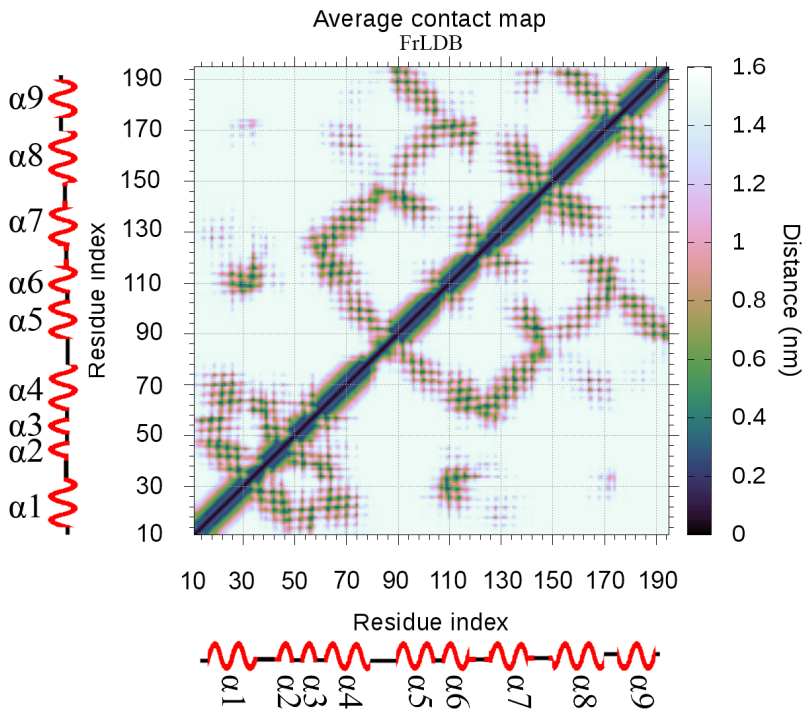

Figure 3.3.10: *Average inter-residue distance of intramolecular interactions of FrLDB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

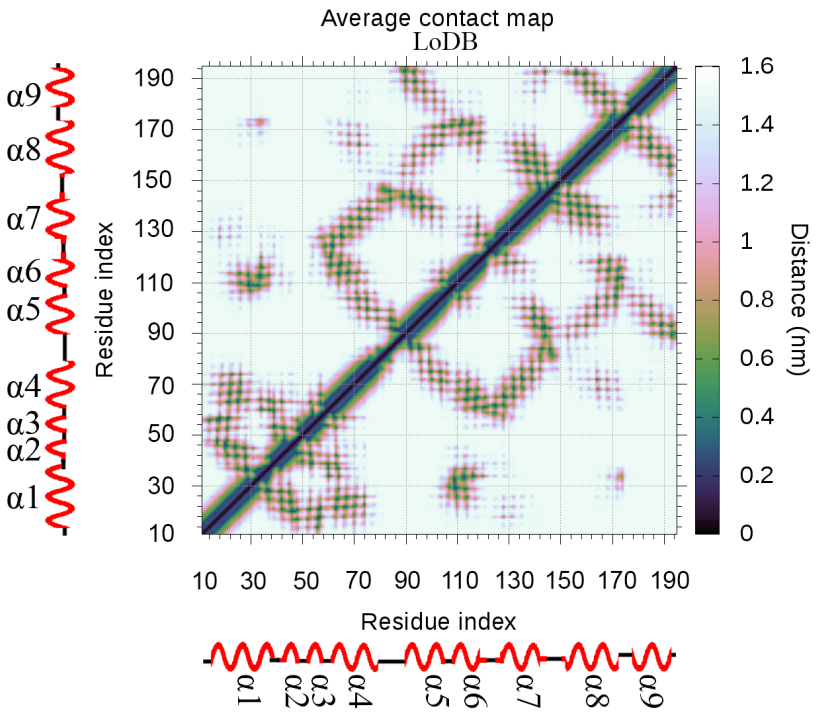

Figure 3.3.11: *Average inter-residue distance of intramolecular interactions of LoDB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

Removal of DNA resulted in fewer and more distant intramolecular con-

tacts across KstR.

3.3.6 DNA removal impacts the secondary structure of KstR

The secondary structure time series for FrLDB KstR was calculated as detailed in Section 2.2.6. By comparing the secondary structure of *apo*, FrLDB, and LoDB KstR how DNA removal affects the secondary structure of KstR is revealed. The differences in secondary structure as a consequence of DNA removal are highlighted when Figures 3.3.12, 3.2.13, and 3.2.14 are compared.

Compared to *apo* KstR and LoDB KstR, FrLDB KstR has a shorter $\alpha 1$ helix and the N- and C-terminus are assigned 3_{10} helices. The FrLDB KstR has a shorter $\alpha 4$ than LoDB and *apo* KstR. The $\alpha 4$ helix for *apo* and LoDB KstR is 17 long residues while for FrLDB it is 16. FrLDB KstR displays a longer $\alpha 4$ - $\alpha 5$ linker domain than LoDB KstR, akin to *apo* KstR assignment at the region. This indicates that the removal of DNA disturbs the structure of $\alpha 1$ and $\alpha 4$, and causes the $\alpha 4$ - $\alpha 5$ linker domain to become more flexible.

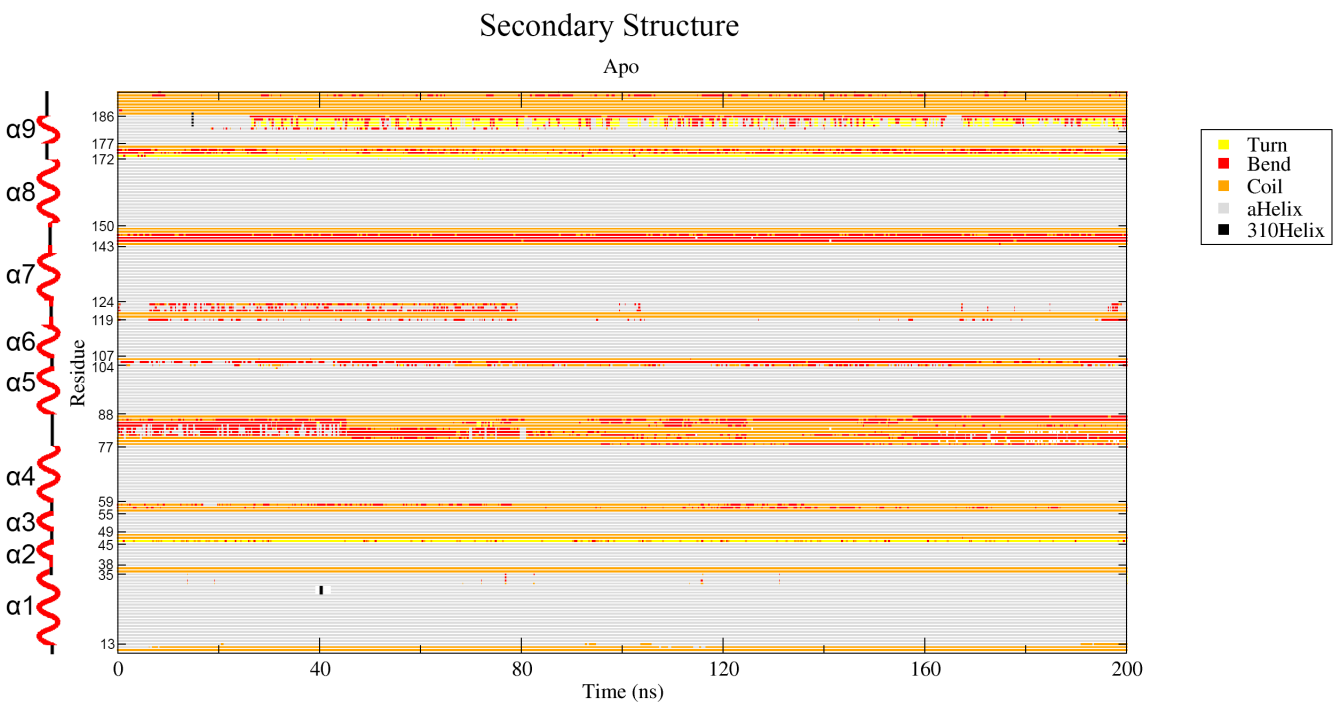

Figure 3.3.12: *Secondary structure assignment of apo KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

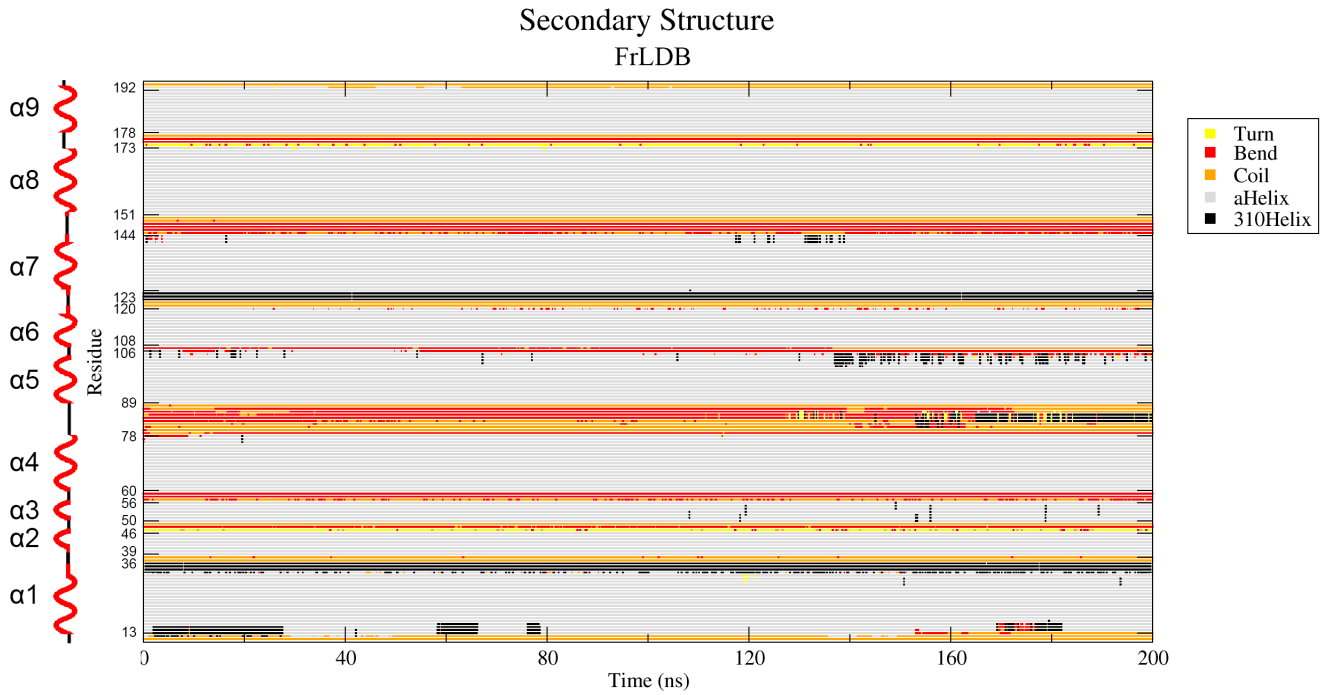

Figure 3.3.13: *Secondary structure assignment of FrLDB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

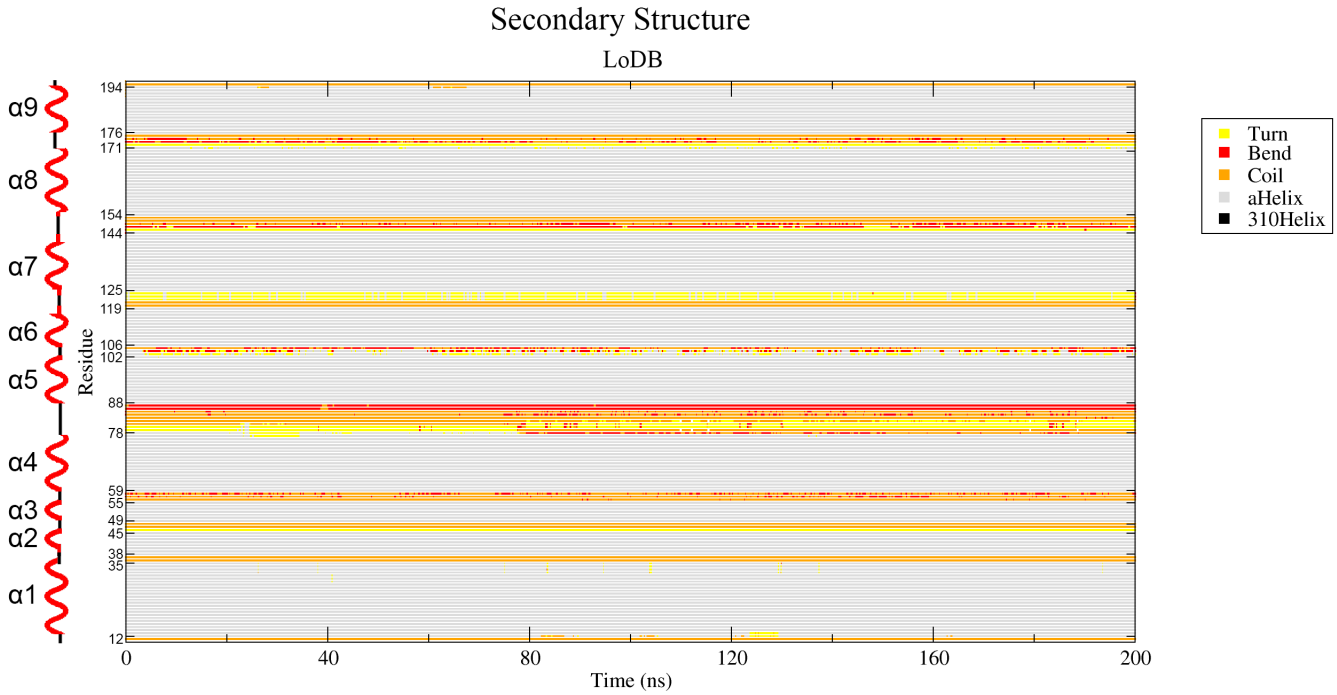

Figure 3.3.14: *Secondary structure assignment of LoDB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

The overall FrLDB secondary structure assignment represents an intermediate between *apo* KstR and LoDB KstR. LoDB KstR is the most helical (78%), followed by FrLDB (76.2%) and *apo* (71.7%) KstR (Table 3.3.6). This indicates that the removal of DNA causes KstR to transition from LoDB KstR to *apo* KstR secondary structure.

Table 3.3.6: *Average secondary structure assignment for apo, FrLDB, and LoDB KstR over time (10 - 200 ns).*

Secondary structure distribution of *apo*, FrLDB, and LoDB KstR					
	α-helices	Coil	Turn	Bend	3_{10}-helices
Apo	70.5%	14.3%	6.1%	7.9%	1.2%
FrLDB	74.0%	11.8%	4.8%	6.7%	2.2%
LoDB	76.6%	8.7%	5.3%	7.5%	1.4%

3.3.7 DNA removal affects Tyr-54 flipping

The Tyr-54 dihedral angle for FrLDB KstR was calculated as detailed in Section 2.2.8. Tyr-54 is located at the DBD and has close contacts with DNA. By comparing the dihedral angle of *apo*, FrLDB, and LoDB Tyr-54 KstR how DNA affects the Tyr-54 dihedral angle is revealed.

There are four states of Tyr-54 sampled across the three systems. FrLDB and *apo* KstR Tyr-54 sample three of these states which it dynamically samples between, while LoDB KstR samples the other state consistently over the course of the simulation (Figure 3.3.18). In addition the Tyr-54 dihedral standard deviation for LoDB KstR is much smaller than that of FrLDB KstR and *apo* KstR (Table 3.3.7). Together this shows that when DNA is bound to KstR Tyr-54 takes on a significantly different and more frequently changing form.

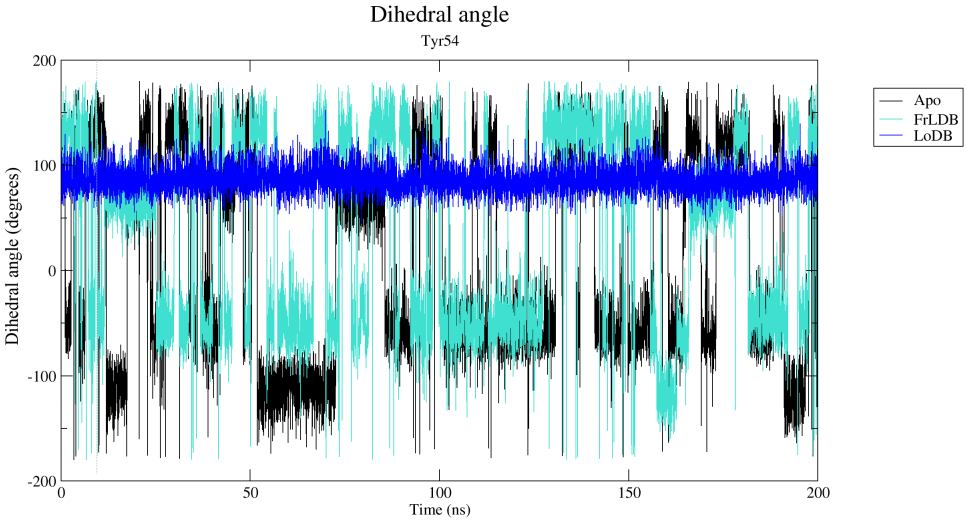

Figure 3.3.15: *Dihedral angle across Tyr-54 for apo (black), FrLDB (cyan), and LoDB (blue) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.4 for the residue dihedral reference.*

Table 3.3.7: *Average, standard deviation and standard error, and range of apo, FrLDB, and LoDB KstR Tyr-54 dihedral angle over time (10 - 200 ns).*

Statistics of Tyr-54 dihedral angle				
	$\bar{x}$ (°)	σ (°)	$\sigma_{\bar{x}}$ (°)	Range (max - min)
Apo	6.9	95.1	0.9	-178.8 - 179.9 (358.7)
FrLDB	32.4	90.5	0.9	-179.4 - 179.4 (358.8)
LoDB	86.4	11.9	0.1	45.6 - 152.4 (106.8)

3.4 Differences in ligand size impact KstR conformational dynamics

To investigate if ligand size alters the conformational dynamics of KstR a shorter ligand bound to KstR was prepared. The short ligand bound (SLB) is comprised of 3OChA bound in the ligand binding cavity of KstR (as seen in crystal structure PDB: 5CW8) and represents a short ligand-bound KstR.

The SLB KstR was then solvated as described in Section 2.2.3, and the full atomic composition is detailed in Table 2.2.1. After solvation, the system was equilibrated as described in Section 2.1.4. Finally, the system was simulated for a total of 200 ns as described in Section 2.1.5. Once the simulation had been completed various analysis were conducted as described in Section 2.2

By comparing the FLB, and SLB KstR simulations the aim was to determine how the conformational dynamics of KstR is altered by ligand size, partially resolving the third objective (Section 1.4)

3.4.1 Ligand size does not impact the overall shape of KstR

The Rg of SLB KstR was calculated according to Section 2.2.1. By comparing the Rg of FLB and SLB KstR how ligand size affects the shape and size of KstR is shown.

The SLB KstR exhibits a larger KstR Rg on average compared to FLB (Table 3.4.1). The FLB KstR samples a bigger range than SLB KstR. The FLB KstR produced a greater average Rg standard deviation than SLB. While the standard error indicates all these differences are significant they are extremely small differences Although, Figure 3.4.1 shows a different trend for both systems, the overall behaviour of KstR Rg is not altered by the size of ligand. SLB and FLB KstR have similar shapes and sizes despite differences in the size of ligand bound.

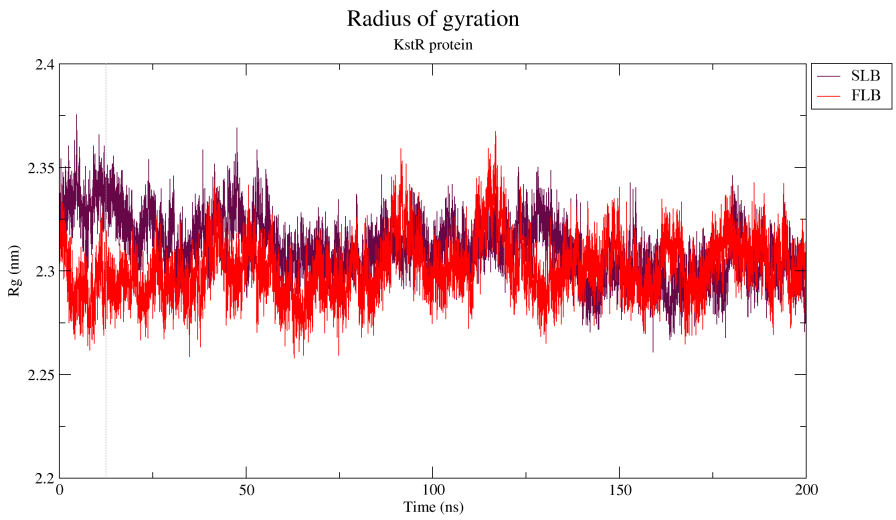

Figure 3.4.1: *KstR Rg of FLB (red) and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.4.1: *Average, standard deviation, standard error, and range of FLB and SLB KstR Rg over time (10 - 200 ns).*

Statistics of KstR Rg				
	$\bar{x}(nm)$	$\sigma(nm)$	$\sigma_{\bar{x}}$ (nm)	Range (max - min)
FLB	2.301	0.014	0.000	2.257 - 2.367 (0.109)
SLB	2.310	0.013	0.000	2.260 - 2.369 (0.108)

The SASA of SLB KstR was calculated according to Section 2.2.2. By comparing the SASA of FLB and SLB KstR how ligand size affects the shape and size of KstR is shown.

The FLB KstR exhibited a smaller SASA than SLB KstR on average (Table 3.4.2). The SLB KstR has a larger range than FLB KstR although they heavily overlap (Figure 3.4.2). The FLB KstR maintained a greater average SASA standard deviation compared to SLB KstR. The standard error shows that these differences are all significant although only slightly Despite Figure 3.4.2 displaying slight trend differences between the two systems, the size of the ligand scarcely changed the overall KstR shape.

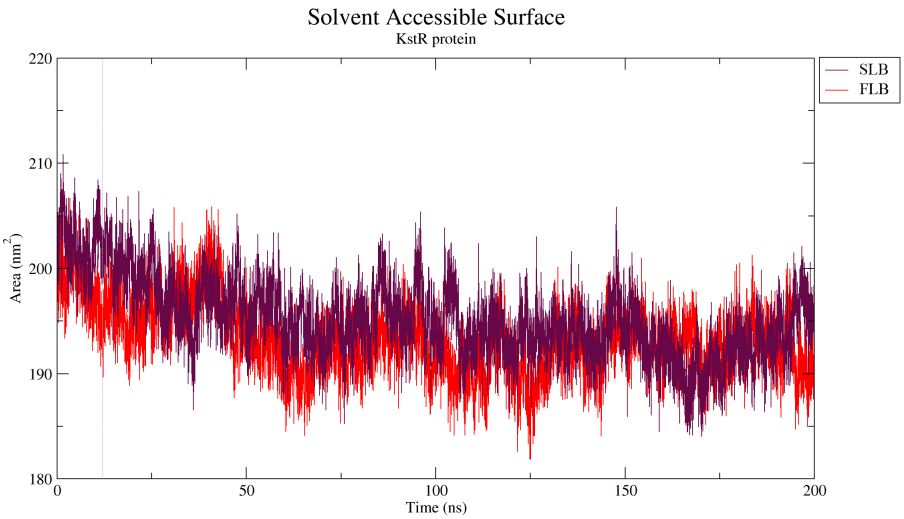

Figure 3.4.2: *KstR SASA of FLB (red) and SLB (maroon) KstR systems over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.4.2: *Average, standard deviation and standard error, and range of FLB and SLB KstR SASA over time (10 - 200 ns).*

Statistics of KstR SASA				
	$\bar{x}(nm^2)$	$\sigma(nm^2)$	$\sigma_{\bar{x}}$ (nm^2)	Range (max - min)
FLB	193.05	3.18	0.03	181.87 - 205.87 (24.00)
SLB	194.86	3.16	0.03	184.04 - 208.41 (24.37)

The volume of SLB KstR was calculated according to Section 2.2.2. By comparing the volume of FLB and SLB KstR how ligand size affects the size of KstR is shown.

The SLB KstR produces a larger average volume than FLB KstR (Table 3.4.3) and both volume ranges overlap heavily (Figure 3.3.3). This volume is significantly different, as evidenced by the standard error. This demonstrates that the size of ligand bound to KstR has a slight impact on KstR size.

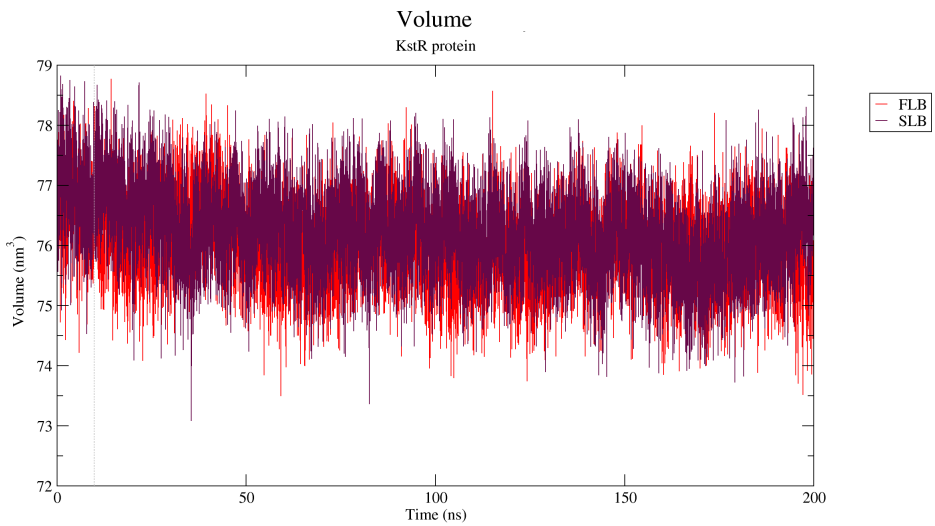

Figure 3.4.3: *Volume of FLB (red), and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.4.3: *Average, standard deviation and standard error, and range of FLB SLB KstR volume over time (10 - 200ns).*

Statistics of KstR volume				
	$\bar{x}(nm^3)$	$\sigma(nm^3)$	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
FLB	76.004	0.679	0.007	73.497 - 78.769 (5.272)
SLB	76.209	0.683	0.007	73.089 - 78.704 (5.615)

The ligand binding cavity volume of SLB KstR was calculated according to Section 2.2.10. By comparing the ligand binding cavity volume of FLB and SLB KstR how ligand size affects the size of the ligand binding cavity is shown.

The FLB average ligand binding cavity volume is larger than that of SLB KstR (Table 3.4.4). The FLB KstR presented a slightly larger ligand binding cavity average volume standard deviation than SLB KstR. The SLB KstR exhibited a greater range of ligand binding cavity volumes compared to FLB KstR and also samples the minimum ligand binding cavity volumes and maximum. The standard error indicates that the differences detailed are significant. This demonstrates that while a shorter ligand induced a smaller ligand binding cavity volume it also changed less often, and has a larger range of ligand binding cavity volumes. However, Figure 3.4.4 displayed that the differences in ligand binding cavity volume trend are slight.

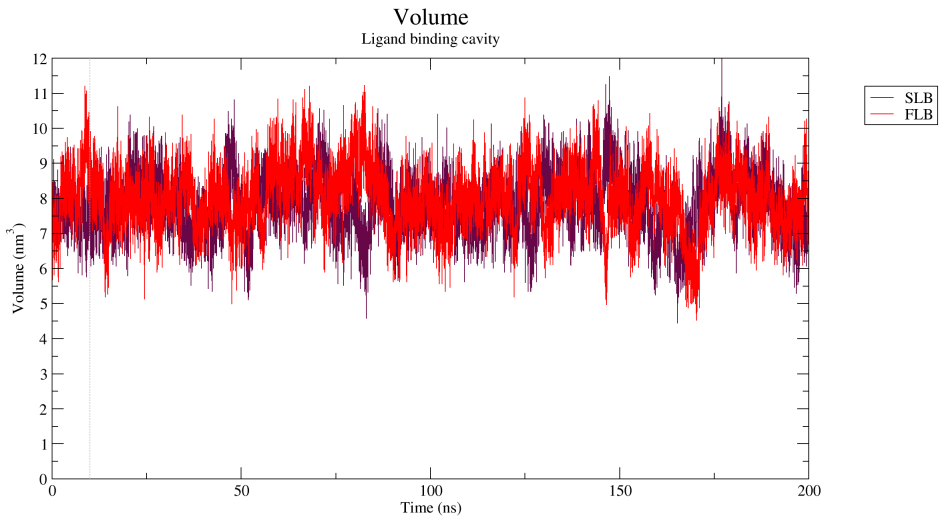

Figure 3.4.4: *Ligand binding cavity volume of FLB (red), and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.4.4: *Average, standard deviation and standard error, and range of FLB and SLB of KstR average ligand binding cavity volume over time (10 - 200 ns).*

Statistics of KstR ligand binding cavity volume				
	$\bar{x}$ (nm^3)	σ (nm^3)	$\sigma_{\bar{x}}$ (nm^3)	Range (min-max)
FLB	7.998	0.892	0.009	4.528 - 11.223 (6.996)
SLB	7.823	0.885	0.009	4.445 - 12.074 (7.628)

Together this shows that a shorter ligand induced a slightly less dynamic KstR shape and a slightly smaller and less dynamic ligand binding cavity.

3.4.2 Ligand size does not alter the conformational pool of KstR

The conformational ensemble of FLB and SLB KstR were produced as described in Section 2.2.4. By comparing the conformational ensembles between FLB and SLB KstR how ligand size impacts the conformational freedom of KstR is highlighted, as shown in Figure 3.4.5. This provides conformational information to a finer detail while also supplementing the findings of Section 3.4.1.

The SLB KstR samples 31 clusters which it frequently fluctuates among while FLB KstR samples 19 clusters however primarily cluster 1 (swamp green) and 2 (blue). This demonstrates that the breadth of conformational pool of KstR becomes much larger and fluctuates more when bound to a shorter ligand.

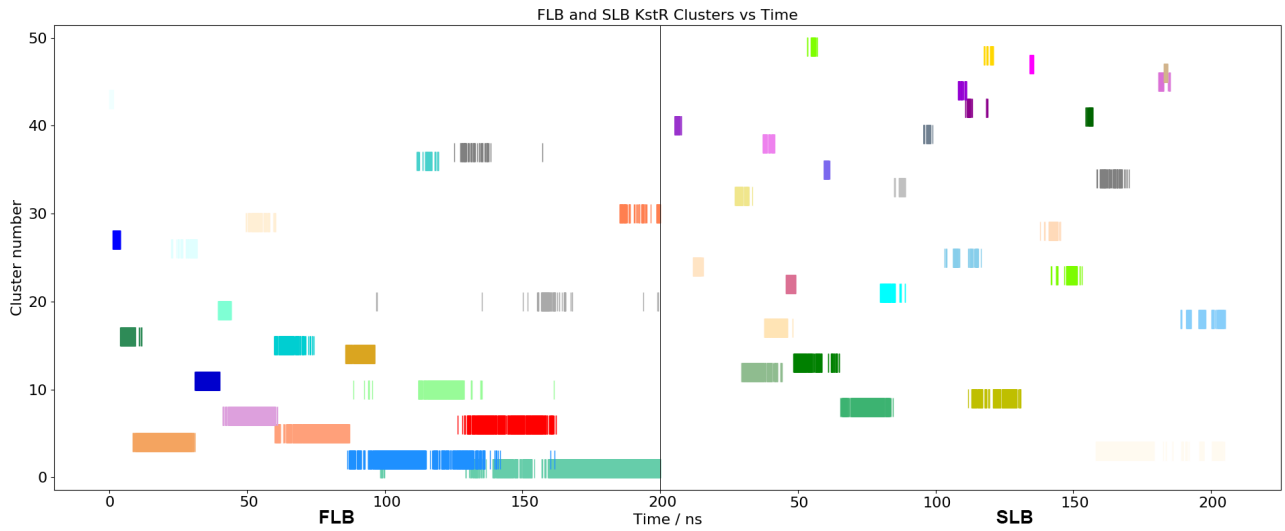

Figure 3.4.5: *49 conformation cluster groups of SLB (left), and FLB KstR (right) over time (0 - 200 ns). Presentation limited to clusters with more than 50 conformations in their group with a RMSD cut-off of 0.2 nm.*

3.4.3 Ligand size impacts the flexibility of KstR

The RMSF of SLB KstR was calculated as detailed in Section 2.2.5. By comparing the FLB and SLB RMSF how ligand size affects the flexibility of KstR is demonstrated. Furthermore, measurements of the KstR RMSF indicate which regions of KstR change in flexibility as a response to ligand shortening and proposes how the changes seen in Section 3.4.1 and 3.4.2 were conceived.

Figure 3.4.6 shows that throughout residues 10 to 45 (α1 - α2) the KstR RMSF was not affected by differences in ligand size. For residues 46 to 58 (α2 - α3 linker and α3) the FLB RMSF increases and is largest. The FLB RMSF then gradually decreases, converging with SLB RMSF where they are similar until residue 70. This demonstrates that while the flexibility of the DBD is largely naive to ligand size, α3 is less flexible when KstR is bound to a shorter ligand.

Residues 70 to 90 (α4 - α5 linker domain) displayed a difference in flexibility between FLB and SLB RMSF. The SLB RMSF peaked at 0.59 nm at residue 82 while FLB KstR peaked at 0.41 nm at residue 82. For residues lying within 115 to 140 (α6 - α7) both FLB and SLB KstR displayed a similar trend in RMSF but FLB KstR produced greater RMSF. This indicates that a shorter ligand causes the KstR α4 - α5 linker domain to become more flexible, and causes the α6 - α7 region of the LBD to be less flexible.

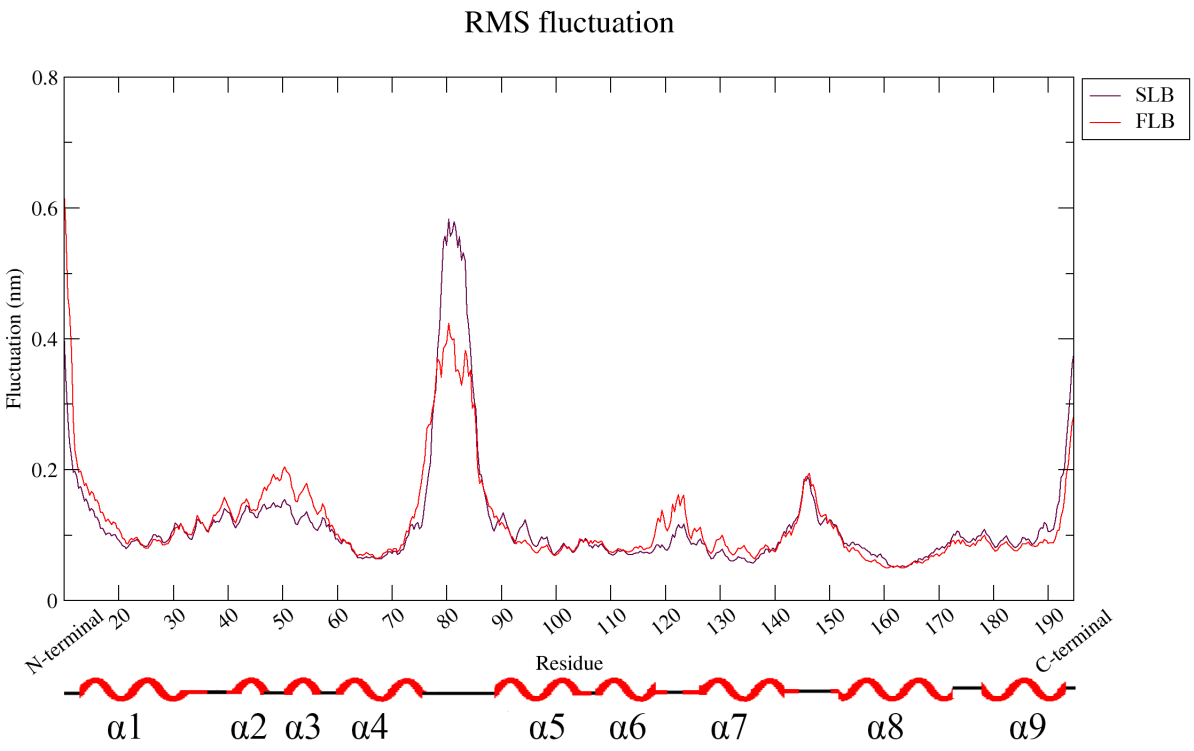

Figure 3.4.6: *Backbone RMSF of KstR chain for FLB (red) and SLB (maroon) KstR. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

Together this shows that the size of ligand has little impact on the dynamics of DBD or LBD aside from around the ligand binding cavity.

3.4.4 Ligand size affects how the DBD and LBD of KstR are arranged

The distances between residue pairs for SLB KstR were calculated as described in Section 2.2.7. By comparing distances between residues in FLB and SLB DBDs how the arrangement of the KstR DBD is altered by varied ligand size is shown. The locations of the DBD residue pairs measured are shown in Figure 2.2.2.

Table 3.4.5 shows that for all residue pairs FLB KstR produced a larger average distance than SLB KstR. For all residues pairs but Arg-45 and Tyr-54, FLB KstR produced a larger average distance standard deviation and range than SLB KstR. For all residue pairs FLB KstR sampled the largest distance while SLB KstR sampled the smallest (Figure 3.4.7).

Together this demonstrates that a shorter ligand induces smaller average distances between residue pairs at the DBD and also causes the distances between residue pairs at the DBD to move less.

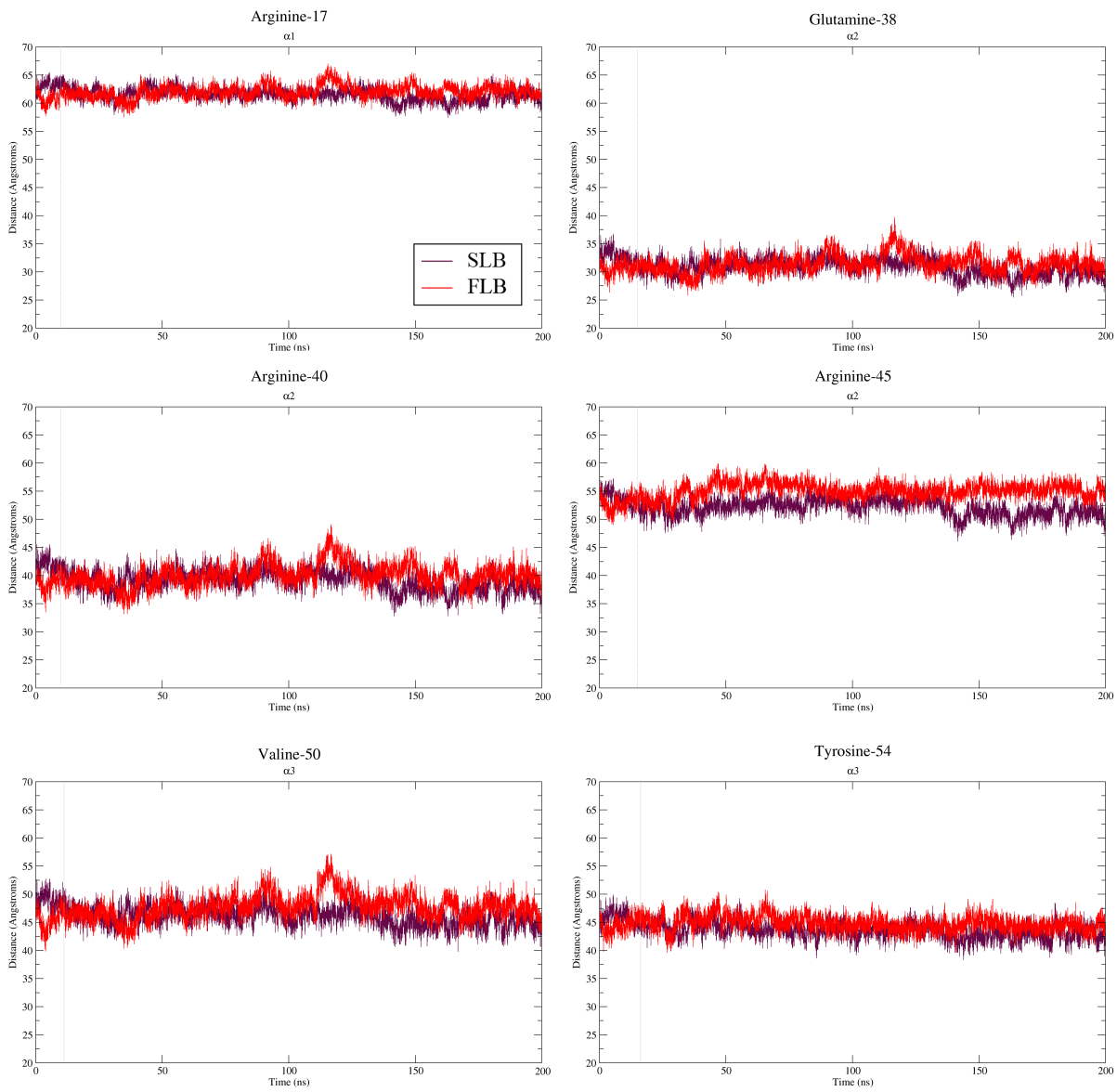

Figure 3.4.7: *Distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 of each DBD of FLB (red) and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.2 for the residue position reference.*

Table 3.4.5: *Average, standard deviation and standard error, and range of distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 for FLB and SLB KstR over time (10 - 200 ns).*

Statistics of Arg-17 distance				
	$\bar{x}$ (Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	62.08	1.16	0.01	57.52 - 67.02 (9.50)
SLB	61.48	1.08	0.01	57.43 - 65.21 (7.78)
Statistics of Gln-38 distance				
FLB	31.45	1.64	0.02	25.84 - 39.70 (13.86)
SLB	30.79	1.47	0.02	25.47 - 35.84 (10.37)
Statistics of Arg-40 distance				
FLB	40.09	2.05	0.02	33.19 - 49.07 (15.88)
SLB	39.04	1.78	0.02	32.74 - 44.94 (12.20)
Statistics of Arg-45 distance				
FLB	55.17	1.33	0.01	49.62 - 59.91 (10.29)
SLB	52.07	1.51	0.02	46.11 - 57.16 (11.05)
Statistics of Val-50 distance				
FLB	47.81	2.20	0.02	40.37 - 57.14 (16.77)
SLB	46.02	1.69	0.02	39.79 - 52.53 (12.74)
Statistics of Tyr-54 distance				
FLB	45.47	1.39	0.01	39.63 - 50.75 (11.12)
SLB	43.53	1.43	0.02	38.31 - 49.59 (11.28)

By comparing distances between residues in FLB and SLB LBDs how the arrangement of the KstR LBD is altered by ligand size is shown. The locations of the LBD residue pairs measured are shown in Figure 2.2.2.

Table 3.4.6 shows that for all residue pairs aside from Leu-135, SLB KstR produced the largest average distances. for all residues pairs but Met-138 and Asp-162, FLB KstR produced the highest average distance standard deviation between residue pairs. for all residue pairs FLB KstR sampled the smallest distance.

Together this indicates that a smaller ligand at the LBD induces larger distances between residue that change less dynamically at the LBD. Furthermore, that a smaller ligand allows for larger distances to be sampled at the LBD.

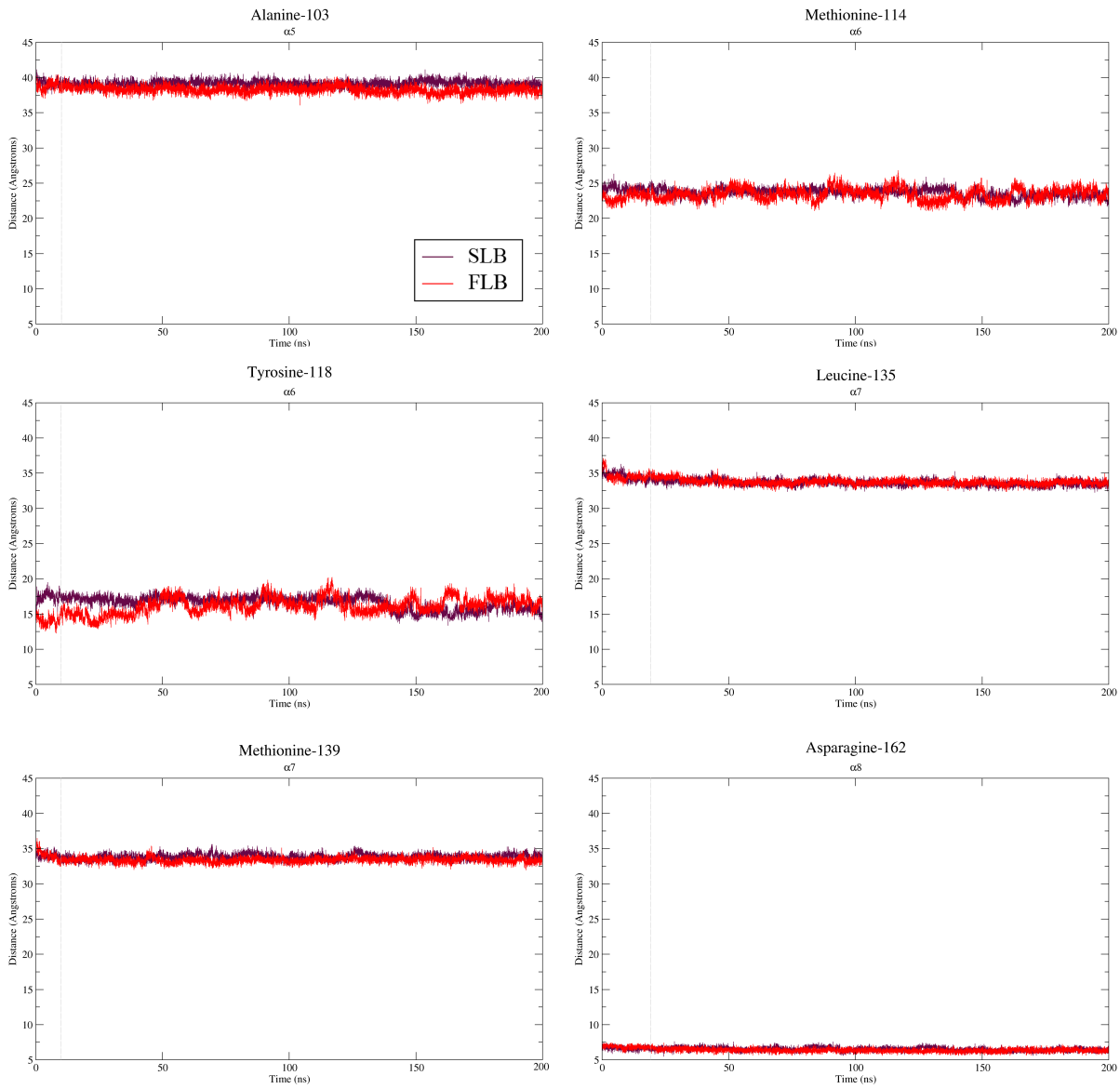

Figure 3.4.8: *Distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 of each LBD of FLB (red) and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.3 for the residue position reference.*

Table 3.4.6: *Average, standard deviation and standard error, and range of distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 for FLB and SLB KstR over time (10 - 200 ns).*

Statistics of Ala-103 distance				
	$\bar{x}$ (Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
FLB	38.285	0.519	0.005	36.099 - 40.189 (4.090)
SLB	39.130	0.462	0.005	37.346 - 41.126 (5.256)
Statistics of Met-114 distance				
FLB	23.448	0.846	0.009	20.981 - 26.772 (5.791)
SLB	23.632	0.591	0.006	21.175 - 25.868 (4.692)
Statistics of Tyr-118 distance				
FLB	16.22	1.11	0.01	12.88 - 20.19 (7.31)
SLB	16.59	0.91	0.01	13.38 - 18.79 (5.41)
Statistics of Leu-135 distance				
FLB	33.711	0.518	0.005	32.240 - 36.032 (3.791)
SLB	33.674	0.408	0.004	32.294 - 35.429 (3.135)
Statistics of Met-139 distance				
FLB	33.387	0.389	0.004	31.976 - 35.364 (3.388)
SLB	33.837	0.421	0.004	32.329 - 35.606 (3.277)
Statistics of Asp-162 distance				
FLB	6.367	0.231	0.002	5.457 - 7.397 (1.939)
SLB	6.518	0.282	0.003	5.509 - 7.603 (2.094)

Overall a shorter ligand results in smaller distances between residue pairs at the DBD, but larger distances for residue pairs at the LBD. Furthermore, at both domains the distances changed less when shorter ligand was bound to KstR. Although, in all cases the differences in average distance were significant, they were subtle.

3.4.5 Ligand size does not perturb the structure of KstR

The inter-residue contact maps for SLB KstR were produced according to Section 2.2.9. Differences in the inter-residue distance between SLB and FLB KstR indicate how the length of ligand affects the structure of KstR, as seen in Figure 3.4.9 compared to Figure 3.4.10.

The SLB KstR results in longer distances between residues 10 to 30 (α1) and 122 to 138 (α7), and between residues 62 to 74 (α4) and 94 to 106 (α6 - α7) but the furthest distances between residues 90 to 110 (α5 - α6 and α8) compared to FLB KstR. This demonstrates that when a smaller ligand is bound to KstR there are changes between helices at the DBD and LBD and within the LBD itself.

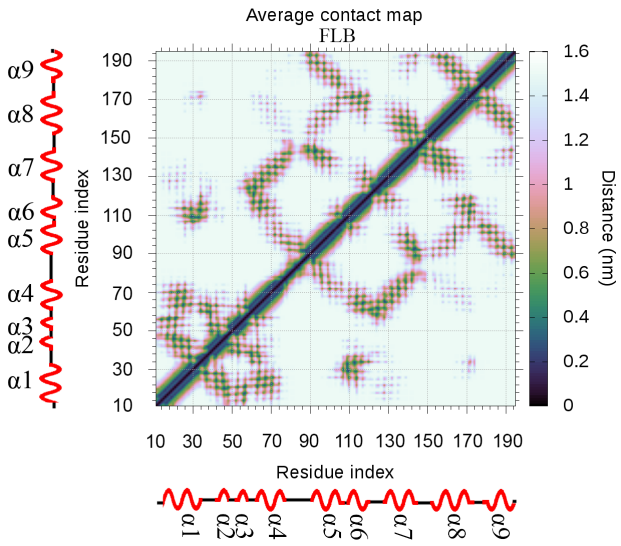

Figure 3.4.9: *Average inter-residue distance of intramolecular interactions of FLB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

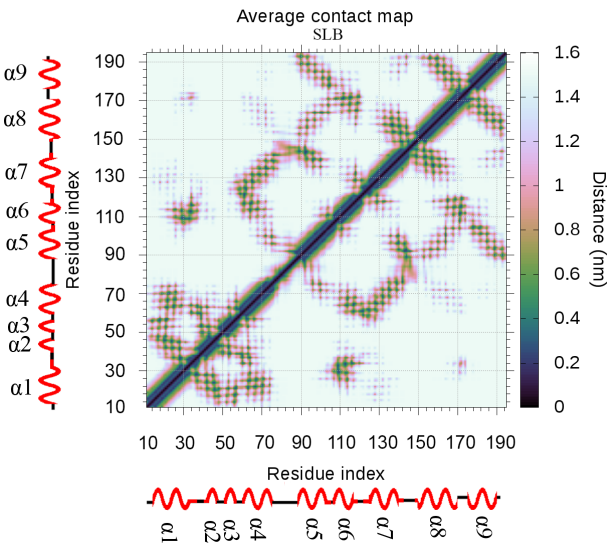

Figure 3.4.10: *Average inter-residue distance of intramolecular interactions of SLB KstR over time (0 - 200ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

This shows that when ligand is shortened there are minor changes to the organisation of residues between the DBD and LBD and near the ligand binding cavity.

3.4.6 Ligand size does not impact the secondary structure of KstR

The secondary structure time series for SLB KstR was calculated as described in Section 2.2.6. By comparing the DSSP assignments of FLB and SLB KstR how ligand size impacts the structure of KstR is revealed, as shown in Figure 3.4.11 compared to Figure 3.4.12.

The SLB KstR exhibits a smaller $\alpha 1$ than FLB KstR. FLB KstR $\alpha 1$ is composed of 27 residues while SLB KstR is composed of 26. The SLB KstR $\alpha 6$ is comprised of 11 residues while FLB KstR is composed of 12 residues. This demonstrates that smaller ligand size causes α-helices at the DBD to become smaller.

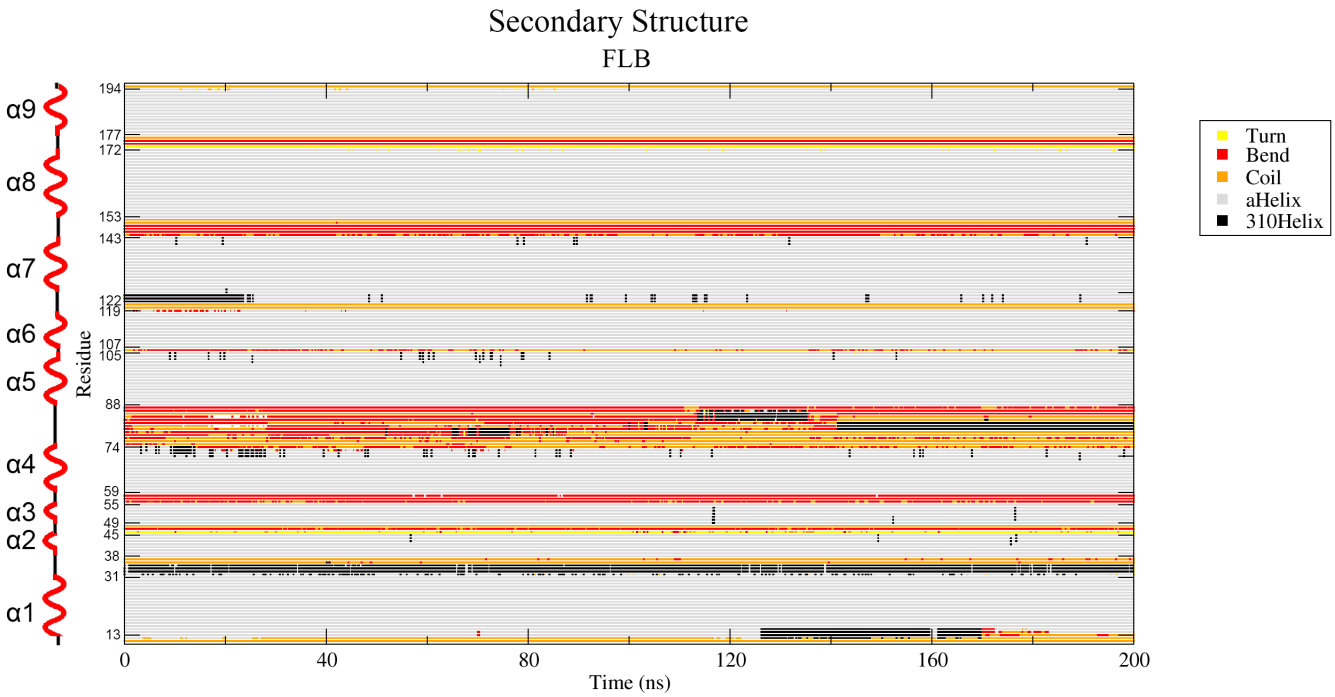

Figure 3.4.11: *Secondary structure assignment of FLB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

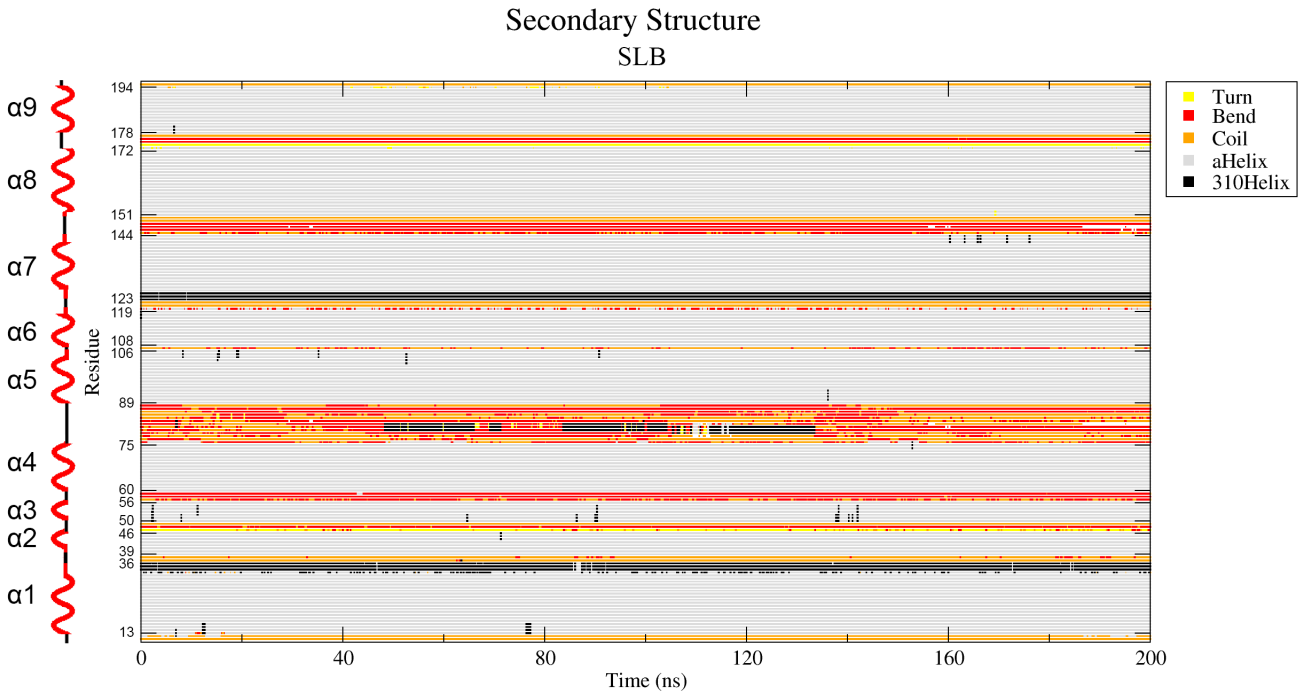

Figure 3.4.12: *Secondary structure assignment of SLB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

The SLB KstR retains slightly more helical assignments than FLB KstR overall. Table 3.4.7 shows that SLB KstR structure is 76.5% helices while for FLB KstR it is 75.6%. This indicates that a smaller ligand induces greater helical assignment than full ligand KstR.

Table 3.4.7: *Average secondary structure assignment for FLB and SLB KstR over time (10 - 200 ns).*

Secondary structure distribution of FLB and SLB KstR					
	α-helix	Coil	Turn	Bend	3_{10}-helix
FLB	74.3%	11.9%	5.6%	6.3%	1.3%
SLB	74.7%	12.8%	3.9%	6.1%	1.9%

3.4.7 Ligand size affects Trp-164 flipping

The Trp-164 dihedral angles for SLB KstR was calculated as detailed in Section 2.2.8. Trp-164 is located within the ligand binding cavity and has close contacts with the ligand. By comparing the FLB and SLB KstR Trp-164 dihedral angle how ligand size impacts Trp-164 conformation is revealed, as shown in Figure 3.4.13.

The FLB KstR presented a significantly more severe Trp-164 dihedral angle on average compared SLB KstR. The SLB KstR demonstrates a larger average

dihedral angle standard deviation of than FLB KstR (Table 3.4.8). Together this indicates that a shorter ligand causes Trp-164 to take on a different organisation which moves about more dynamically than FLB KstR Trp-164.

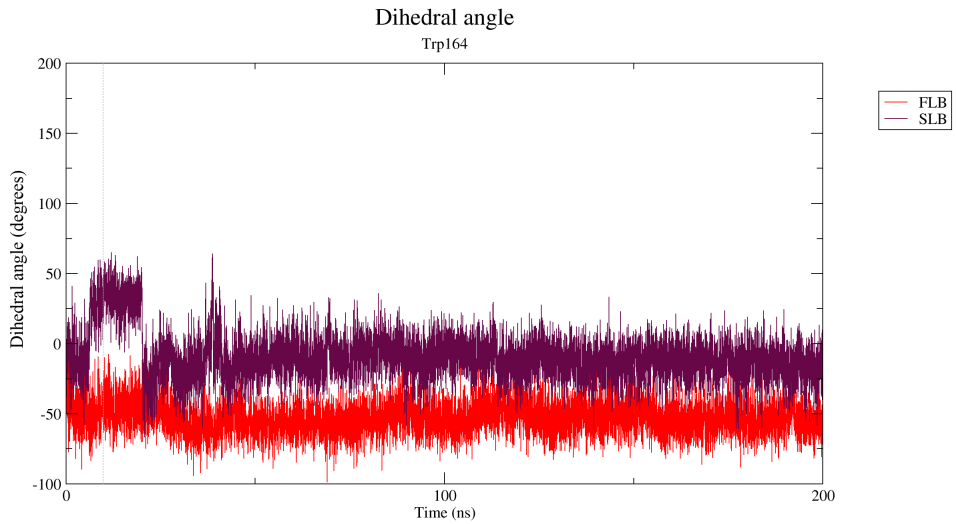

Figure 3.4.13: *Dihedral angle across Trp-164 for FLB (red) and SLB (maroon) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.4 for the residue dihedral reference reference.*

Table 3.4.8: *Average, standard deviation and standard error, and range of FLB, and SLB KstR Trp-164 dihedral angle over time (10 - 200 ns).*

Statistics of Trp-164 dihedral angle				
	$\bar{x}$ (°)	σ (°)	$\sigma_{\bar{x}}$ (°)	Range (max - min)
FLB	-52.9	11.6	0.1	-98.7 - -7.1 (91.6)
SLB	-9.4	13.9	0.1	-66.6 - 64.9 (131.5)

3.4.8 Ligand size affects the enthalpic relationship between KstR and ligand

The enthalpy calculations between SLB KstR and ligand are described in Section 2.2.3. By measuring the enthalpy energy between ligand and KstR how ligand size affects the binding relationship between ligand and KstR is exhibited.

Throughout KstR a smaller ligand resulted in smaller enthalpy between ligand and KstR (Table 3.4.9). The shorter ligand resulted in a 74% decrease in enthalpy to the DBD and a 53% decrease in enthalpy with the LBD. This indicates that the shorter ligand has a smaller enthalpic contribution to KstR and when shortened the enthalpic energy between the DBD is diminished disproportionately.

Table 3.4.9: *Average and error estimate of the contribution of the KstR-ligand enthalpy to the overall enthalpy between KstR and ligand for FLB or SLB KstR respectively over time (0 - 200 ns).*

Enthalpy between KstR and ligand or shortened ligand		
	Average (kJ/mol)	Error estimate
FLB	-1390	24
SLB	-565	18
Enthalpy between KstR DBD and ligand or shortened ligand		
FLB	-400	50
SLB	-106	22
Enthalpy between KstR LBD and ligand or shortened ligand		
FLB	-990	30
SLB	-459	28

3.5 Differences in DNA size impact KstR conformational dynamics

To examine if the size of DNA impacts the conformational dynamics of KstR a 18 bp DNA bound to KstR was prepared. The short long DNA bound (SLDB) is comprised of a 18 bp DNA bound to KstR. The SLDB was produced by shortening the 26 bp DNA from LoDB to 18 bp, and represents a shortened DNA-bound KstR.

The SLDB KstR was then solvated as described in Section 2.2.3, and the full atomic composition is detailed in Table 2.2.1. After solvation, the system was equilibrated as described in Section 2.1.4. Finally, the system was simulated for a total of 200 ns as described in Section 2.1.5. Once the simulation had been completed various analysis were conducted as described in Section 2.2.

By comparing the LoDB, and SLDB KstR simulations the aim was to determine how conformational dynamics of KstR is altered by DNA size, partially resolving the third objective (Section 1.4).

3.5.1 DNA size does not impact the overall shape of KstR

The Rg of SLDB KstR was calculated as described in Section 2.2.1. By comparing Rg of LoDB and SLDB KstR how DNA size impacts the shape and size of KstR is revealed, as shown in Figure 3.5.1.

The SLDB KstR has a larger average Rg than LoDB KstR at 2.189 nm, the SLDB KstR Rg range was also larger than that of LoDB KstR (Table 3.5.1). SLDB KstR and LoDB KstR have the same average Rg standard deviation. Although the Rg differences between LoDB and SLDB KstR are small the standard error indicates it is significant.

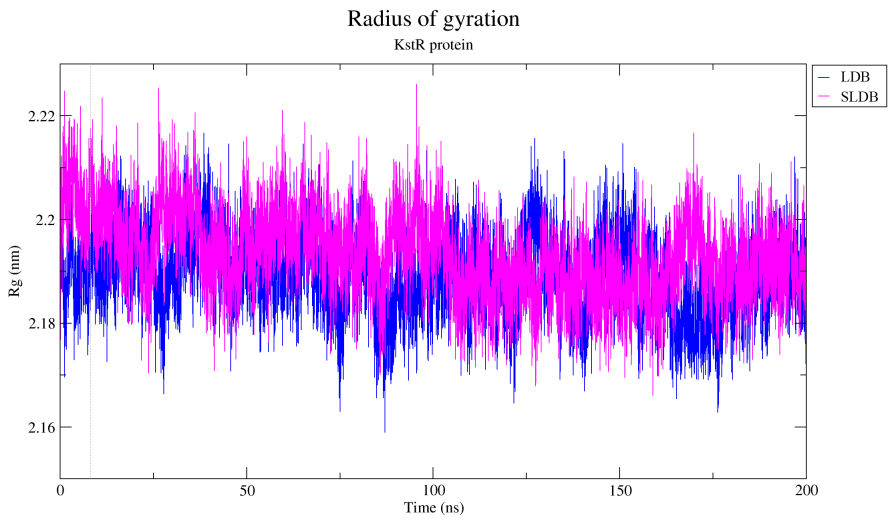

Figure 3.5.1: *KstR Rg of LoDB (blue) and SLDB (magenta) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.5.1: *Average, standard deviation, standard error, and range of LoDB and SLDB KstR Rg over time (10 - 200 ns).*

Statistics of KstR Rg				
	$\bar{x}(nm)$	$\sigma(nm)$	$\sigma_{\bar{x}}$ (nm)	Range (max - min)
LoDB	2.189	0.008	0.000	2.158 - 2.216 (0.058)
SLDB	2.192	0.008	0.000	2.157 - 2.226 (0.069)

The SASA of SLDB KstR was calculated as explained in Section 2.2.2. By comparing SASA of LoDB and SLDB KstR how DNA size impacts the shape and size of KstR is revealed.

The SLDB KstR has a larger average SASA than LoDB KstR, but a smaller average SASA standard deviation and range than LoDB KstR (Table 3.5.2). While the standard error indicates the differences are small they are still significant, demonstrating that DNA size has a small affect on KstR shape (Figure 3.5.2).

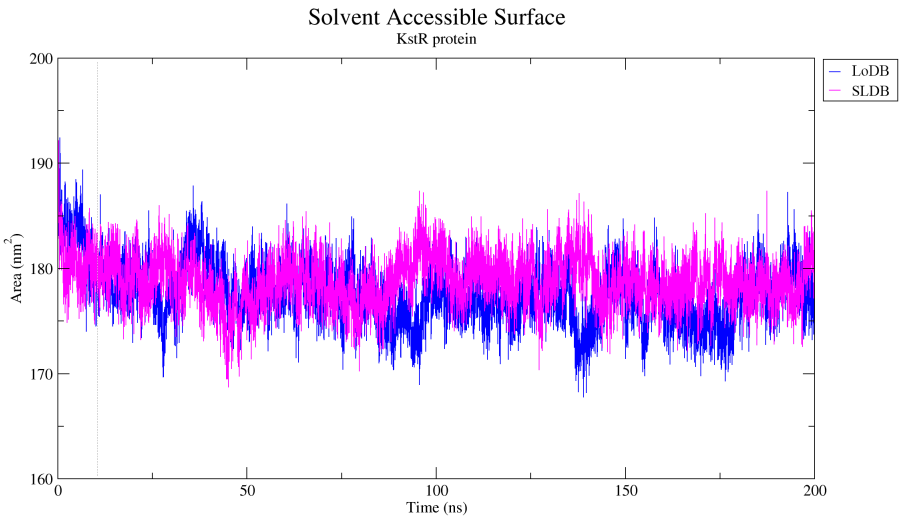

Figure 3.5.2: *KstR SASA of LoDB (blue) and SLDB (magenta) KstR systems over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.5.2: *Average, standard deviation and standard error, and range of LoDB and SLDB KstR SASA over time (10 - 200 ns).*

Statistics of KstR SASA				
	$\bar{x}(nm^2)$	$\sigma(nm^2)$	$\sigma_{\bar{x}}$ (nm^2)	Range (max - min)
LoDB	177.39	2.58	0.03	167.76 - 187.85 (20.09)
SLDB	178.73	2.29	0.02	168.72 - 187.37 (18.65)

The volume of SLDB KstR was calculated as explained in Section 2.2.2. By comparing volume of LoDB and SLDB KstR how DNA size impacts the size of KstR is revealed.

The average volume of LoDB KstR is smaller than that of SLDB, while the difference is small the standard error indicates it is significant (Table 3.5.3). This indicates that the length of DNA has a small impact on the volume of KstR.

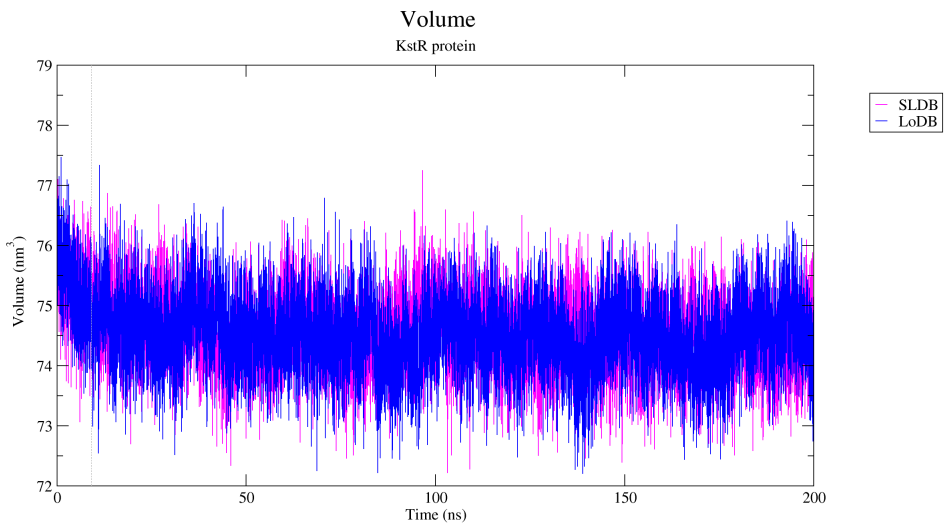

Figure 3.5.3: *Volume of LoDB (blue) and SLDB (magenta) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line.*

Table 3.5.3: *Average, standard deviation and standard error, and range of LoDB and SLDB KstR volume over time (10 - 200 ns).*

Statistics of KstR volume				
	$\bar{x}(nm^3)$	$\sigma(nm^3)$	$\sigma_{\bar{x}}$ (nm^3)	Range (max - min)
LoDB	74.490	0.629	0.007	72.209 - 77.334 (5.125)
SLDB	74.527	0.621	0.006	72.222 - 77.244 (5.022)

Together this demonstrates that when short DNA is bound to KstR there are scarce differences in the shape and size of KstR.

3.5.2 DNA size does not impact the conformational ensemble of KstR

The conformational ensembles of LoDB and SLDB KstR were calculated as detailed in Section 2.2.4. By comparing the conformational ensembles of KstR between SLDB and LoDB KstR how DNA length affects conformational ensemble of KstR is shown, as seen on Figure 3.5.4.

The LoDB KstR samples 11 conformational clusters while SLDB KstR samples 8. LoDB KstR predominantly resides within cluster 1 (beige) while SLDB predominantly samples cluster 2 (light-purple) and samples cluster 3 (beige) for a substantial duration. Furthermore, both systems sample cluster 8 (blue) albeit briefly. This indicates that DNA size has little impact on the conformational dynamics of KstR although did induce a slightly smaller breadth of conformations sampled and less restrained conformational behaviour.

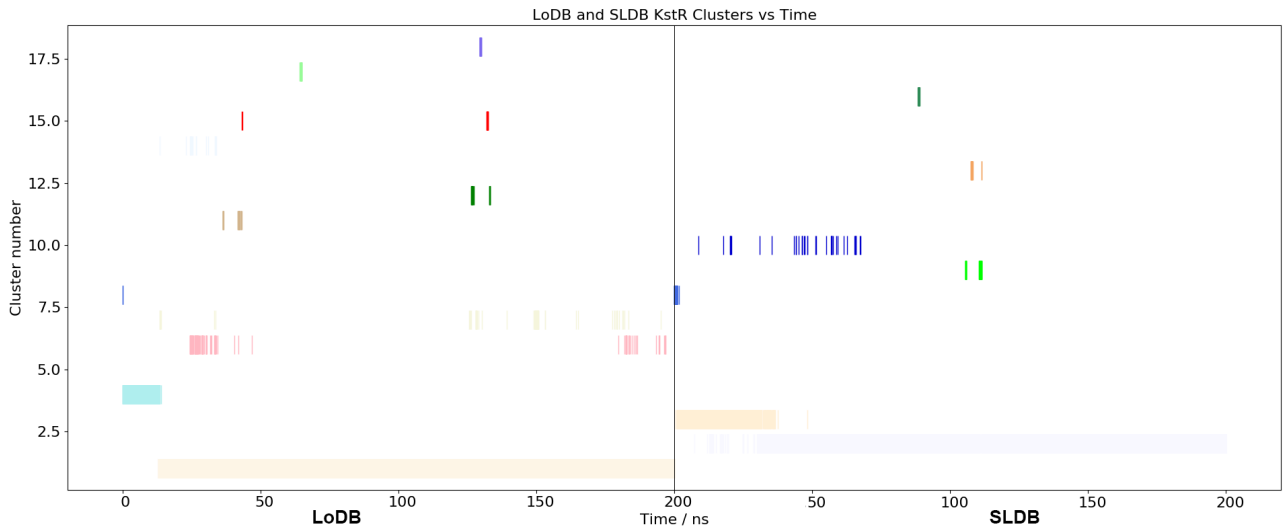

Figure 3.5.4: *18 conformation cluster groups of LoDB (left), and SLDB (right) KstR over time (0 - 200 ns). Presentation limited to clusters with more than 10 conformations in their group with a RMSD cut-off of 0.2 nm.*

3.5.3 DNA size impacts the flexibility of KstR

The RMSF for SLDB KstR was calculated as detailed in Section 2.2.5. By comparing the RMSF of LoDB and SLDB KstR whether DNA length impacts the flexibility, to what extent, and where is shown. This aids findings of Section 3.5.1 and 3.5.2 by detailing which regions of KstR may be responsible for the DNA size affects seen, as shown in Figure 3.5.5.

The LoDB and SLDB KstR feature the same RMSF for the majority of the DBD except between residues 30 to 40 (α1 C-terminus). While the RMSF trend through this region is the same for LoDB and SLDB KstR, LoDB shows a larger fluctuation. This shows that when DNA is shortened α1 fluctuates less, although flexibility of the KstR DBD is otherwise unchanged.

There is variability in RMSF between LoDB and SLDB KstR throughout the LBD. Between residue 75 to 85 (α4 - α5 linker domain) LoDB KstR displays a much larger RMSF peaking at approximately 0.35 nm while SLDB KstR has a RMSF peak of 0.15 nm. Throughout residues 90 to 160 of the SLDB KstR LBD displayed more flexibility than LoDB KstR. Residues 90 to 110 (α5 - α6), 120 to 160 (α7 - α8 N-terminus) displayed a larger RMSF for SLDB KstR. Residues 165 to 175 (α8) show a slightly larger RMSF for LoDB than SLDB KstR. This indicates that a shorter DNA restricts the flexibility of KstR α4 - α5 and α8 C-terminus while enhanced the flexibility of α5 - α6 and α7 - α8 N-terminus.

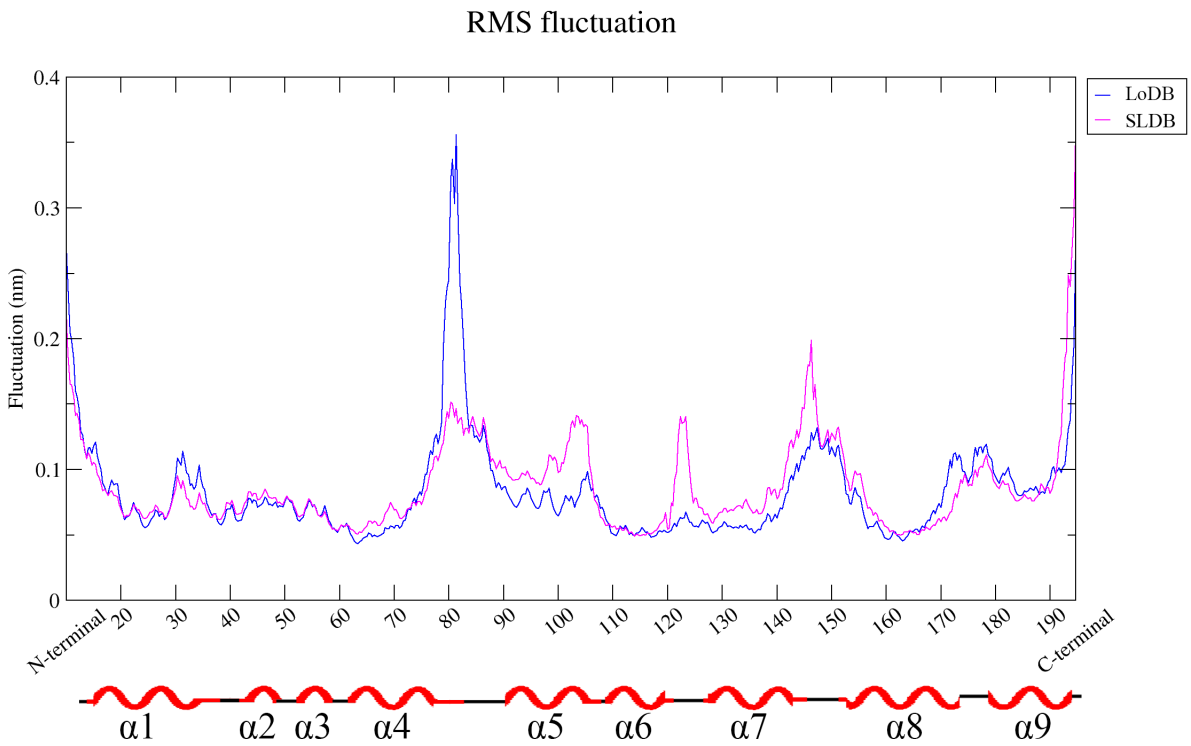

Figure 3.5.5: *Backbone RMSF of entire KstR chain for LoDB (blue) and SLDB (magenta) KstR. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

Shortened DNA has a small impact on the flexibility of the KstR DBD, while across the KstR LBD a smaller DNA results in a more flexible LBD, for some regions aside from the $\alpha 4$ - $\alpha 5$ linker domain.

3.5.4 DNA size does not affect how the DBD and LBD of KstR are arranged

The distances between residue pairs for SLDB KstR were calculated as described in Section 2.2.7. By comparing the distance between DBD residues in LoDB and SLDB KstR how DNA size affects the arrangement of the KstR DBD is shown, as shown in Figure 3.5.6. The locations of the pairs of residues is shown in Figure 2.2.2.

For all distances between residue pairs at the DBD LoDB KstR had a smaller average distances, average distance standard deviation, and range and sampled the smallest distances while SLDB KstR sampled the largest (Table 3.5.4). Although, the distance values sampled by LoDB KstR and SLDB KstR frequently overlap and display similar trends over the time series (Figure 3.5.6). Together this shows that a shorter DNA induces subtly smaller and less restrained distances between residue pairs at the DBD.

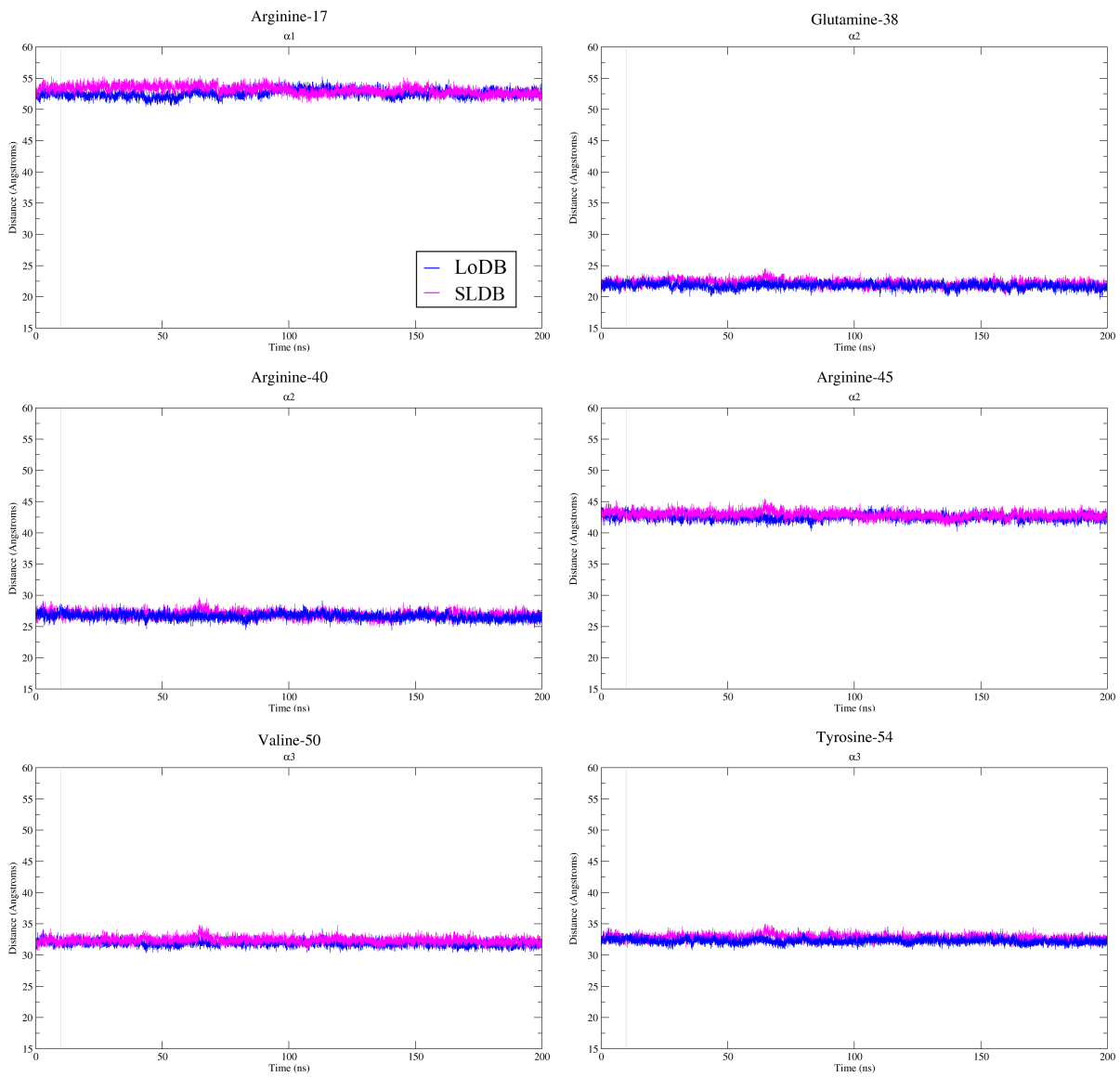

Figure 3.5.6: *Distance between Arg-17, Gln-38, Arg-40, Arg-45, Va-50, and Tyr-54 of each chain of LoDB (blue) and SLDB (magenta) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.2 for the residue position reference.*

Table 3.5.4: *Average, standard deviation and standard error, and range of distance between Arg-17, Gln-38, Arg-40, Arg-45, Val-50, and Tyr-54 for LoDB and SLDB KstR over time (10 - 200 ns).*

Statistics of Arg-17 distance				
	$\bar{x}$(Å)	σ(Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
LoDB	52.575	0.613	0.006	50.470 - 55.187 (4.717)
SLDB	53.145	0.673	0.007	50.655 - 55.373 (4.718)
Statistics of Gln-38 distance				
LoDB	21.776	0.473	0.005	19.614 - 23.390 (3.775)
SLDB	22.240	0.514	0.005	20.178 - 24.543 (4.365)
Statistics of Arg-40 distance				
LoDB	26.63	0.53	0.05	24.45 - 29.13 (4.68)
SLDB	26.91	0.55	0.01	24.91 - 29.66 (4.75)
Statistics of Arg-45 distance				
LoDB	42.489	0.508	0.005	40.246 - 44.531 (4.285)
SLDB	42.908	0.554	0.006	40.951 - 45.460 (4.508)
Statistics of Val-50 distance				
LoDB	31.885	0.455	0.005	30.389 - 34.192 (3.803)
SLDB	32.382	0.498	0.005	30.591 - 34.778 (4.186)
Statistics of Ty-54 distance				
LoDB	32.245	0.415	0.004	30.791 - 34.525 (3.733)
SLDB	32.828	0.462	0.005	31.190 - 34.998 (3.808)

By comparing the distance between LBD residues in LoDB and SLDB KstR how DNA size affects the arrangement of the KstR LBD is shown. The locations of the pairs of residues is shown in Figure 2.2.3.

For distances between residues at the LBD LoDB KstR exhibited a smaller average distance (aside for Ala-103), standard deviation, and range (Table 3.5.5) than SLDB KstR. LoDB KstR and SLDB KstR both heavily overlap in values sampled, and the trends between the systems are almost indiscernible (Figure 3.5.7). Together this shows that a shorter DNA induces slightly larger distances that fluctuate more between residues of the LBD.

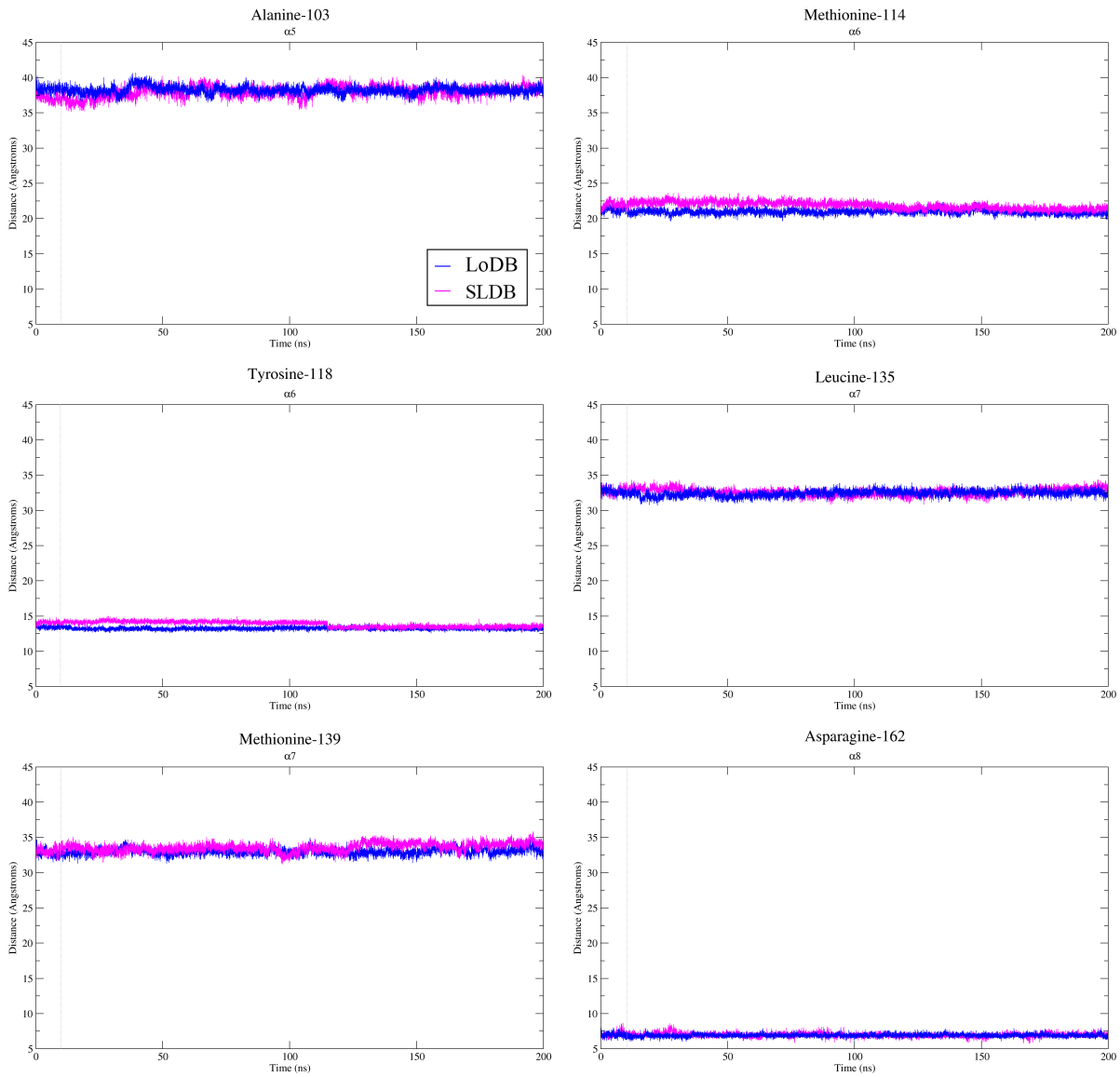

Figure 3.5.7: *Distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 of each chain of LoDB (blue) and SLDB (magenta) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See Figure 2.2.3 for the residue position reference.*

Table 3.5.5: *Average, standard deviation and standard error, and range of distance between Ala-103, Met-114, Tyr-118, Leu-135, Met-139, and Asp-162 for LoDB and SLDB KstR over time (10 - 200 ns).*

Statistics of Ala-103 distance				
	$\bar{x}$(Å)	σ (Å)	$\sigma_{\bar{x}}$ (Å)	Range (max - min)
LoDB	38.234	0.553	0.006	36.369 - 40.655 (4.285)
SLDB	37.997	0.713	0.007	35.172 - 40.307 (5.134)
Statistics of Met-114 distance				
LoDB	20.939	0.351	0.004	19.633 - 22.550 (2.917)
SLDB	21.857	0.519	0.005	20.261 - 23.624 (3.363)
Statistics of Tyr-118 distance				
LoDB	13.224	0.188	0.002	12.597 - 14.161 (1.563)
SLDB	13.834	0.394	0.004	12.714 - 14.961 (2.247)
Statistics of Leu-135 distance				
LoDB	32.403	0.430	0.004	30.725 - 33.934 (3.208)
SLDB	32.588	0.492	0.005	30.838 - 34.353 (3.514)
Statistics of Met-139 distance				
LoDB	32.998	0.471	0.005	31.323 - 34.936 (3.612)
SLDB	33.603	0.614	0.006	31.207 - 35.814 (4.607)
Statistics of Asp-162 distance				
LoDB	6.902	0.217	0.002	6.014 - 7.963 (1.948)
SLDB	6.961	0.278	0.003	5.760 - 8.577 (2.817)

DNA size has a small impact on the distance between residues at the DBD. Throughout KstR a shorter DNA resulted in larger distances, less consistent and larger ranges of distances between residues. However, in for both regions the LoDB and SLB range of values overlap immensely.

3.5.5 DNA size does not perturb the structure of KstR

The inter-residue contact maps of SLDB KstR were generated according to 2.2.9. By comparing the LoDB and SLDB contact maps whether DNA length affects the structure of KstR is shown.

Figure 3.5.8 and 3.5.9 show there are distances between residues 14 to 22 (α1) and 74 to 82 (α4 - α5 linker domain) that are longer when KstR is bound to shorter DNA. This indicates that shorter DNA affects contacts between α1 and α4 - α5 linker domain.

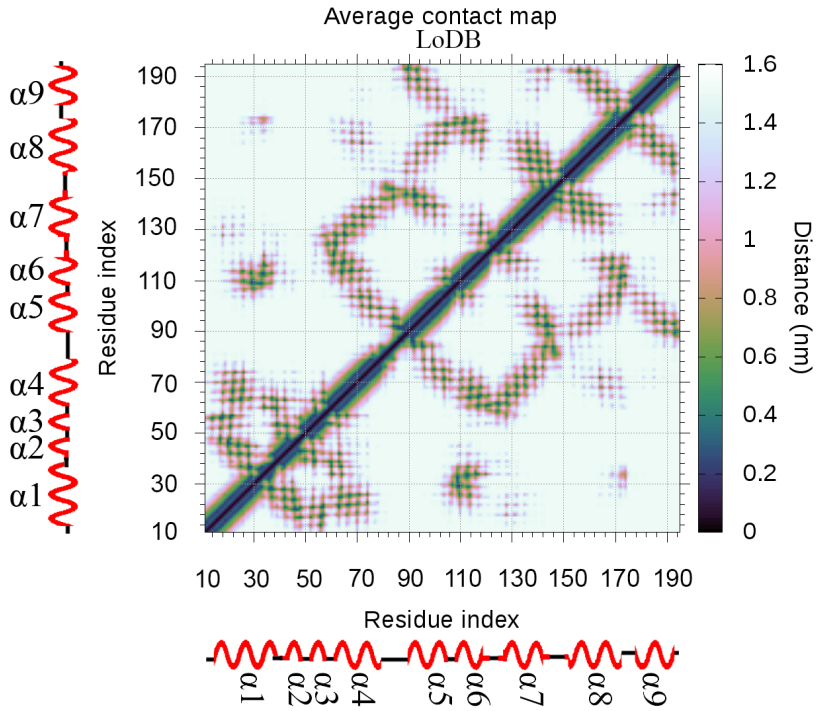

Figure 3.5.8: *Average inter-residue distance of intramolecular interactions of LoDB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

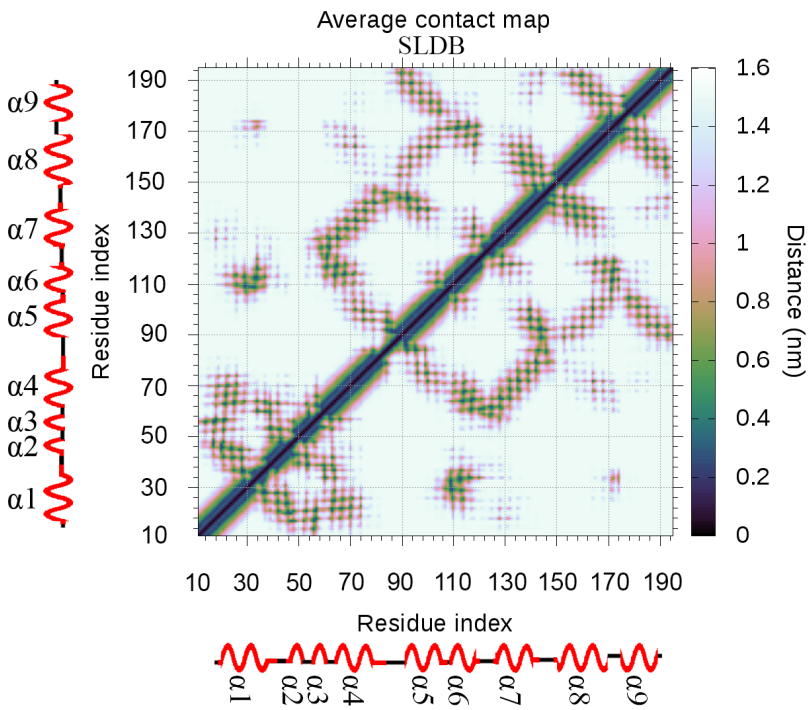

Figure 3.5.9: *Average inter-residue distance of intramolecular interactions of SLDB KstR over time (0 - 200 ns). Cartoons at the left and bottom of the figure illustrates helices (red) and loops (black).*

3.5.6 DNA size does not impact the KstR secondary structure

The secondary structure time series of SLDB KstR were calculated as explained in Section 2.2.6. By comparing the LoDB and SLDB KstR DSSP assignment to residues through KstR displays whether DNA size affects the structure of KstR is shown

Figure 3.5.10 and Figure 3.5.11 indicate that the $\alpha 4$ - $\alpha 5$, $\alpha 5$ - $\alpha 6$, $\alpha 7$ - $\alpha 8$, and $\alpha 8$ - $\alpha 9$ linkers are all shorter for SLDB KstR than LoDB KstR. This indicates that by shortening the DNA helical regions at the KstR LBD were extended while linkers shortened.

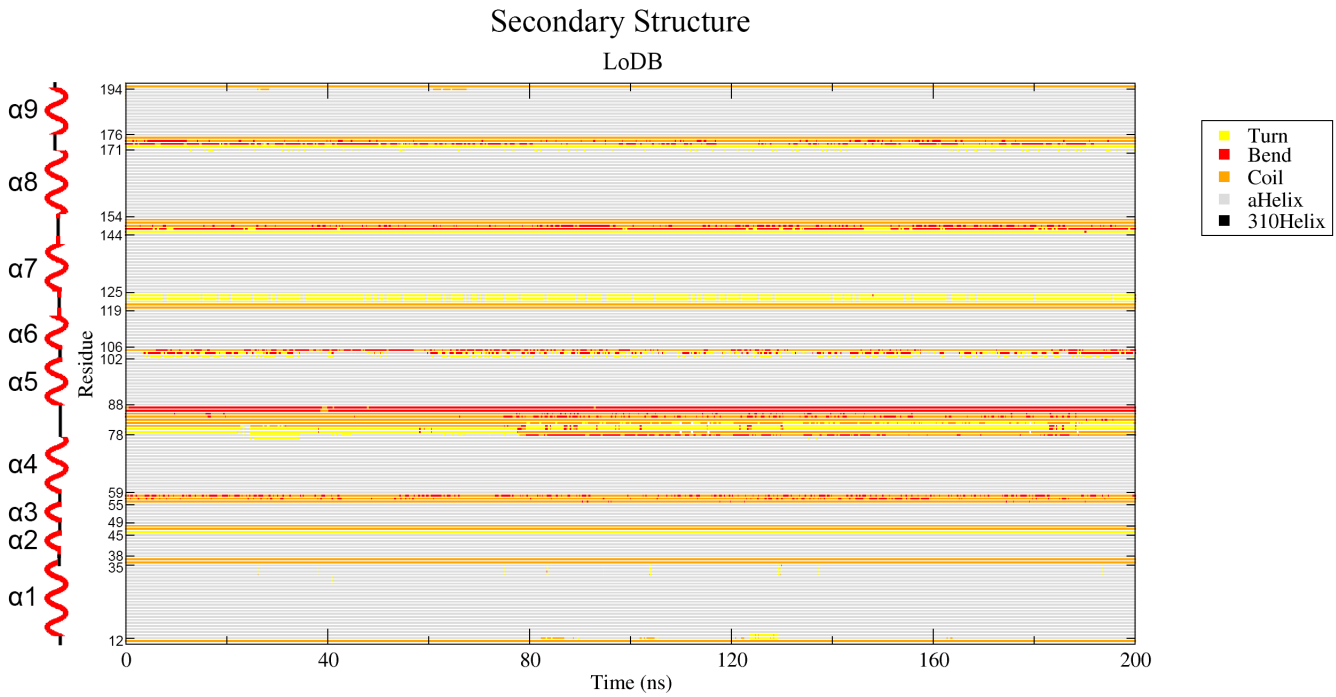

Figure 3.5.10: *Secondary structure assignment of LoDB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

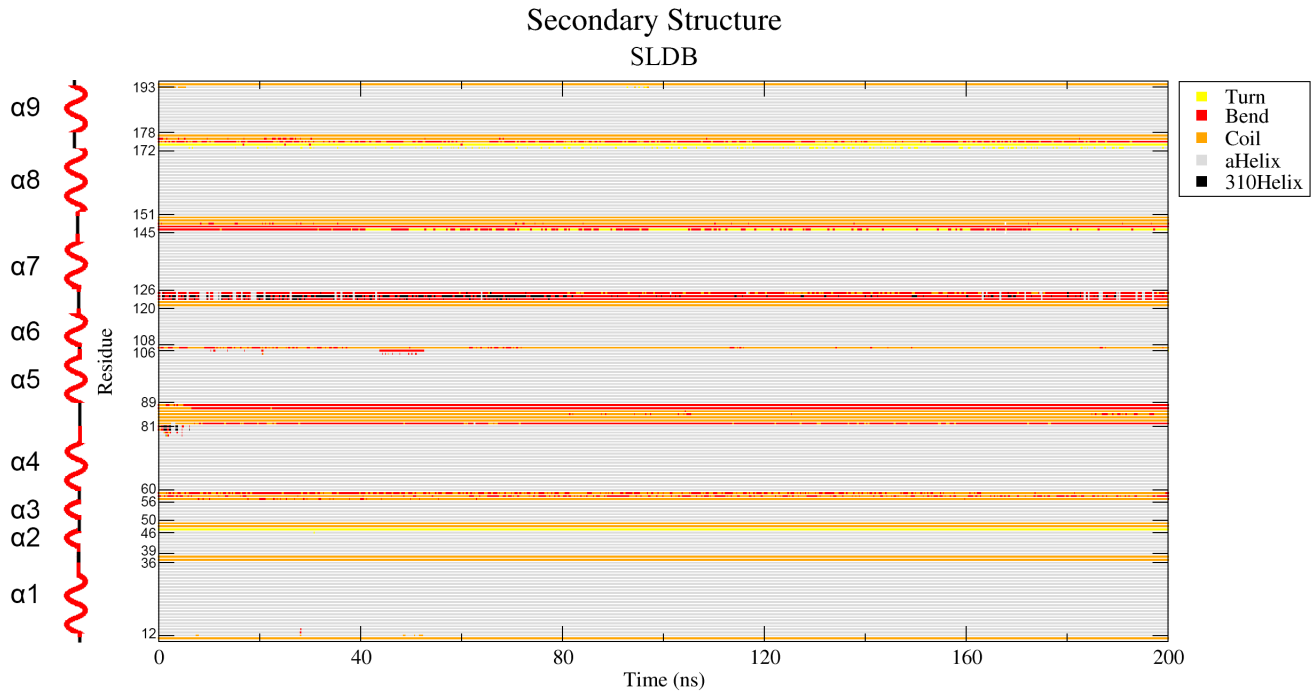

Figure 3.5.11: *Secondary structure assignment of SLDB KstR residues over time (0 - 200 ns). Cartoon at the left of the figure illustrates helices (red) and loops (black).*

The SLDB KstR is composed of a greater proportion of helical regions (80.8%) then LoDB (78%) KstR. This indicates that when DNA was shortened the architecture of KstR favoured more helical secondary structure assignments.

Table 3.5.6: *Table providing average secondary structure assignment for LoDB and SLDB KstR over time (10 - 200 ns).*

Secondary structure distribution of LoDB and SLDB KstR					
	α-helix	Coil	Turn	Bend	3_{10}-helix
LoDB	76.6%	8.7%	5.3%	7.5%	1.4%
SLDB	79.9%	8.7%	4.0%	6.1%	0.9%

3.5.7 DNA size affects Tyr-54 flipping

The Tyr-54 dihedral angle for SLDB KstR were calculated according to Section 2.2.8. Tyr-54 is located at the DBD and has close contacts with DNA. By comparing the Tyr-54 dihedral angle of LoDB to that of SLDB KstR how DNA length affects the orientation of Tyr-54 is shown, as seen in Figure 3.5.12.

The average Tyr-54 dihedral angle was larger for LoDB KstR than SLDB KstR (Table 3.5.6). The SLDB KstR had a larger Tyr-54 dihedral angle average standard deviation compared to LoDB KstR. The standard error indicates the differences are significant. Together this demonstrates that when DNA

was shortened the Tyr-54 dihedral angle was smaller and changed more dynamically, although the difference is slight.

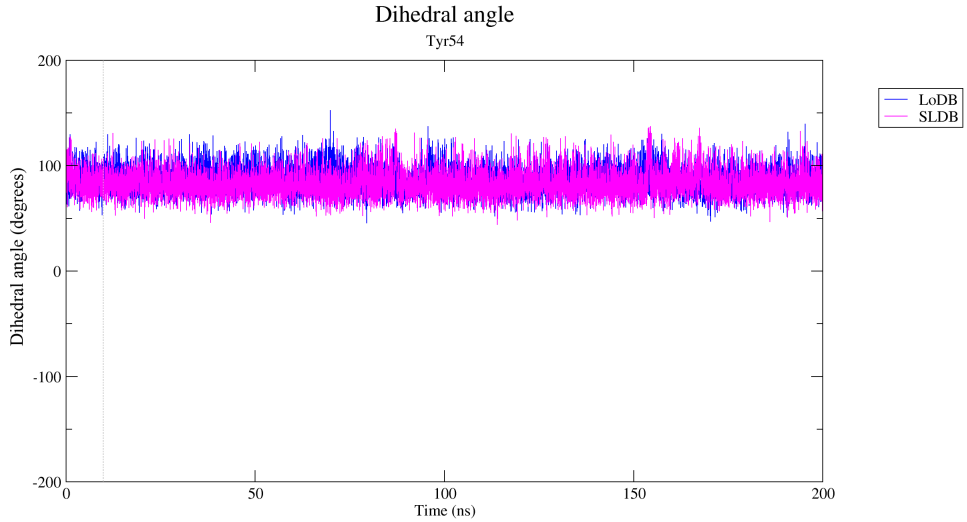

Figure 3.5.12: *Dihedral angle across Tyr-54 for LoDB (blue) and SLDB (magenta) KstR over time (0 - 200 ns). The first 10 ns are taken as an equilibration period, marked by the dashed line. See figure 2.2.4 for the residue dihedral reference.*

Table 3.5.7: *Average, standard deviation and standard error, and range of LoDB and SLDB KstR Tyr-54 dihedral angle over time (10 - 200 ns).*

Statistics of Tyr-54 dihedral angle				
	$\bar{x}$ (°)	σ (°)	$\sigma_{\bar{x}}$ (°)	Range (max - min)
LoDB	86.4	11.9	0.1	45.6 - 152.4 (106.8)
SLDB	82.9	12.2	0.1	44.0 - 137.2 (93.2)

3.5.8 DNA size affects the enthalpic relationship between KstR and DNA

The enthalpy calculations between SLDB KstR and DNA are detailed in Section 2.2.3. By comparing the enthalpy between KstR and DNA for LoDB and SLDB KstR how DNA size affects the DNA interaction with KstR is shown.

The DNA:SLDB KstR interaction enthalpy is more favourable than that of between DNA:LoDB KstR (Table 3.5.7). The DBD:SLDB KstR and the DBD:LoDB KstR are the same as that for DNA and the whole protein. The DNA:LBD SLDB KstR. Together this indicates that shorter DNA results in greater enthalpic potential between KstR and DNA, although the binding between the LBD DNA is less unfavourable for shortened DNA compared to long DNA.

Table 3.5.8: *Average and error estimate of contribution of the KstR-DNA enthalpy to the overall enthalpy between KstR and DNA for SLDB or LoDB KstR respectively over time (0 - 200 ns).*

Enthalpy between KstR and long DNA or shortened DNA		
	Average (kJ/mol)	Error estimate
LoDB	-1980	31
SLDB	-2000	16
Enthalpy between KstR DBD and long DNA or shortened DNA		
LoDB	-1990	3
SLDB	-2000	15
Enthalpy between KstR LBD and long DNA or shortened DNA		
LoDB	7	1
SLDB	1	0.5

3.6 Residues responsible for transitioning between KstR states

To identify residues important for KstR conformational transitions the enthalpy between KstR and residues across KstR was calculated. This highlights which KstR residues are most important for the binding between ligand or DNA and KstR.

Additionally, mutant forms of LoDB KstR were prepared to compare key residues (Section 3.1). Ala-17, Arg-40, and Tyr-54 are known to be important in the interaction between KstR and DNA and were mutated in Pymol (*Ho et al. manuscript in preparation*). Seven systems were prepared each had a different combination of mutations as detailed in Table 3.6.2.

The seven KstR mutants were then solvated as described in Section 2.1.3, the full atomic composition of each system is detailed in Table 2.2.1. After solvation, all three systems were equilibrated as described in Section 2.1.4. Finally, the systems were simulated for a total of 100 ns each, as described in Section 2.1.5. Once the simulations had been completed the energy between different parts of KstR and DNA were calculated, detailed in Section 2.2.3.

By comparing the enthalpic contributions between KstR residues and DNA or ligand which regions are critical for conformational rearrangement are highlighted. Furthermore, by comparing the LoDB and mutant KstR simulations enthalpy between KstR and DNA whether these residues are important in KstR dynamics is revealed. Together, this resolves the fourth objective (Section 1.4).

3.6.1 Enthalpic contribution between ligand and KstR residues and DNA and KstR residues

The enthalpy between residues and DNA or ligand demonstrates to what extent DNA and ligand affect different regions of KstR. By measuring the enthalpy between each residue and either DNA or ligand which regions are affected by DNA or ligand and to what extent is highlighted, as described in Section 2.2.3.

Residues 11 to 21 (α1) of KstR have favourable enthalpic interactions with DNA, but not with ligand. Instead, residues 22 to 31 (α1) have favourable enthalpic interactions with ligand, whereas these residues scarcely interact with DNA. Residue 33 to 64 (α2 - α4) of KstR exhibit a favourable enthalpic interaction with DNA while ligand only produces a small favourable enthalpy for some of those residues. Together this shows that the KstR DBD has a more favourable enthalpic interactions with DNA than with ligand.

Residue 64 to 85 (α4 and α4 - α5 linker domain), 95 to 119 (α5 - α6, 128 to 140 (α7), and 157 to 186 (α8 - α9) all demonstrated favourable enthalpic potentials between the ligand, while there were few enthalpic interactions between DNA and KstR throughout the LBD. This indicates that the LBD has more favourable interactions with ligand than with DNA.

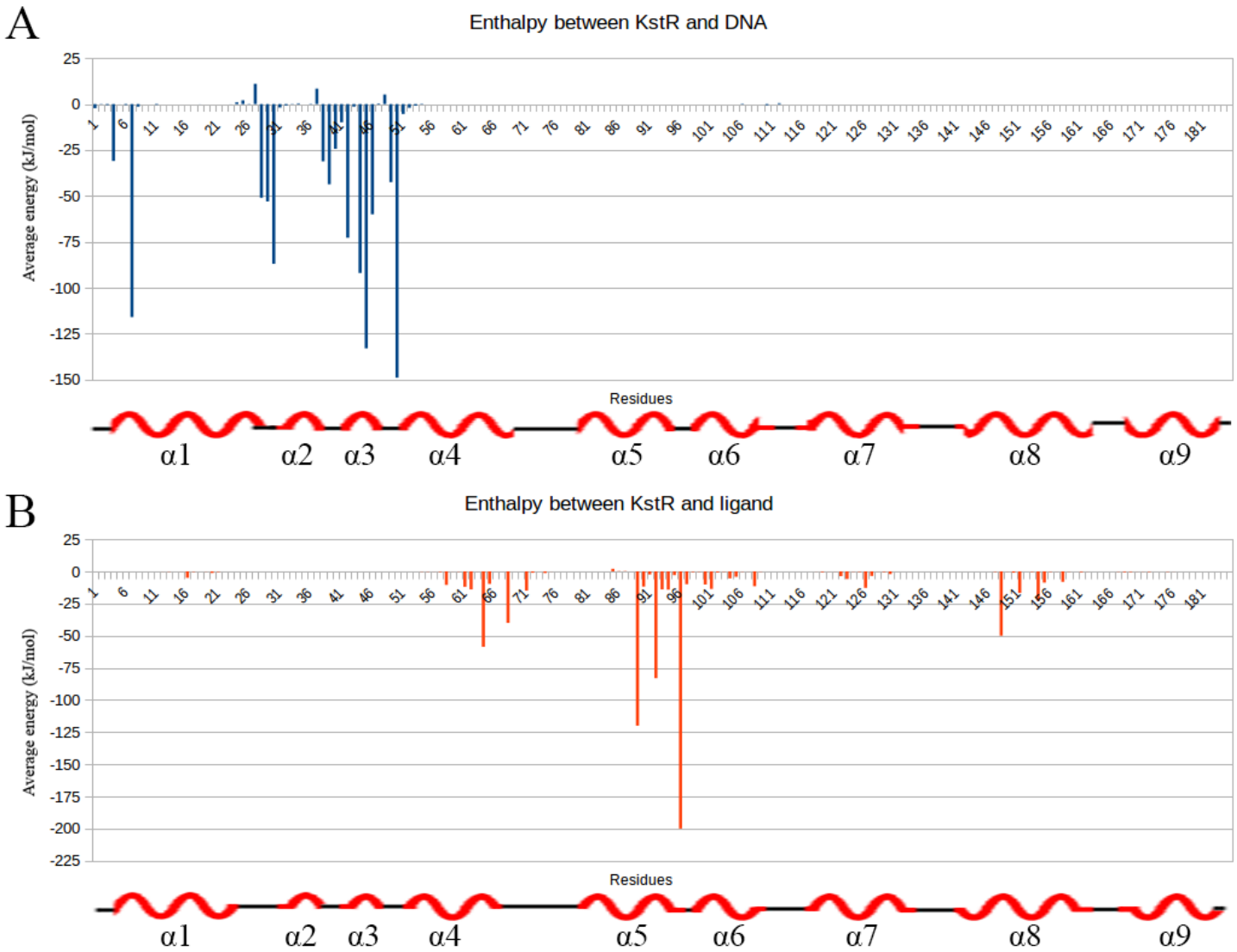

Figure 3.6.1: *A shows the average enthalpy energy between each residue of KstR and DNA, while B shows the enthalpy between each residue of KstR and ligand. Cartoon at the bottom of the figure illustrates helices (red) and loops (black).*

3.6.2 Enthalpic contributions between mutated KstR and DNA

Gln-38, Arg-40, and Tyr-54 have been highlighted as residues which have crucial polar contacts with the phosphate backbone of KstR (*Ho et al. manuscript in preparation*). Specific residues across the KstR DBD were mutated and simulations run of each system to determine to what degree these residues impact the enthalpy between KstR and DNA. Table 3.6.2 describes which KstR residues were mutated to Ala.

Table 3.6.1: *System names and residues mutated to Ala.*

Mutations included in each system	
System	Mutations
LoDB	N/A
Q38A	Gln-38
R40A	Arg-40
Y54A	Tyr-54
Q38A/R40A	Gln-38 and Arg-40
Q38A/Y54A	Gln-38 and Tyr-54
R40A/Y54A	Arg-40 and Tyr-54
Q38A/R40A/Y54A	Gln-38, Arg-40, and Tyr-54

Table 3.6.3 shows that R40A and Y54A resulted less favourable interaction between KstR and DNA while Q38A resulted in a more favourable enthalpic contribution. This indicates that Ala at residue 38 improves binding with DNA while mutating Arg-40 and Tyr-54 to Ala results in a weaker binding, Arg-40 more so.

In cases where two residues were mutated R40A/Y54A displayed the least favourable interactions with DNA. This indicates that mutating Gln-38 has little impact on how KstR interacts with DNA while Arg-40 and Tyr-54 contribute significantly. This was further demonstrated with Q38A/R40A/Y4A where the difference in enthalpic potential between R40A/Y54A and R38A/R40A/Y54A is small. This shows that once Arg-40 and Tyr-54 were mutated there was a drastic change in enthalpy while Gln-38 mutation has little impact on the overall binding relationship between KstR and DNA.

Table 3.6.2: *Average and error estimate of the contribution of KstR-DNA enthalpy to overall enthalpy between KstR and DNA for LoDB or all mutated KstR respectively over time (0 - 200 ns).*

Enthalpy between KstR and DNA		
	Average (kJ/mol)	Error estimate
LoDB	-1980	31
Q38A	-2000	12
R40A	-1820	12
Y54A	-1940	32
Q38A/R40A	-1770	18
Q38A/Y54A	-1760	55
R40A/Y54A	-1580	23
Q38A/R40A/Y54A	-1590	21

By comparing the enthalpy between only the residue that is mutated and the DNA, the impact of the mutation is determined.

Residue 38 has less favourable interactions with DNA when mutated to Ala (Table 3.6.3). Although R40A/Y54A does not possess a residue 38 mutation it also results in a less favourable interaction with between residue 38 and DNA. While Q38A/R40A/Y54A possesses a mutation of all three however have a less favourable enthalpy between residue 38 and DNA. This indicates that while Q38A in isolation cause DNA binding with KstR to be less favourable it does not operate synergystically with R40A and Y54A mutations to affect enthalpy.

Mutation of Arg-40 to Ala results in substantially less favourable enthalpic interactions between residue 40 and DNA (Table 3.6.3) . Mutations which did not affect Arg-40 resulted in more favourable interactions between residue 40 and DNA. Where Arg-40 was mutated the enthalpy between residue 40 and DNA was less favourable by over 90%. The Q38A, Y54A, and Q38A/Y54A all resulted in a more favourable enthalpy between residue 40 and DNA.

Only systems where Tyr-54 was mutated alone exhibit a less favourable interaction between residue 54 and DNA (Table 3.6.3). The enthalpy between residue 54 and KstR is not impacted by mutations at residue 38 or residue 40.

Table 3.6.3: *Average and error estimate of enthalpy between DNA and individual residues over time (0 - 200 ns).*

Enthalpy energy between KstR Residue 38 and DNA		
	Average (kJ/mol)	Error estimate
LoDB	-51	2
Q38A	-31	2
R40A	-58	2
Y54A	-58	2
Q38A/R40A	-26	3
Q38/Y54AA	-21	2
R40A/Y54A	-20	2
Q38A/R40A/Y54A	-30.3	0.3
Enthalpy energy between KstR Residue 40 and DNA		
LoDB	-87	3
Q38A	-115	5
R40A	-6.3	0.3
Y54A	-114	2
Q38A/R40A	-10	1
Q38A/Y54A	-160	21
R40A/Y54A	-8.5	0.2
Q38A/R40A/Y54A	-7.9	0.1
Enthalpy energy between KstR Residue 54 and DNA		
LoDB	-92	2
Q38A	-93.3	0.7
R40A	-90	2
Y54A	-0	0
Q38A/R40A	-88	2
Q38A/Y54A	-0.145	0
R40A/Y54A	-0	0
Q38A/R40A/Y54A	-0.20	0.4

Together this demonstrates that Arg-40 and Tyr-54 contribute substantially to the interaction between KstR and DNA while Gln-38 scarcely does. This indicates that Arg-40 and Tyr-54 are important residues for KstR to bind with DNA.

Chapter 4

Discussion

4.0.1 Limitations of simulation procedure and analysis

There are several points to keep in mind when interpreting the MD results.

Firstly, there are slight differences between each system that affect how comparable they are. Namely, that each system has a unique dispersion of velocities by which their conformational trajectory may have been skewed. Furthermore, no system is identical to the other in terms of atom composition as shown in Table 2.1.1 and this may skew their molecular dynamics.

Secondly, while all simulations were limited to either 100 ns or 200 ns, in general molecular dynamics benefits from longer simulations. For the ligand removed (Section 3.2) and DNA removed (Section 3.3) systems, longer simulations would have better indicated if KstR ever fully takes on *apo* KstR characteristics. In both Sections 3.2.1 and 3.2.1 that where ligand or DNA is removed from KstR, the shape of KstR becomes more similar to that of *apo* KstR but still not close enough to indicate it is identical. Upon the removal of ligand or DNA from KstR, aside from water and ion atom number the composition of the system would be the same- thereby we would expect the ligand or DNA removed KstR to eventually reproduce *apo* KstR conformational dynamics. The same is seen for Sections 3.2.2 and 3.3.2 where upon removal of ligand or DNA KstR begins to take on a larger pool of conformations for shorter time periods as if transitioning to *apo* KstR. Sections 3.2.4 and 3.3.4 also show that the removal of ligand or DNA from KstR causes the distances between domains to become more like those of *apo* KstR although not identical at the simulation timescale. Given a longer timescale it would be expected that the ligand or DNA removed simulations would reproduce the conformational dynamics of *apo* KstR.

Thirdly, endogenous KstR interacts with DNA that is not fragmented. The conformational dynamics of DNA when bound to KstR are also important and specific (*Ho et al. manuscript in preparation*). However, a 26 bp fragment of DNA does not accurately perform or represent the conformational dynamics of unfragmented DNA so is a insufficient approximation.

Fourthly, it has been shown that the steroid, 3OChA, can not bind to KstR

without the CoA tail, and therefore modelling the conformational dynamics of short ligand-bound KstR is purely speculative [4]. A comparison between 3OChA-CoA (FLB) and 3OChA (SLB) was pursued to determine if the CoA tail causes a unique effect on KstR once it is bound. It was found that the contribution of the CoA tail is slight if any.

4.1 Conformation of KstR is affected by ligand or DNA binding

Sections 3.1, 3.2 and 3.3 investigated the conformation and dynamics of KstR when bound to either ligand-CoA or DNA or neither, as sought out by objective 1. This has shown that ligand and DNA binding with KstR induces KstR to take on distinct conformations.

4.1.1 Ligand or DNA bound KstR introduce distinctive KstR conformations

The findings from 3OChA-CoA and DNA-bound KstR simulations reinforced the observations from the KstR crystal structures [4] [54], showing that 3OChA-CoA binding induces KstR to take on larger shapes, and DNA binding induces KstR to take on compact shapes.

When ligand is bound to KstR it samples larger shapes than *apo* KstR. The average SASA and average *R*g of KstR when bound to ligand were larger than that of *apo* KstR (Sections 3.1.1 and 3.4.1). Conversely, DNA had a pronounced contracting effect on the shape of KstR compared to *apo* KstR (Sections 3.1.1 and 3.5.1). Furthermore, the volume of KstR was subtly different between ligand-bound, DNA-bound and *apo* KstR, although not large enough to explain the differences in SASA and *R*g. Therefore, changes measured by KstR *R*g and SASA were a function of shape as opposed to size. The *apo* KstR range of shapes frequently overlapped with ligand-bound KstR while rarely overlapping with DNA-bound KstR. This shows that it is unfavourable for the shape of DNA-bound KstR to be sampled by KstR if DNA is absent. However, *apo* KstR is not excluded from exhibiting ligand-bound KstR shapes. Unfortunately, the shape measurements in themselves do not demonstrate which regions of KstR are responsible for the changes in the shape of KstR.

The effects of ligand or DNA binding on KstR structure were also shown by the measurements of distance between residue pairs in KstR. Ligand-bound KstR favoured larger distances which changed more between residue pairs while DNA binding induced smaller distances which changed less. This demonstrates that ligand and DNA binding affect the distance between residue pairs within KstR in distinct and opposing ways (Sections 3.1.4, 3.4.4, and 3.5.4). Furthermore, the effects ligand or DNA binding to KstR had on the distance

between residue pairs were more pronounced at the DBD compared to the LBD. Despite the low number of residue pairs sampled, and sparse dispersion of residue pairs within the LBD which limit the robustness of the findings, these results demonstrated that DNA or ligand binding had opposing affects on KstR, and that while DNA or ligand bind at different sites on KstR the DBD is what exhibits the change and not the LBD.

The impact of ligand binding on KstR conformation is shown by how ligand alters the ligand binding cavity (Sections 3.1.1, 3.2.1, and 3.4.1). Where ligand is present in the ligand binding cavity the ligand binding cavity exhibits a larger average volume. However, when the ligand was shortened the largest volume of the ligand binding cavity increased. The trj cavity program only quantifies the volume at the pit of the pocket of the ligand binding cavity which both 3-OChA and 3OChA-CoA occupy. Therefore contributions from the CoA tail of the ligand towards the ligand binding cavity volume are not accounted for (Figure 2.2.6). Nevertheless, what is demonstrated is that the volume of the ligand binding cavity, and therefore conformation of the LBD, is affected by the presence of ligand.

Ligand or DNA binding induce KstR to take on different conformational behaviour. Ligand-bound and *apo* KstR yielded a greater number of conformational ensembles, which were sampled briefly (Sections 3.1.2 and 3.4.2). DNA-bound KstR exhibited far fewer conformational ensembles, generally residing within one cluster for the duration of the simulation (Sections 3.1.2 and 3.5.2). While there were differences in conformational ensembles between ligand-bound and *apo* KstR, these differences were subtle compared to the difference between *apo* and DNA-bound KstR conformational ensembles.

4.1.2 Ligand or DNA binding affect how frequently KstR conformation can change

RMSF, standard deviation and rate of transition between clusters were used to measure the structural flexibility of KstR. These measurements showed that the flexibility of KstR was affected by DNA or ligand binding. The variation in flexibility induced by ligand binding has been documented for TFRs [39] [53]. MD studies of the conformational dynamics and intramolecular binding of TetR elucidated the TetR mechanism of action [88]. The TetR MD research showed that upon tetracycline binding to the tetracycline binding domain of TetR, the shape of TetR is made more flexible and takes on more diverse motions, while DNA binding to TetR causes the conformation of TetR move less freely indicating that DNA restricts the dynamics of TetR [88]. The MD findings for KstR here demonstrate that ligand or DNA binding affect the conformational dynamics of KstR in similar ways to how they affected TetR.

The RMSF of KstR is affected by ligand or DNA binding. Ligand binding to KstR enhanced the flexibility of KstR, most prominently at the LBD and around the ligand binding cavity (Sections 3.1.3 and 3.4.3), while DNA-bound

KstR induced a less flexible KstR structure throughout the DBD and around the ligand binding cavity (Sections 3.1.3 and 3.5.3). The differences in flexibility were more pronounced for DNA-bound KstR than ligand-bound KstR. This may be a consequence of DNA being larger than ligand and therefore introducing a larger enthalpy between KstR. Together this demonstrates that DNA binding or ligand binding had opposing effects on the flexibility of KstR and affected different regions of KstR. Otherwise the difference in flexibility between ligand bound and DNA bound KstR may be due to DNA restraining the most flexible and larger regions of KstR.

Aspects of KstR flexibility could also be gleaned from the conformational clustering of ligand-bound and DNA-bound KstR. Ligand-bound KstR frequently fluctuated between conformations within its conformational ensemble while DNA-bound KstR resided within one conformation among its few ensembles for the majority of the simulation (Sections 3.1.2, 3.4.2, and 3.5.2). This shows that ligand-bound KstR changes more frequently than DNA-bound KstR.

The standard deviation values when measuring shape metrics of KstR also provide insight into how ligand and DNA binding affect the flexibility of KstR. The standard deviation of properties that report the shape of ligand-bound KstR were greater than those of *apo* KstR (Sections 3.1.1 and 3.4.1). Shape measurements of DNA-bound KstR exhibited smaller standard deviation than those of *apo* KstR (Sections 3.1.1 and 3.5.1). However, the standard deviation for residue pairs between KstR monomers did not reproduce this trend as the standard deviation for *apo* KstR was greater than that of ligand-bound KstR (Sections 3.1.4 and 3.4.4). This indicates that ligand binding affects KstR flexibility through a culmination of movements resulting in a greater conformational flexibility compared to *apo* KstR. Together this shows that ligand-binding to KstR induced a more flexible KstR structure while DNA-binding to KstR induced a less flexible KstR structure.

4.1.3 Ligand or DNA binding affect the intramolecular organisation of KstR

Ligand and DNA binding induced similar changes in the secondary structure of KstR. Ligand or DNA-bound KstR favour larger helices compared to *apo* KstR (Sections 3.1.6, 3.4.6 and 3.5.6). However, while ligand and DNA-bound KstR both favour helices compared to *apo* KstR, DNA-bound KstR is compromised of longer helical domains than ligand-bound, although the difference in helices composition between each system is slight. This indicates that the secondary structure of KstR may be a consequence of the enthalpy between ligand or DNA and KstR. MD experiments have investigated how the helix-stability of a soluble molecule like KstR is dependent on crowding agents and entropic factors [89]. The results presented here, the fact that DNA has a larger favourable enthalpic interactions with KstR propagates structural changes within KstR such as differences in the length of helical domains.

The inter-residue distance of KstR was also affected by ligand or DNA binding. Ligand-bound KstR had fewer inter-residue contacts at the DBD and the LBD and DBD (Sections 3.1.5 and 3.4.5), while DNA binding resulted in more inter-residue contacts within the LBD and between the DBD and LBD (Sections 3.1.5 and 3.5.5). Changes in intramolecular arrangement were most prominent between domains of the LBD and DBD opposed to within the LBD or DBD. This demonstrates that ligand binding reduces the intramolecular bond network of KstR favouring it to take on larger shapes while DNA binding to KstR has the opposite effect on the shape of KstR, favouring a more contracted and smaller shape.

The calculations of enthalpy between KstR and ligand or DNA present to what degree binding events are favoured by the system. DNA produces a much greater enthalpy energy between KstR than ligand does (Sections 3.1.8, 3.4.8 and 3.5.8). This shows why DNA has a greater impact on the intramolecular organisation of KstR than ligand does. Although, as mentioned earlier the enthalpy between KstR and DNA is almost entirely located at the DBD while the enthalpy between KstR and 3OChA-CoA is more evenly dispersed between the LBD and DBD. Which demonstrates that 3OChA-CoA has a direct impact on both domains of KstR while DNA only directly impacts the DBD.

4.2 Removal of ligand or DNA causes KstR to take on apo KstR-like features

It was found that ligand or DNA binding affected the flexibility of KstR, as sought out by objective 2. In addition, once DNA or ligand was removed from KstR, KstR took on conformational dynamics more like *apo* KstR.

4.2.1 Ligand or DNA removal induced apo KstR-like conformations

The removal of 3OChA-CoA from KstR caused the average Rg and SASA of KstR to become smaller and closer to that of *apo* KstR while the average volume of KstR did not change (Section 3.2.1). The removal of DNA from KstR caused the average Rg and SASA of KstR to become larger like that of *apo* KstR, while the average volume of KstR did not change (Section 3.3.1). Removal of 3OChA-CoA or DNA has a similar effect on the distance between residue pairs. The average distance between the residue pairs decreased when 3OChA-CoA was removed while the average distance between residues increased upon DNA removal (Sections 3.2.4 and 3.3.4). Furthermore, the differences were more pronounced for residue pairs at the DBD than they were for those at the LBD. This shows that removal of either 3OChA-CoA or DNA from KstR caused the shape of KstR to become more *apo* KstR-like.

Furthermore, the DNA-bound and ligand-bound KstR conformational dynamics can only be achieved if either DNA or ligand is bound to KstR and are not frequently sampled within the conformational behaviour of *apo* KstR.

The conformational clusters when 3OChA-CoA or DNA are removed from KstR inform how the conformation of KstR is affected. When either 3OChA-CoA or DNA were removed KstR sampled the largest pool of conformations ensembles (Sections 3.2.2 and 3.3.2). The 3OChA-CoA removed KstR sampled more ensembles, more briefly than both the 3OChA-CoA-bound or *apo* KstR. Removal of 3OChA-CoA and DNA did not result in KstR sampling the same conformational clusters as *apo* KstR but this is most likely due to the limited length of the simulations. This trend is also reflected in how the ligand binding cavity is affected by removal of 3OChA-CoA. The volume of the ligand binding cavity never completely replicated the volume of *apo* KstR. Ypon 3OChA-CoA removal from KstR, the ligand binding cavity immediately collapsed and then retained an intermediate volume between 3OChA-CoA-bound and *apo* KstR (Section 3.2.1). This also demonstrates that the reorganisation of the ligand binding cavity occurs on a longer timescale than the simulation allowed for.

4.2.2 Ligand or DNA removal affects the flexibility of KstR

When 3OChA-CoA was removed from KstR some regions of KstR became more flexible while others less. 3OChA-CoA removal from KstR caused regions local to the ligand binding cavity became less flexible, like that of *apo* KstR while those distant to become more flexible than 3OChA-CoA-bound KstR (Section 3.2.3). This shows that 3OChA-CoA removal perturbs the flexibility of the DBD more than the flexibility of the LBD, showing aspects of the KstR mechanism of action: when 3OChA-CoA is bound to KstR the specific conformational dynamics required for KstR binding to DNA are unlikely to occur. DNA removal from KstR enhanced the flexibility of large portions of the KstR (Section 3.3.3). This demonstrates that DNA impacts more of the KstR DBD and LBD than 3OChA-CoA. Furthermore, some residues remain unchanged upon the removal of 3OChA-CoA or DNA from KstR indicating that 3OChA-CoA and DNA only manipulate the flexibility of KstR at specific domains.

How often KstR moves among its conformational ensembles within a cluster indicates the flexibility of KstR. 3OChA-CoA removal from KstR causes KstR to change conformation more often than 3OChA-CoA-bound or *apo* KstR (Section 3.2.2). Upon DNA removal from KstR, KstR more frequently fluctuates between its ensembles compared to both DNA-bound and *apo* KstR (Section 3.3.2). This indicates that once either 3OChA-CoA or DNA is removed from KstR the structure of KstR becomes more flexible causing it to sample different ensembles more, taking on dynamic conformational behaviour like that of *apo* KstR.

The removal of 3OChA-CoA or DNA causes more pronounced differences in distances between residue pairs that transition on longer timescales than those of overall KstR shape (Sections 3.2.1, 3.2.4, 3.3.1 and 3.3.4). This indicates that the overall shape of KstR responds more quickly to changes in binding mode than individual domains do.

4.2.3 Ligand or DNA removal affects the intramolecular organisation of KstR

3OChA-CoA removal perturbed the structure of KstR in novel ways. 3OChA-CoA removal from KstR reduced the helical composition of KstR and presented a compact shape comparable to *apo* KstR (Section 3.2.6). Furthermore, 3OChA-CoA removal from KstR induced a small helices to form between $\alpha 1$ and $\alpha 2$. It is unclear what drove the formation of the small helices although it demonstrates that 3OChA-CoA removal perturbs the organisation of residues at the DBD. Similarly, DNA removal from KstR caused KstR to take on more *apo* KstR like secondary structure composition (Section 3.3.6). In both cases the removal of 3OChA-CoA or DNA from KstR was the impetus to cause more *apo* KstR-like secondary structure features to emerge.

The intramolecular contacts within KstR demonstrate that 3OChA-CoA or DNA removal from KstR cause *apo*-like KstR features. Upon 3OChA-CoA removal from KstR more contacts and closer contacts within the DBD and between the LBD and DBD emerge (Section 3.2.5). Upon DNA removal from KstR fewer contacts and longer distances within the LBD and between the DBD and LBD were produced (Section 3.3.5).

4.3 Shortening of ligand or DNA has a small effect on KstR

The conformational dynamic differences between 26 bp and 18 bp bound KstR are indiscernible, as are the differences between or 3OChA-CoA or 3OChA bound KstR, as sought out by objective 3.

4.3.1 Ligand or DNA shortening scarcely impact the shape of KstR

The shape and size of KstR was not affected by the size of ligand or DNA bound to KstR. The average Rg, SASA, and volume for 3OChA-bound KstR was slightly larger than that of 3OChA-CoA-bound (Section 3.4.1). The average Rg, SASA, and volume for short DNA-bound KstR was not different than that of long DNA-bound KstR (Section 3.5.1). This indicates that when the size of either ligand or DNA bound to KstR become smaller KstR does not

change in shape or size.

The average distance between residue pairs of KstR when ligand or DNA bound to KstR were shortened did not change. 3OChA-bound KstR induced insignificantly larger average distances between residues at the DBD of KstR compared to 3OChA-CoA-bound KstR, there was no consistent impact at the LBD (Section 3.4.4). Where DNA bound to KstR was shortened the average distance between residue pairs at the DBD were the same as those of long DNA-bound KstR, and differences at the LBD were inconsistent (Section 3.5.4). This exhibits that a smaller ligand or DNA did not induce significant differences in distance between residue pairs at across KstR.

The conformational clustering between 3OChA and 3OChA-CoA-bound KstR, or short DNA-bound and long DNA-bound KstR never shared a single conformation (Sections 3.4.2 and 3.5.2). This indicates that the conformational ensemble pool of KstR is not obviously altered if ligand or DNA are smaller. Additionally, the shape, size, flexibility measurements all also indicate that the changes made to ligand or DNA had little impact on the conformational dynamics of KstR (Sections 3.4.1, 3.4.3, 3.5.1 and 3.5.3). This suggests the conformational cluster RMSD cut-off of 0.2 nm may have been too small and sensitive to group similarities in conformation (3.4.1 and 3.5.1)

The shortening of ligand bound to KstR did not have an immense impact on the volume of the ligand binding cavity. The 3OChA-KstR resulted in a slightly smaller average ligand binding cavity volume although did sample the largest ligand binding cavity volumes (Section 3.4.1). The ligand binding cavity volume calculation tool, trj cavity, does not recognise volumes near the opening of the ligand binding cavity as a cavity and therefore does not calculate it. As a consequence much of the 3OChA-CoA contribution towards cavity size is left unappreciated.

4.3.2 Ligand or DNA shortening scarcely affects the flexibility of KstR

3OChA or shorter DNA bound to KstR affects the flexibility of KstR. For 3OChA-bound KstR the RMSF in a small region of the DBD and about the ligand binding cavity was reduced compared to 3OChA-CoA-bound KstR (Section 3.4.3). When DNA bound to KstR was shortened the flexibility of KstR increased at linker loops across the LBD and $\alpha 5$ (Section 3.5.3). This exhibits that 3OChA bound to KstR causes KstR to become less flexible than when bound to 3OChA-CoA, while shortening of DNA bound to KstR causes KstR to become more flexible compared to when bound to long DNA.

The standard deviation for size and shape measurements of 3OChA or shortened DNA-bound KstR also inform how shortening ligand or DNA affects the structure of KstR. The standard deviation for shape measurements of KstR were smaller for 3OChA-bound KstR compared to 3OChA-CoA-bound KstR

(Section 3.4.1). This indicates that 3OChA-CoA binding causes the structure of KstR to fluctuate more frequently than 3OChA-bound KstR. The shorter DNA produced greater standard deviation on shape measurements of KstR than long DNA (Section 3.5.1). This indicates that a shorter DNA allows the shape of KstR to change more often than when bound to longer DNA. Although, in all cases the difference in standard deviation is only slight.

4.3.3 Ligand or DNA shortening has a small impact on the intramolecular bonding of KstR

Differences in secondary structure and in the nature of intramolecular interactions between 3OChA-bound KstR and 3OChA-CoA-bound KstR are trivial (Section 3.4.5).

DNA shortening introduced differences in the secondary structure of KstR compared to long DNA-bound KstR. Most prominent were the differences in helical assignment, that upon DNA shortening KstR favoured a greater composition of helical residues than long DNA-bound KstR (Section 3.5.6).

4.4 Enthalpic contributions between DNA or ligand and KstR

The enthalpy was used to identify how ligand or DNA affect the conformational dynamics of KstR. This also highlighted residues thought to be important in the mechanism of KstR, as sought out by objective 4.

4.4.1 DNA has greater enthalpic interactions with KstR than ligand

The enthalpy between KstR and DNA is larger than that between KstR and 3-OChA-CoA (Section 3.1.8). The reduced conformational sampling of DNA-bound KstR reflects a drop in entropy, which comes at a energetic cost. This cost seems to be balanced by the favourable enthalpic interactions between KstR and DNA. The enthalpic interactions with DNA are almost entirely concentrated among residues in the DBD (99.7%), while the enthalpy between 3OChA-CoA and KstR is more evenly dispersed among the two domains: 28.9% at the DBD, while 71.1% is dispersed among the LBD. Although, the entropy between ligand or DNA and KstR is also affected by the proximity: enthalpy is only calculated for atom pairs which are within 1.2 nm of one another to the enthalpy. Additionally, the size of the contributions decreases with the distance between two atoms. Furthermore, the enthalpy between ligand or DNA and KstR is also a function of the size of ligand or DNA. Therefore a smaller ligand or DNA will implicitly generate a smaller enthalpy between

KstR (Sections 3.4.8 and 3.5.8).

4.4.2 Residues important for the enthalpy between KstR and ligand or DNA

Specific residues contribute more towards the enthalpy between KstR and 3OChA-CoA or DNA more than others. Residues within the $\alpha 1$ helix of KstR have a favourable interaction between KstR and 3OChA-CoA, although no others in the DBD contribute. For the LBD residues within the $\alpha 4$, the $\alpha 4$ - $\alpha 5$ linker domain, $\alpha 5$, $\alpha 6$, $\alpha 7$, and $\alpha 8$ contributed towards the enthalpy between KstR and 3OChA-CoA (Section 3.6.1). Most prominently were domains local to the 3OChA-CoA and residues within the $\alpha 4$ - $\alpha 5$ linker domain. The residues which presented substantial favourable enthalpy contributions between ligand (less than -30kJ/mol) include: Arg-74, Lys-78, Lys-99, Arg-102, Arg-106, and Arg-158, illustrated on Figure 4.4.1. This shows that the greatest enthalpy between the residues and 3OChA-CoA are residues which are long positively charged residues local to the 3OChA-CoA. However, there are the residues within $\alpha 1$ which are distant from 3OChA-CoA yet still produce an enthalpy.

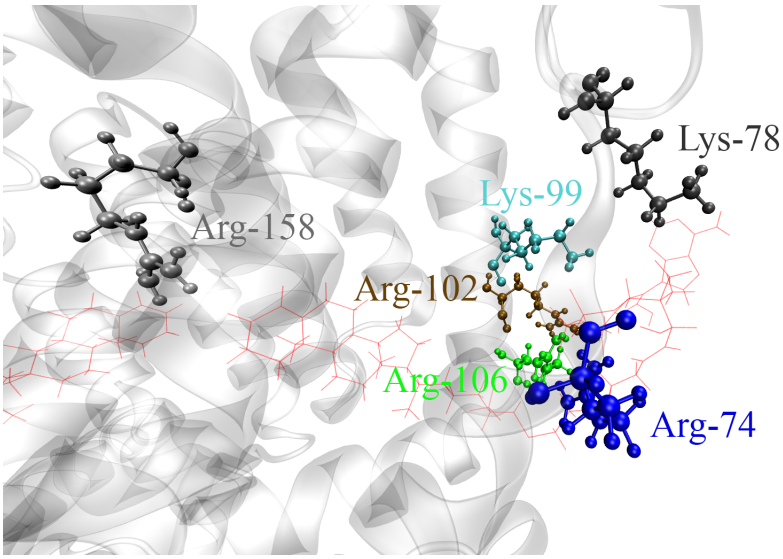

Figure 4.4.1: *Annotated cartoon showing residues which have the greatest contribution to the enthalpic interactions between 3OChA-CoA and KstR.*

Most of the largest enthalpy contributions between KstR and DNA occurred around $\alpha 3$, those which are prominent (less than -30 kJ/mol) include: Gln-14, Arg-17, Gln-38, Met-39, Arg-40, Val-48, Ala-49, Thr-52, Tyr-54, Arg-55, Tyr-56, Ser-59, and Lys-60, as seen on Figure 4.4.2.

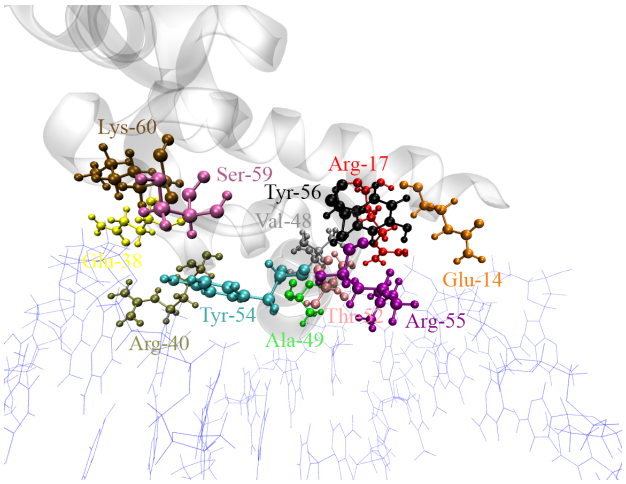

Figure 4.4.2: *Annotated cartoon showing residues which have the greatest contribution to the enthalpic interactions between DNA and KstR.*

4.4.3 Key residues required for DNA binding

Several residues have already been proposed as important in the binding interaction between DNA and KstR. These include Gln-38, Arg-45, and Tyr-54 which bind to the DNA phosphate backbone (*Ho et al. manuscript in preparation*).

Computational mutational studies were pursued to investigate to what degree these residues are essential for DNA binding. Variants of KstR in which mutations of Gln-38, Arg-40, and Tyr-54 or combinations of the three were simulated. In all simulations where Arg-40 and Tyr-54 were mutated the enthalpy between KstR and DNA was attenuated while when Gln-38 was mutated there was only a subtle difference in enthalpy (Section 3.6.2). Furthermore, mutation of Gln-38 in combination with either Arg-40 or Tyr-54 did not reduce the enthalpy between KstR and DNA anymore than mutation of Arg-40 or Tyr-54 did alone. This shows that Arg-40 and Tyr-54 are essential for strong DNA binding whereas Gln-38 is not.

4.5 Conformational selection and induced fit

The behaviour of KstR presents elements of both conformational selection and induced fit. It has been shown in previous studies that ligand binding presents an induced fit model to TFR proteins [39] [46].

4.5.1 Conformational selection

The MD simulations of KstR behaviour shown that KstR demonstrates aspects of the conformational selection model of binding, as explained by the following discussion [46] [90].

The shape of *apo* KstR can oscillate within the range of 3OChA-CoA-bound and DNA-bound KstR, demonstrating it is possible for the shape of *apo* KstR to exhibit shapes of either ligand or DNA-bound KstR (Section 3.1.1). However it is only the extremes of the *apo* KstR distribution of Rg and SASA which overlap with those of 3OChA-CoA-bound or DNA-bound KstR. This demonstrates that while 3OChA-CoA-bound and DNA-bound KstR have exclusive conformational dynamics, *apo* KstR can sample conformations of either and perhaps select for the conformation if either 3OChA-CoA or DNA are within proximity.

Measurements of the ligand binding cavity also suggest that KstR might function via conformational selection. *Apo* KstR exhibits ligand binding cavity volumes within the range of 3OChA-CoA-bound KstR ligand binding cavity volumes (Section 3.1.1). At different times the ligand binding cavity is either open or closed, displaying a breathing like motion [**?**]. This demonstrates that 3OChA-CoA is not volumetrically excluded from occupying the *apo* KstR ligand binding cavity. However, conformational selection depends on *apo* KstR forming the right ligand binding cavity conformational shape and that the 3OChA-CoA is primed for moving into the cavity.

4.5.2 Induced fit

It has been shown that TFRs demonstrate induced fit binding models [37] [44] [46]. Aspects of KstR behaviour also demonstrate the induced fit model of binding, as evidenced by the following data.

Ligand binding to KstR induces KstR to take on unique larger and more frequently fluctuating conformational dynamics (Sections 3.1.1 and 3.1.4). It is demonstrated that ligand is what induces this change as once ligand is removed from KstR, KstR takes on more *apo* KstR like dynamics (Sections 3.2.1 and 3.2.4). DNA binding to KstR induces unique conformational dynamics of KstR, making it smaller and less flexible. Furthermore, it is shown that DNA is what induces the conformational dynamics of KstR as upon removal of DNA from KstR is immediately reverts to more *apo* KstR like behaviour (Section 3.3.1 and 3.3.4). Together this demonstrates that it is binding with ligand or DNA that induces the conformational dynamics which are different to each other and those of *apo* KstR.

When ligand is absent Trp-164 takes on a different orientation compared to when ligand is bound (Sections 3.1.7, and 3.2.7). Furthermore, the orientation of Trp-164 is less mobile when 3OChA-CoA is bound compared to 3OChA (Section 3.4.7). This demonstrates that the orientation of Trp-164 in the KstR ligand binding cavity was induced as a consequence of ligand binding.

When DNA is bound to KstR Tyr-54 takes on a unique and consistent dihedral angle while when it is absent it fluctuates within a range close to 360° (Sections 3.1.7 and 3.3.7). Furthermore, when DNA was shortened Tyr-54 was

more mobile (Section 3.5.7). This demonstrates that the orientation of Tyr-54 was induced by the presence of DNA, which may have greater ramifications for other regions of KstR.

The two different systems never sampled the same conformational cluster, aside from the first ns, between LoDB and SLDB KstR (Sections 3.1.2, 3.2.2, 3.3.2, 3.4.2, and 3.5.5). Ligand-bound KstR results in the most changeable structure and greatest number of conformational ensembles, *apo* KstR presents similar behaviour but not to the same extent, while DNA-bound KstR had the smallest number of conformational clusters and rarely changed. This indicates that the ligand or DNA binding induces KstR to occupy different pools of conformational ensembles than each other and *apo* KstR

4.6 KstR mechanism of action

The introduction outlined the hypothesised mechanism of action of KstR, shown in Figure 1.2.1.

As has been discussed, KstR takes on specific conformations depending on what it is bound to (Section 3.1.1 and 3.1.2). It has also been shown that the removal of ligand or DNA causes KstR to take on *apo* KstR like conformational properties (Section 4.5.2). Additionally, it was shown that ligand binding induces larger and more flexible KstR shapes while DNA binding induces contracted and less flexible KstR shapes, ligand or DNA binding have opposing effects on the conformation of KstR. Together this demonstrates that it is ligand or DNA that skew the conformational dynamics of KstR causing it to either bind to DNA and disallow transcription of the DNA, or bind to ligand and disallow KstR interaction with DNA.

When KstR is bound to DNA, the ligand binding cavity is larger than that of *apo* KstR on average and is never closed (Section 3.1.1). This indicates that DNA binding at the DBD induces the shape of the ligand binding cavity to take on distinct conformations. It also indicates that perhaps the opportunity for ligand to bind to KstR is enhanced when KstR interacts with DNA.

The availability of the ligand binding cavity also informs the mechanism of action. When DNA binds to KstR the average ligand cavity binding volume is larger than that of *apo* KstR, within the range of ligand binding cavity volumes of 3OChA-CoA-bound KstR, and never closes (Section 3.1.1). DNA binding to KstR induces the shape of the ligand binding cavity to be left available for 3OChA-CoA to bind to KstR. This suggests ligand may still be able to bind while KstR is DNA-bound, making the cell more responsive to increases in ligand concentration.

The mechanism of action of KstR is also suggested through how the ligand or DNA affect KstR flexibility is altered. Ligand or DNA binding have opposing effects on the flexibility of KstR that enhance their proposed function (Section 4.5.1). It is speculated that the DNA phosphate backbone must be

in a specific conformation along with KstR in a contracted DBD conformation to binding between DNA and KstR [4]. Therefore ligand-bound KstR is unlikely to bind to DNA as ligand induces larger distances between the DBD and change structure more frequently. Consequently, ligand-bound KstR occludes KstR from taking on the correct shape, and in the unlikely event it would it took the correct conformation, it would only be for a brief amount of time before changing again. Together this reduces the possibility of taking on the specific conformation required for DNA binding.

It is also shown that ligand-removed KstR, or DNA-removed KstR conformational behaviour represents the intermediate state between ligand-bound or DNA-bound KstR, respectively, and *apo* KstR. Ligand-bound and DNA binding conformational properties rarely sample within the same range, however *apo* KstR often exhibits shapes within the range of both (Section 3.1.1, 3.2.1, and 3.3.1). This indicates that *apo* KstR has a conduit-like role within the mechanism of KstR. *Apo* KstR is capable of conformations of either ligand-bound and DNA-bound KstR and thereby either binding to ligand or DNA.

The mechanism of action is also reinforced by how the intramolecular interactions of KstR are affected by ligand or DNA. Ligand binding favours the dissociation and general erosion of intramolecular contacts whereas DNA favours the construction and intensification of intramolecular contacts (section 3.1.5, 3.2.5 and 3.3.5). Interestingly the change in intramolecular contacts as a consequence of DNA or ligand binding is largely centred within the LBD or between the DBD and LBD. However it is shown that the DBD is where conformational changes are most pronounced. This indicates that it is the LBD where changes take place but the DBD which manifests the changes in the form of conformational rearrangement.

Together this demonstrates that DNA, ligand and *apo* KstR take on distinctive conformations along a spectrum. Ligand-bound KstR induces more expanded conformations, and DNA-bound KstR induces more contracted conformations, and while *apo* KstR shape oscillates between the two.

4.6.1 Allosteric mechanism

The presence of allosteric communication between regions of TetR is known and has been investigated [38] [91] [92]. The presence of allosteric communication between regions of the KstR has been speculated [4]. This research indicates that the allosteric mechanism of KstR is a consequence of changes at the LBD affecting the conformation and dynamics of the DBD, while changes at the DBD have little bearing on the conformation and dynamics of the LBD. Fundamentally, the KstR DBD is a slave to the LBD, while the LBD provides the impetus to manifest change at the DBD.

The intramolecular structure of KstR is affected by DNA or ligand (Section 3.1.5). 3OChA-CoA binding causes a dramatic difference in the shape of the

ligand binding cavity, located within the LBD (Sections 3.1.1 and 3.2.1). The difference in residue distances between the LBD were small between ligand, DNA or *apo* KstR, while at the DBD ligand bound KstR displayed larger distances between residues while DNA-bound KstR induced smaller distances (Section 3.1.4 and 3.3.4). DNA binding induced shorter distances mostly within the DBD and between the LBD and DBD however not within the LBD itself. Furthermore, DNA binding reduced the flexibility of the DBD and domains near the ligand binding cavity and not other domains within the LBD (Section 3.1.3 and 3.3.3). Finally it was shown that 3OChA-CoA-bound KstR causes a few residues within the DBD to take on a non-helical composition while DNA binding induced subtle shortening in the $\alpha 4$ - $\alpha 5$ linker domain (Section 3.1.6 and 3.2.6). Together, this shows that despite the DBD being further away from the ligand binding cavity compared to domains at the LBD, ligand binding affected the conformation of the KstR DBD while effects of DNA binding were more pronounced at the DBD, not affecting the LBD.

The effects of 3OChA-CoA removal on KstR show how allosteric function operates within KstR (Section 3.2). While 3OChA-CoA removal has an explicit effect on the shape of KstR it is the RMSF at the DBD which shows the most pronounced changes (Sections 3.2.1 and 3.2.3). Additionally, the changes in the distances between residue pairs at the LBD displayed no difference upon ligand removal while the distances between residue pairs at the DBD became smaller on average (Section 3.2.4). Additionally, 3OChA-CoA removal from KstR causes differences in secondary structure which were most prominent at the DBD (Section 3.2.6). Together this demonstrates that when 3OChA-CoA is removed it is the conformation and dynamics of the DBD that are altered to be more *apo* KstR-like.

The removal of DNA does not impact the LBD as fast as the DBD (Section 3.3). The shape and size of KstR overall is impacted by DNA removal from KstR (Section 3.3.1). It is the flexibility about the DBD which is most affected by DNA removal from KstR with partial flexibility differences about the ligand binding cavity (Section 3.3.3). This is consistent with how DNA removal from KstR produces consistent lengthening of the average distance between residue pairs at the DBD while there is no distinctive trend for distances between residues at the LBD (Section 3.3.4). Additionally, the DNA binding induced closer contacts within the DBD and between the LBD and DBD however not within the LBD itself (Section 3.3.5). Overall, when DNA was removed the DBD itself was affected while the LBD was scarcely altered.

Together this demonstrates that DNA and ligand induce different effects on KstR. Section 3.1 demonstrates that ligand and DNA binding to KstR affects the conformation and dynamics of the LBD and DBD. Section 3.2 shows that ligand removal from KstR also affects the conformation and dynamics of the DBD with slight changes to the LBD. Section 3.3 shows that when DNA is removed from KstR it is exclusively the conformation and dynamics of the DBD which are altered.

4.7 Future work

There are a range of areas in which this work could be extended in the future.

Extend the length of DNA. Endogenously, KstR interacts with DNA which is seemingly endless compared to the size of KstR. While shortening and removal effects of DNA on KstR dynamics were tested here elongating effects were not. This presents an opportunity to investigate a system which is more representative of what KstR is exposed to endogenously. One alternative includes extending the length of DNA to the periodic borders of the system thereby creating a seemingly infinite strand of DNA.

Investigate the effect of other ligands. Through this whether KstR displays unique dynamics depending on the structure of ligand could be observed. Furthermore while ligand shortening and ligand removal affects have been observed what has yet to be seen is if the size of the ligand is extended beyond the endogenous ligand. Alternatively other cholesterol metabolites which are produced during the metabolisation of cholesterol could be tested to see if they bind and change the conformational dynamics of KstR.

Compute the binding free energies. Advanced sampling techniques such as Umbrella Sampling allow the free energy of binding of KstR to ligand or DNA to be computed. During this work, an attempt was made to compute the difference between the binding free energy of KstR to DNA with and without ligand bound. Unfortunately there were insufficient computational hours and time to complete the Umbrella Sampling.

Test a greater number of computational mutations. A number of residues thought to be important in ligand or DNA binding were highlighted in section 3.6.1. Additionally other features of the simulations of mutated KstR variants could be investigated. Other aspects of the conformation or dynamics of KstR may have changed upon mutation of specific residues. There was not enough time or resources to adequately observe these changes with the same level of detail for the other systems.

Ligand and DNA dynamics. The KstR dynamics have been analysed however the dynamics of ligand and DNA have been left unstudied. I think there is a substantial amount of compelling research available to investigate the dynamics of ligand and DNA dynamics in their own right and how KstR impacts their dynamics.

Cooperatively between KstR monomers. The relationship and comparison of dynamics between the KstR monomers was not examined. The movements of each monomer, and movements of ligand and DNA bound at each monomer are not identical to the each other. By investigating how each monomer responds to different perturbations would examine the cooperatively within KstR, an aspect which may contribute towards the proteins mechanism of action.

4.8 Conclusion

The conformational dynamics of KstR is dependent on whether KstR is bound to ligand or DNA. The removal of ligand or DNA from KstR affects the conformational dynamics of KstR. Changes in the length of DNA or ligand have insufficient effects on the conformational dynamics of KstR. Specific residues across KstR are critical in its association with DNA and ligand, and can be used as analytical markers of the binding relationships of KstR.

The structure and conformational dynamics of KstR is configured by its association with ligand and DNA. DNA contracts and restrains the overall shape of KstR, while ligand expands and allows it to be more mobile. This demonstrates that ligand or DNA binding induce the conformation and dynamics of KstR in different ways, indicating an induced fit mechanism. However, extreme values of *apo* KstR sample within the conformational behaviour of both ligand-bound and DNA-bound KstR. This indicates that if the DNA or ligand is present at the same time as KstR is in the right conformation, the appropriate binding conformation will be selected for, indicating a conformational selection binding mode. Ligand binding has a small impact on the LBD itself but transmits conformational changes to the DBD, while changes at the DBD do not impact the LBD to the same extent. This indicates that the KstR allosteric mechanism is triggered by ligand and operates in one direction. The architecture of ligand-bound KstR is not dependent on a specific composition of ligand at the binding cavity. Ligand binding to KstR causes the shape and behaviour of KstR to disfavour association with ligand, while when KstR binds to DNA the ligand binding cavity is induced to open, perhaps to more frequently accept ligand. Thus the KstR mechanism of action also favours aspects of induced fit.

Bibliography

[1] WHO, *Global Tuberculosis Report 2017*. 2017.

[2] K. H. Z. Bentrup and D. G. Russell, "Mycobacterial persistence: Adaptation to a changing environment," *Trends in Microbiology*, vol. 9, no. 12, pp. 597–605, 2001.

[3] S. L. Kendall, M. Withers, C. N. Soffair, N. J. Moreland, S. Gurcha, B. Sidders, R. Frita, A. Ten Bokum, G. S. Besra, J. S. Lott, and N. G. Stoker, "A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis," *Molecular Microbiology*, vol. 65, no. 3, pp. 684–699, 2007.

[4] N. A. T. Ho, S. S. Dawes, A. M. Crowe, I. Casabon, C. Gao, S. L. Kendall, E. N. Baker, L. D. Eltis, and J. S. Lott, "The structure of the transcriptional repressor kstr in complex with coa thioester cholesterol metabolites sheds light on the regulation of cholesterol catabolism in mycobacterium tuberculosis," *Journal of Biological Chemistry*, vol. 291, no. 14, pp. 7256–7266, 2016.

[5] M. Abraham, B. Hess, D. van der Spoel, and E. Lindahl, "GROMACS User Manual version 5.0.7," *Www.Gromacs.Org*, 2015.

[6] A. M. Talaat, R. Lyons, S. T. Howard, and S. A. Johnston, "The temporal expression profile of Mycobacterium tuberculosis infection in mice.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 13, pp. 4602–4607, 2004.

[7] H. Ouellet, J. B. Johnston, and P. R. O. de Montellano, "Cholesterol catabolism as a therapeutic target in Mycobacterium tuberculosis," *Trends in Microbiology*, vol. 19, no. 11, pp. 530–539, 2011.

[8] C. J. Cambier, S. Falkow, and L. Ramakrishnan, "Host evasion and exploitation schemes of Mycobacterium tuberculosis," *Cell*, vol. 159, no. 7, pp. 1497–1509, 2014.

[9] A. K. Pandey and C. M. Sassetti, "Mycobacterial persistence requires the utilization of host cholesterol.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 11, pp. 4376–80, 2008.

[10] T. C. Zahrt, "Molecular mechanisms regulating persistent Mycobacterium tuberculosis infection," *Microbes and Infection*, vol. 5, no. 2, pp. 159–167, 2003.

[11] Y. Liu, S. Tan, L. Huang, R. B. Abramovitch, K. H. Rohde, M. D. Zimmerman, C. Chen, V. Dartois, B. C. VanderVen, and D. G. Russell, "Immune activation of the host cell induces drug tolerance in Mycobacterium tuberculosis both in vitro and in vivo," *The Journal of Experimental Medicine*, p. jem.20151248, 2016.

[12] S. L. Kendall, S. C. G. Rison, F. Movahedzadeh, R. Frita, and N. G. Stoker, "What do microarrays really tell us about M. tuberculosis?," *Trends in Microbiology*, vol. 12, no. 12, pp. 537–544, 2004.

[13] A. Lenaerts, C. E. Barry, and V. Dartois, "Heterogeneity in tuberculosis pathology, microenvironments and therapeutic responses," *Immunological Reviews*, vol. 264, no. 1, pp. 288–307, 2015.

[14] P. Peyron, J. Vaubourgeix, Y. Poquet, F. Levillain, C. Botanch, F. Bardou, M. Daffé, J. F. Emile, B. Marchou, P. J. Cardona, C. De Chastellier, and F. Altare, "Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for M. tuberculosis persistence," *PLoS Pathogens*, vol. 4, no. 11, pp. 1–14, 2008.

[15] N. M. Parrish, J. D. Dick, and W. R. Bishai, "Mechanisms of latency in Mycobacterium tuberculosis," *Trends in Microbiology*, vol. 6, no. 3, pp. 107–112, 1998.

[16] S. Ehrt, "Reprogramming of the Macrophage Transcriptome in Response to Interferon-gamma and Mycobacterium tuberculosis: Signaling Roles of Nitric Oxide Synthase-2 and Phagocyte Oxidase," *Journal of Experimental Medicine*, vol. 194, no. 8, pp. 1123–1140, 2001.

[17] J. E. Griffin, A. K. Pandey, S. A. Gilmore, V. Mizrahi, J. D. McKinney, C. R. Bertozzi, and C. M. Sassetti, "Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations," *Chemistry and Biology*, vol. 19, no. 2, pp. 218–227, 2012.

[18] S. L. Kendall, P. Burgess, R. Balhana, M. Withers, A. Ten Bokum, J. S. Lott, C. Gao, I. Uhia-Castro, and N. G. Stoker, "Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2," *Microbiology*, vol. 156, no. 5, pp. 1362–1371, 2010.

[19] A. M. Crowe, P. J. Stogios, I. Casabon, E. Evdokimova, A. Savchenko, and L. D. Eltis, "Structural and functional characterization of a ketosteroid transcriptional regulator of Mycobacterium tuberculosis," *Journal of Biological Chemistry*, vol. 290, no. 2, pp. 872–882, 2015.

[20] W. Lee, B. C. VanderVen, R. J. Fahey, and D. G. Russell, "Intracellular Mycobacterium tuberculosis exploits host-derived fatty acids to limit metabolic stress," *Journal of Biological Chemistry*, vol. 288, no. 10, pp. 6788–6800, 2013.

[21] K. C. Yam, I. D. Angelo, R. Kalscheuer, H. Zhu, J.-x. Wang, L. H. Ly, P. J. Converse, W. R. J. Jr, N. Strynadka, and L. D. Eltis, "Studies of a Ring-Cleaving Dioxygenase Illuminate the Role of Cholesterol Metabolism in the Pathogenesis of Mycobacterium tuberculosis," vol. 5, no. 3, 2009.

[22] S. Sisay, B. Mengistu, W. Erku, and D. Woldeyohannes, "Ten years' experience of Directly Observed Treatment Short-course (DOTS) in Gambella Regional State, Ethiopia: An evaluation of tuberculosis control program," in *International Journal of Mycobacteriology*, vol. 5, pp. S117–S118, 2016.

[23] Y. Beyene, B. Geresu, and A. Mulu, "Mortality among tuberculosis patients under DOTS programme: A historical cohort study," *BMC Public Health*, vol. 16, no. 1, pp. 371–378, 2016.

[24] Y. Beyene, B. Geresu, and A. Mulu, "Mortality among tuberculosis patients under DOTS programme : a historical cohort study," *BMC Public Health*, pp. 4–9, 2016.

[25] R. Sivaraj, S. Umarani, S. Parasuraman, and P. Muralidhar, "Revised National Tuberculosis Control Program regimens with and without directly observed treatment , short- course : A comparative study of therapeutic cure rate and adverse reactions," vol. 5, no. 1, pp. 16–19, 2014.

[26] N. Willand, B. Dirié, X. Carette, P. Bifani, A. Singhal, M. Desroses, F. Leroux, E. Willery, V. Mathys, R. Déprez-Poulain, G. Delcroix, F. Frénois, M. Aumercier, C. Locht, V. Villeret, B. Déprez, and A. R. Baulard, "Synthetic EthR inhibitors boost antituberculous activity of ethionamide," *Nature Medicine*, vol. 15, no. 5, pp. 537–544, 2009.

[27] S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, a. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. a. Quail, M. a. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell, "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.," *Nature*, vol. 393, no. 6685, pp. 537–544, 1998.

[28] J. G. Rodríguez, A. C. Hernández, C. Helguera-Repetto, D. A. Ayala, R. Guadarrama-Medina, J. M. Anzóla, J. R. Bustos, M. M. Zambrano, J. González-y Merchand, M. J. García, and P. D. Portillo, "Global adaptation to a lipid environment triggers the dormancy- related phenotype of mycobacterium tuberculosis," *mBio*, vol. 5, no. 3, pp. 1–13, 2014.

[29] K. J. Aldred, R. J. Kerns, and N. Osheroff, "Mechanism of quinolone action and resistance," *Biochemistry*, vol. 53, no. 10, pp. 1565–1574, 2014.

[30] N. M. Nesbitt, X. Yang, P. Fontán, I. Kolesnikova, I. Smith, N. S. Sampson, and E. Dubnau, "A thiolase of Mycobacterium tuberculosis is required for virulence and production of androstenedione and androstadienedione from cholesterol," *Infection and Immunity*, vol. 78, no. 1, pp. 275–282, 2010.

[31] N. a. Lack, A. Kawamura, E. Fullam, N. Laurieri, S. Beard, A. J. Russell, D. Evangelopoulos, I. Westwood, and E. Sim, "Temperature stability of proteins essential for the intracellular survival of Mycobacterium tuberculosis.," *The Biochemical journal*, vol. 418, no. 2, pp. 369–78, 2009.

[32] A. M. Upton and J. D. McKinney, "Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis," *Microbiology*, vol. 153, no. 12, pp. 3973–3982, 2007.

[33] M. C. Anderton, S. Bhakta, G. S. Besra, P. Jeavons, L. D. Eltis, and E. Sim, "Characterization of the putative operon containing arylamine N-acetyltransferase (nat) in Mycobacterium bovis BCG," *Molecular Microbiology*, vol. 59, no. 1, pp. 181–192, 2006.

[34] R. J. C. Balhana, A. Singla, M. H. Sikder, M. Withers, and S. L. Kendall, "Global analyses of TetR family transcriptional regulators in mycobacteria indicates conservation across species and diversity in regulated functions," *BMC Genomics*, vol. 16, no. 1, p. 479, 2015.

[35] P. Orth, D. Schnappinger, W. Hillen, W. Saenger, and W. Hinrichs, "Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system.," *Nature structural biology*, vol. 7, no. 3, pp. 215–9, 2000.

[36] F. Haberl, H. Lanig, and T. Clark, "Induction of the tetracycline repressor: Characterization by molecular-dynamics simulations," *Proteins: Structure, Function and Bioinformatics*, vol. 77, no. 4, pp. 857–866, 2009.

[37] T. B. K. Le, M. A. Schumacher, D. M. Lawson, R. G. Brennan, and M. J. Buttner, "The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding," *Nucleic Acids Research*, vol. 39, no. 21, pp. 9433–9447, 2011.

[38] T. B. K. Le, C. E. M. Stevenson, H. P. Fiedler, A. Maxwell, D. M. Lawson, and M. J. Buttner, "Structures of the tetr-like simocyclinone efflux pump repressor, SimR, and the mechanism of ligand-mediated derepression," *Journal of Molecular Biology*, vol. 408, no. 1, pp. 40–56, 2011.

[39] L. Cuthbertson and J. R. Nodwell, "The TetR Family of Regulators," *Microbiology and Molecular Biology Reviews*, vol. 77, no. 3, pp. 440–475, 2013.

[40] B. C. VanderVen, R. J. Fahey, W. Lee, Y. Liu, R. B. Abramovitch, C. Memmott, A. M. Crowe, L. D. Eltis, E. Perola, D. D. Deininger,

T. Wang, C. P. Locher, and D. G. Russell, "Novel Inhibitors of Cholesterol Degradation in Mycobacterium tuberculosis Reveal How the Bacterium's Metabolism Is Constrained by the Intracellular Environment," *PLoS Pathogens*, vol. 11, no. 2, pp. 1–20, 2015.

[41] J. Jiang, X. Sun, W. Wu, L. Li, H. Wu, L. Zhang, G. Yu, and Y. Li, "Construction and application of a co-expression network in Mycobacterium tuberculosis," *Scientific Reports*, vol. 6, no. June, pp. 1–18, 2016.

[42] F. Frénois, J. Engohang-Ndong, C. Locht, A. R. Baulard, and V. Villeret, "Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis," *Molecular Cell*, vol. 16, no. 2, pp. 301–307, 2004.

[43] A. Aleksandrov and T. Simonson, "Binding of tetracyclines to elongation factor Tu, the Tet Repressor, and the ribosome: A molecular dynamics simulation study," *Biochemistry*, vol. 47, no. 51, pp. 13594–13603, 2008.

[44] A. Aleksandrov, L. Schuldt, W. Hinrichs, and T. Simonson, "Tetracycline-Tet repressor binding specificity: Insights from experiments and simulations," *Biophysical Journal*, vol. 97, no. 10, pp. 2829–2838, 2009.

[45] P. Csermely, R. Palotai, and R. Nussinov, "Induced fit, conformational selection and independent dynamic segments: An extended view of binding events," *Trends in Biochemical Sciences*, vol. 35, no. 10, pp. 539–546, 2010.

[46] A. D. Vogt and E. Di Cera, "Conformational selection or induced fit? A critical appraisal of the kinetic mechanism," *Biochemistry*, vol. 51, no. 30, pp. 5894–5902, 2012.

[47] A. D. Vogt, N. Pozzi, Z. Chen, and E. Di Cera, "Essential role of conformational selection in ligand binding," *Biophysical Chemistry*, vol. 186, pp. 13–21, 2014.

[48] R. L. Gomez, L. Jose, R. Ramachandran, S. Raghunandanan, B. Muralikrishnan, J. B. Johnson, K. C. Sivakumar, S. Mundayoor, and R. A. Kumar, "The multiple stress responsive transcriptional regulator Rv3334 of Mycobacterium tuberculosis is an autorepressor and a positive regulator of kstR," *FEBS Journal*, vol. 283, pp. 3056–3071, 2016.

[49] J. L. Ramos, M. Martínez-bueno, J. Antonio, W. Terán, K. Watanabe, M. T. Gallegos, R. Brennan, M. Mart, A. J. Molina-henares, and W. Tera, "The TetR Family of Transcriptional Repressors The TetR Family of Transcriptional Repressors," *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 2, pp. 326–356, 2005.

[50] M. Yang, R. Lu, K. E. Guja, M. F. Wipperman, J. R. St. Clair, A. C. Bonds, M. Garcia-Diaz, and N. S. Sampson, "Unraveling Cholesterol Catabolism in Mycobacterium tuberculosis: ChsE4-ChsE5 ??2??2 Acyl-CoA Dehydrogenase Initiates ??-Oxidation of 3-Oxo-cholest-4-en-26-oyl CoA," *ACS Infectious Diseases*, vol. 1, no. 2, pp. 110–125, 2016.

[51] J. K. Capyk, I. Casabon, R. Gruninger, N. C. Strynadka, and L. D. Eltis, "Activity of 3-ketosteroid 9??-hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in Mycobacterium tuberculosis," *Journal of Biological Chemistry*, vol. 286, no. 47, pp. 40717–40724, 2011.

[52] P. O. Nikiforov, S. Surade, M. Blaszczyk, V. Delorme, P. Brodin, A. R. Baulard, T. L. Blundell, and C. Abell, "A fragment merging approach towards the development of small molecule inhibitors of Mycobacterium tuberculosis EthR for use as ethionamide boosters," *Org. Biomol. Chem.*, vol. 14, no. 7, pp. 2318–2326, 2016.

[53] S. E. Reichheld, Z. Yu, and A. R. Davidson, "The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor," *Proceedings of the National Academy of Sciences*, vol. 106, no. 52, pp. 22263–22268, 2009.

[54] I. Uhía, B. Galán, F. J. Medrano, and J. L. García, "Characterization of the KstR-dependent promoter of the gene for the first step of the cholesterol degradative pathway in Mycobacterium smegmatis," *Microbiology*, vol. 157, no. 9, pp. 2670–2680, 2011.

[55] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," vol. 459, no. 1959, 1959.

[56] McCammon, J.A., Gelin B.R and K. M., "Dynamics of folded proteins," 1977.

[57] L.-W. Yang, S. Matysiak, S.-T. D. Hsu, G. Mustata Wilson, and Y. Joti, "Functional dynamics of proteins.," *Computational and mathematical methods in medicine*, vol. 2012, p. 242903, 2012.

[58] M. A. Schumacher, M. C. Miller, and S. Grkovic, "Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR," *EMBO Journal*, vol. 21, no. 5, pp. 1210–1218, 2002.

[59] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell, "Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges," *Journal of Chemical Information and Modeling*, vol. 52, no. 12, pp. 3155–3168, 2012.

[60] D. G. Truhlar, "Valence Bond Theory for Chemical Dynamics," *Journal of computational chemistry*, vol. 28, no. 1, pp. 73–86, 2009.

[61] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer, and P. Varnai, "Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps," *Biophysical Journal*, vol. 89, no. 6, pp. 3721–3740, 2005.

[62] N. Foloppe and A. D. MacKerell, Jr., "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data," *Journal of Computational Chemistry*, vol. 21, no. 2, pp. 86–104, 2000.

[63] M. Dorn, M. B. E Silva, L. S. Buriol, and L. C. Lamb, "Three-dimensional protein structure prediction: Methods and computational strategies," *Computational Biology and Chemistry*, vol. 53, no. PB, pp. 251–276, 2014.

[64] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[65] V. Tsui, A. David, and D. a. Case, "Molecular Dynamics Simulations of Nucleic Acids with a Generalized Born Solvation Model," *Journal of the American Chemical Society*, vol. 122, no. 11, pp. 2489–2498, 2000.

[66] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[67] M. Boggio-Pasqua, C. F. Burmeister, M. A. Robb, and G. Groenhof, "Photochemical reactions in biological systems: probing the effect of the environment by means of hybrid quantum chemistry/molecular mechanics simulations.," *Physical Chemistry Chemical Physics*, vol. 14, no. 22, pp. 7912–28, 2012.

[68] H. a. Carlson and J. a. McCammon, "Accommodating protein flexibility in computational drug design.," *Molecular pharmacology*, vol. 57, no. 2, pp. 213–218, 2000.

[69] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, "Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality," *Journal of Chemical Physics*, vol. 119, no. 6, pp. 3559–3566, 2003.

[70] J. A. Lemkul, W. J. Allen, and D. R. Bevan, "Practical considerations for building GROMOS-compatible small-molecule topologies," *Journal of Chemical Information and Modeling*, vol. 50, no. 12, pp. 2221–2235, 2010.

[71] H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[72] E. Mutt and R. Sowdhamini, "Molecular dynamics simulations and structural analysis to decipher functional impact of a twenty residue insert in the ternary complex of Mus musculus TdT isoform," *PLoS ONE*, vol. 11, no. 6, pp. 1–24, 2016.

[73] A. Pang, Y. Arinaminpathy, M. S. Sansom, and P. C. Biggin, "Interdomain dynamics and ligand binding: Molecular dynamics simulations of glutamine binding protein," *FEBS Letters*, vol. 550, no. 1-3, pp. 168–174, 2003.

[74] L. Schrödinger, "The PyMOL Molecular Graphics System, Version˜1.5." nov 2015.

[75] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera - A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[76] L. T. Sali, A. and Blundell, "Comparative Protein Modelling by Satisfaction of Spatial Restraints," *Journal of Molecular Biology*, vol. 234, pp. 779 – 815, 1993.

[77] W. Yu, X. He, K. Vanommeslaeghe, and A. D. MacKerell, "Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations," *Journal of Computational Chemistry*, vol. 33, no. 31, pp. 2451–2468, 2012.

[78] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, no. 2, pp. 187–217, 1983.

[79] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, "LINCS: A Linear Constraint Solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[80] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[81] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems," *The Journal of Chemical Physics*, vol. 98, no. 12, pp. 10089–10092, 1993.

[82] F. Eisenhaber, P. Lijnzaad, P. Argos, and C. Sander, "The Double Cubic Lattice Method : Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies," vol. 16, no. 3, pp. 273–284, 1994.

[83] S. Sisay, B. Mengistu, W. Erku, and D. Woldeyohannes, "Ten years' experience of Directly Observed Treatment Short-course (DOTS) in Gambella Regional State, Ethiopia: An evaluation of tuberculosis control program," in *International Journal of Mycobacteriology*, vol. 5, pp. S117–S118, 2016.

[84] S. Pronk, S. P??ll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: A high-throughput and highly parallel

open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.

[85] W. Kabsch and C. Sander, "Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983.

[86] T. Paramo, A. East, D. Garzón, M. B. Ulmschneider, and P. J. Bond, "Efficient characterization of protein cavities within molecular simulation trajectories: Trj-cavity," *Journal of Chemical Theory and Computation*, vol. 10, no. 5, pp. 2151–2164, 2014.

[87] A. D. MacKerell and L. Nilsson, "Molecular dynamics simulations of nucleic acid-protein complexes," *Current Opinion in Structural Biology*, vol. 18, no. 2, pp. 194–199, 2008.

[88] H. Lanig, O. G. Othersen, F. R. Beierlein, U. Seidel, and T. Clark, "Molecular Dynamics Simulations of the Tetracycline-repressor Protein: The Mechanism of Induction," *Journal of Molecular Biology*, vol. 359, no. 4, pp. 1125–1136, 2006.

[89] A. van Giessen, B. Macdonald, and P. Bui, "Protein-Protein Interactions and Secondary Structure affect Helix Stability in Crowded Environments," *Biophysical Journal*, vol. 110, no. 3, p. 213a, 2016.

[90] J. L. Ramos, M. Martínez-bueno, J. Antonio, W. Terán, K. Watanabe, M. T. Gallegos, R. Brennan, M. Mart, A. J. Molina-henares, and W. Tera, "The TetR Family of Transcriptional Repressors The TetR Family of Transcriptional Repressors," *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 2, pp. 326–356, 2005.

[91] C.-J. Tsai, S. Kumar, B. Ma, and R. Nussinov, "Folding funnels, binding funnels, and protein function," *Protein Science*, vol. 8, no. 6, pp. 1181–1190, 1999.

[92] K.-i. Okazaki and S. Takada, "Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms," *Proceedings of the National Academy of Sciences*, vol. 105, no. 32, pp. 11182–11187, 2008.