Molecular Flexibility of DNA as a Major Determinant of RAD51 Recruitment



Federico Paoletti

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

Acknowledgements

Completing my DPhil has been a challenging and rewarding experience. I would like to thank the many people and institutions that have helped me along the way.

I would like to thank the Doctoral Training Centre for introducing me to the world of mathematical biology and computational modelling. I also thank the Engineering and Physical Sciences Research Council (EPSRC) and University College for providing funding. My humble gratitude goes to my supervisors, Professor Fumiko Esashi and Dr Omer Dushek, for their immense support and guidance throughout the project. Specifically, I am grateful to Professor Esashi for her guidance in protein purification and ability to swiftly organise project collaborations, and Dr Dushek for guidance in surface plasmon resonance and mathematical modelling. I thank Dr Afaf El-Sagheer and Professor Tom Brown for providing essential DNA reagents, Professor Jun Allard for advice on SPR data interpretation, Dr Robert Rambo for collecting the SAXS data, and Dr David Staunton for collecting the SEC-MALS data. My thanks also goes to all the members of the Esashi lab for their encouragement and constructive feedback. Special thanks goes to Nicola Trendel of the Dushek lab for sharing essential ABC-SMC template scripts, and to Francesco Fiorentino for assistance with PyMOL.

Finally, I would like to thank several people that supported my scientific endeavours, in good times and bad times. My immense recognition goes to Melis Anatürk, Erika Calabrese, Oliver Ditthardt, Lev Tankelevitch, Benjamin Dodsworth, Anna Clements, Pantelis Antonoudiou, and Lorenzo Palmieri.

Last but not least, none of this would have been possible without the inexorable love and support of my parents, Stefano Paoletti and Elena Bonita Llorente Ruiz.

Contributions

The SAXS data (Fig. 4.2) and SEC-MALS data (Fig. 5.2B-C) were collected by Dr Robert Rambo and Dr David Staunton, respectively. All other experiments and mathematical modelling were carried out by Federico Paoletti. All Figures in chapters 3, 4 and 5 were adapted from the preprint article:

Paoletti et al., ”Molecular flexibility of DNA as a major determinant of RAD51 recruitment.” *BioRxiv* doi: https://doi.org/10.1101/392795, 2018.

Abstract

The timely activation of homologous recombination (HR) is essential for the maintenance of genome stability, in which the RAD51 recombinase plays a central role. *In vivo*, RAD51 polymerises on single-stranded DNA (ssDNA) at double-strand breaks (DSBs) to catalyse HR-dependent DSB repair. Biochemically, human RAD51 polymerises faster on ssDNA compared to double-stranded DNA (dsDNA), raising a key conceptual question: how does it discriminate between them? Throughout this thesis, I tackled this problem by systematically assessing RAD51 binding kinetics on ssDNA and dsDNA differing in length and flexibility using surface plasmon resonance. By fitting detailed polymerisation models informed by the experimental datasets, I suggest that RAD51 is a mechano-sensor that polymerises faster on flexible ssDNA compared to rigid ssDNA or dsDNA. Furthermore, I suggest that the energy released during RAD51 polymerisation is sufficient to overcome any entropic penalty to binding flexible ssDNA. This model presents a new general framework suggesting that the flexibility of DNA, which may increase locally as a result of DNA damage, plays an important role in recruiting repair factors that multimerise at sites of DNA damage.

Contents

List of Figures **v**

List of Tables **viii**

List of Abbreviations **ix**

1 Introduction **1**

1.1 The Repair of DNA Double-Strand Breaks by Homologous Recombination 3

1.1.1 End Resection 3

1.1.2 RAD51 Recruitment 5

1.1.3 Strand Invasion and Homology Search 6

1.1.4 DNA Synthesis and Ligation 7

1.2 The RAD51 Nucleoprotein Filament 10

1.2.1 The Roles of Nucleotide and Cation Binding 10

1.2.2 Active Versus Inactive Filaments 11

1.2.3 RAD51-DNA Interactions 16

1.2.4 RAD51 Protomer-Protomer Interactions 16

1.3 RAD51 Polymerisation 19

1.3.1 Static Analyses 19

1.3.2 Kinetic Analyses 22

1.4 The Impact of DNA Flexibility on Protein-DNA Interactions 25

1.4.1 Higher DNA Flexibility - Faster Recruitment of Repair Factors? 28

1.4.2 Overcoming the Entropic Penalty to Binding DNA 29
1.5 Surface Plasmon Resonance 31
1.5.1 The SPR Principle 31
1.5.2 A Typical Experiment 35
1.5.3 Potential Pitfalls 39
1.6 Fitting Unconventional SPR Data 40
1.7 Summary of Thesis 41

2 Materials and Methods **43**
2.1 RAD51 Molecular Cloning 43
2.2 RAD51 Protein Purification 43
2.3 Multi-Angle Light Scattering 45
2.4 DNA Oligo and Duplex Synthesis 45
2.5 Small-Angle X-Ray Scattering 48
2.6 Surface Plasmon Resonance 48
2.7 SPR Data Processing 51
2.8 ABC-SMC 52
2.9 Dimer Models 55
2.9.1 Assumptions 56
2.9.2 Description of ODEs 57
2.9.3 Model Selection 60
2.9.4 Interpretation of the ssDNA dT-50 $\mathbf{k_{el.f}}$ Rate 61
2.10 n-mer Models 62
2.10.1 Assumptions 64
2.10.2 Description of ODEs - Equilibrium Sub-Model 65
2.10.3 Description of ODEs - Kinetic Sub-Models 66
2.10.4 Model Selection 69

3 Mechanistic Models of RAD51 Polymerisation 73

3.1 Introduction 73

3.2 RAD51 Preferentially Associates with ssDNA 75

3.3 RAD51 exhibits stronger binding to dsDNA when DNA is short, but stronger binding to ssDNA when DNA is long 76

3.4 Mechanistic models of RAD51 polymerisation predict faster adsorption and / or elongation to ssDNA 79

3.4.1 Development of Dimer Models 79

3.4.2 Fitting the Dimer Models 82

3.4.3 Development of n-mers Models 89

3.4.4 Fitting the n-mers Models 91

3.5 Discussion 99

4 RAD51 Polymerises Faster on Flexible DNA 102

4.1 Introduction 102

4.2 The Bend-To-Capture Mechanism 103

4.3 Persistence Length Measurements of DNA molecules using SAXS . 106

4.4 RAD51 Adsorbs and/or Elongates Faster on Flexible DNA 108

4.5 Discussion 115

5 RAD51 Overcomes the Entropic Penalty to Binding Flexible DNA 117

5.1 Introduction 117

5.2 The Entropic Compensation Mechanism 118

5.3 The RAD51 Protomer-Protomer Interaction Compensates for the Entropic Penalty of Flexible DNA 121

5.4 Discussion 136

6 Conclusion 140

6.1 Summary of Thesis 140

6.2 Significance of Results 144

6.2.1 Alternative Mechanism to BTC 144
6.2.2 Experimental Caveats 145
6.3 Comparisons with Other Polymerisation Models 146
6.3.1 The Mcghee & von Hippel Model 146
6.3.2 The Schwarz & Watanabe Model 152
6.4 Future Work 156
6.4.1 Validation of the DUET Model 156
6.4.2 Single Molecule Approaches 158

Appendices **159**

A BioNetGen Code **160**
A.1 ssDNA Dimer Model 160
A.2 dsDNA Dimer Model 163
A.3 n-mer Equilibrium Sub-Model 166
A.4 n-mer General ssDNA and dsDNA Kinetic Sub-Models 170

B Additional Figures **178**

Bibliography **186**

List of Figures

1.1 Homologous recombination and alternative DSB repair pathways . . 8
1.2 HR and the function of RAD51 9
1.3 Active versus inactive filaments 13
1.4 The sequence homology of RAD51 orthologs and the structure of RAD51-DNA filaments 18
1.5 The impact of monovalent cations on pre-synaptic filament formation 21
1.6 RAD51 filament dissassembly at the single molecule level 24
1.7 Schematics of the flexibilities of ssDNA and dsDNA with and without RAD51 27
1.8 The physical basis of SPR 34
1.9 A typical SPR experiment 38
1.10 1:1 interactions versus polymer-polymer interactions 41

2.1 Chemical diagram of the dN-11p double-hexaethylene-glycol-linked oligo 47
2.2 RAD51 polymers can potentially elongate off DNA 50
2.3 Sum-of-squared residuals of the dimer kinetic sub-models 61
2.4 Sum-of-squared residuals of the n-mer kinetic sub-models 71

3.1 RAD51 polymerises faster on ssDNA 75
3.2 WT RAD51 binding kinetics to varying length of ssDNA and dsDNA, with corresponding dimer model fits 78
3.3 Kinetic representations of the dimer models 81

3.4 ABC-SMC probability densities for the ssDNA dimer model fit parameters . . . 83
3.5 ABC-SMC probability densities for the dsDNA dimer model fit parameters . . . 85
3.6 Predicted n-mer formation on DNA oligos using dimer models . . . 88
3.7 WT RAD51 binding kinetics to varying length of ssDNA and dsDNA, with corresponding n-mer model fits . . . 92
3.8 Kinetic representations of the ssDNA and dsDNA n-mer models . . . 93
3.9 ABC-SMC probability densities for the ssDNA and dsDNA n-mer model fit parameters using WT RAD51 . . . 96
3.10 WT RAD51 polymerisation in solution . . . 97

4.1 The Bend-To-Capture (BTC) mechanism . . . 105
4.2 Small-Angle X-Ray Scattering (SAXS) measurements of DNA oligos 107
4.3 WT RAD51 binding kinetics to DNA with varying flexibility and corresponding dimer model fits . . . 109
4.4 Supporting evidence for the Bend-To-Capture (BTC) mechanism using the ssDNA dimer model . . . 110
4.5 WT RAD51 binding kinetics to DNA with varying flexibility and corresponding n-mer model fits . . . 111
4.6 Supporting evidence for the Bend-To-Capture (BTC) mechanism using the ssDNA n-mer model . . . 113

5.1 The Entropic Penalty Compensation (EPC) mechanism . . . 120
5.2 F86E RAD51 is monomeric in solution. . . . 122
5.3 F86E RAD51 binding kinetics to ssDNA and dsDNA. . . . 124
5.4 Simplified dimer and n-mer models used for F86E RAD51 SPR data fitting . . . 126
5.5 F86E RAD51 polymerisation in solution . . . 128

5.6 F86E RAD51 binding kinetics to DNA with varying flexibility with associated dimer model fits 130
5.7 ABC-SMC probability densities for the simplified n-mer model fit parameters for one repeat of dT-50, dN-50, dA-50 and dN-50p datasets using F86E RAD51 132
5.8 F86E RAD51 binding kinetics to DNA with varying flexibility with associated n-mer model fits 133

6.1 The DUET model 143
6.2 The Mcghee & von Hippel model 149
6.3 The Schwarz & Watanabe model 155

B.1 WT RAD51 binding kinetics on dN-5, dN-8, dN-11, dN-14 and dN-17 at low RAD51 concentrations 179
B.2 WT RAD51 binding kinetics on dN-5p, dN-8p, and dN-11p at low RAD51 concentrations 180
B.3 ssDNA dN-X data fits for the ssDNA dimer models A - B, and dsDNA dN-Xp data fits for the dsDNA dimer models A and C 181
B.4 ssDNA dN-X data fits for the ssDNA kinetic sub-models A - D 182
B.5 ssDNA dN-X data fits for the ssDNA kinetic sub-models E, G, H, I . 183
B.6 dsDNA dN-Xp data fits for the dsDNA kinetic sub-models A, B, C, E 184
B.7 dsDNA dN-Xp data fits for the dsDNA kinetic sub-models E - H . . . 185

List of Tables

1.1 Helical parameters of recombinase filaments 14

2.1 DNA molecules used throughout the study 46

2.2 DNA molecule names and lengths, with corresponding maximum RAD51 polymer lengths (b) and dimer model ODEs used for SPR data fitting 56

2.3 DNA molecule names and lengths, with corresponding maximum RAD51 polymer lengths (b) and n-mer model ODEs used for SPR data fitting 63

2.4 The set of ssDNA and dsDNA kinetic sub-models derived from the general ssDNA kinetic sub-model and the dsDNA kinetic sub-model 70

3.1 Fit parameter values for the ssDNA and dsDNA dimer models 86

3.2 Fit parameter values for the ssDNA and dsDNA n-mer models . . . 97

4.1 Persistence lengths (L_p) of the ssDNA dN-50, dA-50, dT-50 and ds-DNA dN-50p molecules 106

List of Abbreviations

L_k	Kuhn length
L_p	persistence length
$\mathbf{K_D}$	RAD51 protomer-protomer dissociation constant
$\mathbf{k_{ad.f}}$	adsorption forward rate constant
$\mathbf{k_{ad.r}}$	adsorption reverse rate constant
$\mathbf{k_{el.f}}$	elongation forward rate constant
$\mathbf{k_{el.r}}$	elongation reverse rate constant
$\mathbf{k_s}$	stabilisation rate constant
[RAD51]	RAD51 SPR injection concentration
ABC-SMC	approximate bayesian computation coupled to sequential monte carlo
ATM	ataxia-telangiectasia mutated
ATP	adenosine triphosphate
ATPD	ATPase domain
BTC	Bend-To-Capture
C-NHEJ	canonical non-homologous end joining
CAP	catabolite gene activator
CDK	cyclin-dependent kinase
Cryo-EM	cryo-electron microscopy
DDR	DNA damage response
DSBs	double-strand breaks

dsDNA	double-stranded DNA
DUET	**D**na molec**U**lar fl**E**xibili**T**y
EMSA	electrophoretic mobility shift assay
EPC	Entropic Penalty Compensation
GCN4	general control protein 4
HMGB	eukaryotic high-mobility group B
HR	homologous recombination
MMEJ	microhomology-mediated end joining
MRN	MRE11-RAD50-NBS1
NTD	N-terminal domain
OB	oligonucleotide/oligosaccharide-binding
ODE	ordinary differential equation
RPA	replication protein A
SAXS	small-angle x-ray scattering
SEC-MALS	size-exclusion chromatography with multi-angle light scattering
SPR	surface plasmon resonance
SSA	single-strand annealing
ssDNA	single-stranded DNA
STR	Stack-To-Reject
TBP	TATA-binding protein
WT	wild-type

Chapter 1

Introduction

Each of our $\approx 10^{13}$ cells suffers tens of thousands of DNA lesions each day [1]. Such lesions arise due to the action of both physiological processes and environmental agents, can block genome replication and transcription, and potentially lead to genome-wide aberrations and tumourigenesis [1]. The DNA Damage Response (DDR) is a highly conserved set of DNA repair mechanisms that both prokaryotic and eukaryotic cells have evolved to repair DNA lesions. Defects in the DDR are often associated with human disease.

DNA double-strand breaks (DSBs) are highly toxic lesions that can lead to chromosomal rearrangements, genomic instability, cell death or tumourigenesis [2]. In mammalian cells, homologous recombination (HR) is an error-free DDR pathway that can accurately repair DSBs by using the sister chromatid as a template for DNA synthesis, and therefore mainly occurs during the S and G_2 cell cycle phases [2].

RAD51 is the key ATP-dependent recombinase that catalyses HR in mammalian cells. *In vivo*, RAD51 polymerises on single-stranded DNA (ssDNA) generated at DSBs to guide DNA strand-invasion onto the sister chromatid, and its recruitment is dependent on the activity of several protein co-factors (e.g. BRCA2, PALB2, RAD52) [3, 4, 5, 6, 7, 8, 9, 10]. *In vitro* upon inhibition of ATP hydrolysis, RAD51

polymerises faster on ssDNA compared to double-stranded DNA (dsDNA) [11, 12], and depolymerises slowly at similar speeds from both ssDNA and dsDNA [11]. This observation raises a key conceptual question: how does RAD51 discriminate between ssDNA and dsDNA?

Throughout this thesis, I provide evidence to suggest that RAD51 preferentially polymerises on ssDNA due to the higher flexibility of ssDNA compared to dsDNA. In particular, in chapter 3 I measure the RAD51 binding kinetics to ssDNA and dsDNA oligos of different lengths via surface plasmon resonance (SPR), develop RAD51 polymerisation models to fit these data and suggest that RAD51 nuclei adsorb and/or elongate faster on ssDNA. In chapter 4, I measure the RAD51 binding kinetics to ssDNA oligos of different flexibilities via SPR, fit the ssDNA model from chapter 3 to these data, and conclude that RAD51 polymerises faster on flexible DNA compared to rigid DNA. In chapter 5, I investigate how RAD51 can stably bind flexible DNA, and measure via SPR the RAD51 binding kinetics to the same oligos as in chapter 4 but using a RAD51 mutant (F86E) with a reduced protomer-protomer affinity. I then fit the ssDNA and dsDNA models from chapter 3 to these data, and conclude that the enthalpic contribution of WT RAD51 polymerisation is sufficient to offset any entropic penalty associated with binding flexible DNA. Finally, in chapter 6 I summarise and discuss the significance of the findings, compare the models developed in chapter 3 to other models in the literature, and propose ideas to further understand how DNA flexibility impacts RAD51 polymerisation.

In this chapter, I will provide an overview of DSB repair and focus on the mechanism of HR. I will then describe the *in vivo* function of RAD51, as well as the role of protein-cofactors in the regulation of RAD51 activity. I will then present the structural features of RAD51-DNA filaments, the biochemical features of RAD51 polymerisation, and introduce the role of DNA flexibility in protein-DNA interactions. I

will then introduce surface plasmon resonance (SPR) and describe the motivation for a combined experimental and mathematical modelling approach to characterise RAD51 polymerisation kinetics on DNA. Finally, I will provide a chapter-by-chapter summary of the thesis.

1.1 The Repair of DNA Double-Strand Breaks by Homologous Recombination

HR is an error-free DSB repair pathway that uses a complementary DNA sequence from a sister chromatid to repair a DSB, and therefore occurs during the S and G_2 phases in somatic cells [13]. While HR is also proposed to promote telomere protection and elongation [14], this thesis focuses on the DDR function of HR. HR competes for DSB substrates with at least three other DSB repair pathways: canonical non-homologous end joining (C-NHEJ), single-strand annealing (SSA) and microhomology-mediated end joining (MMEJ) (Fig. 1.1).

1.1.1 End Resection

End resection is the process by which 3' ssDNA overhangs are generated at sites of DSBs (Fig. 1.2A). The MRE11-RAD50-NBS1 (MRN) complex together with the CtIP protein co-factor binds to DSBs and initiates end resection [15], and cyclin-dependent kinases (CDKs) ensure end resection primarily occurs during the S and G_2 cell cycle phases. In particular, the MRE11 exonuclease of the MRN complex in complex with CtIP catalyses the initial DNA clipping step via a two-step process. First, phosphorylation of CtIP at Thr847 by cyclin-dependent kinases 1/2 [16, 17] promotes an MRE11-dependent endonucleolytic incision away from the DNA break, preferentially on 5' terminated DNA strands [18]. The endonuclease activity of MRE11 is also dependent on ATP hydrolysis by RAD50 [18]. CDK1/2 - dependent phosphorylation of NBS1 [19, 20] during the S and G_2 phases is also

associated with increased end resection and has been proposed to promote the endonucleolytic activity of MRE11. MRE11 then catalyses 3' - 5' end resection via its exonuclease activity from this incision to complete the DNA clipping step [18]. The exonucleases DNA2 (in complex with the helicases WRN or BLM [21]) or EXO1 then catalyse extensive 5' - 3' end resection to generate 3' ssDNA overhangs [21, 22, 15]. In the case of EXO1, up to 3500 bps are resected in human cells [15]. The checkpoint kinase ataxia-telangiectasia mutated (ATM) then prevents excessive end resection by phosphorylating MRE11 and EXO1 [23].

DNA end resection primes DSBs for HR-dependent repair and occurs during the S and G_2 phases. During G_1 phase 53BP1, in complex with several protein co-factors including RIF1 [24] and shieldin [25, 26, 27, 28], accumulates at DSBs and is known to inhibit the CtIP-dependent DNA clipping step of end resection and promote C-NHEJ [29]. 53BP1 specifically binds to histone H4 dimethylated on Lys20 (H4K20me2) via its Tudor domain [30], and to histone H2A ubiquitinated by RNF168 on Lys15 (H2AK15ub) via its ubiquitination-dependent recruitment motif on chromatin flanking DSBs [31]. RIF1 was shown to be an essential protein co-factor for 53BP1 function, and its recruitment is dependent on ATM-dependent phosphorylation of 53BP1 [32]. Additionally, the shieldin complex has been proposed to bind to ssDNA at DSBs via its oligosaccharide-binding (OB) folds and limit nuclease activity required for end resection [28, 27]. During the S and G_2 phases, BRCA1 forms a complex with MRN-CtIP at DSBs that prevents accumulation of 53BP1 and promotes HR [33]. In particular, the BRCA1-CtIP interaction is dependent on the CDK-dependent phosphorylation of CtIP on Ser327 [34]. Indeed, the 53BP1-RIF1 complex and the BRCA1-CtIP complex form a cell-cycle dependent regulatory circuit that dictates DSB pathway choice between HR and C-NHEJ, whereby during S and G_2 the BRCA1-CtIP complex evicts RIF1 from the chromatin flanking the DSB and promotes HR-dependent repair [32]. This was also supported by the observa-

tion that suppression of 53BP1 or RIF1 in BRCA1-deficient cells restores functional end resection [29, 32]. Interestingly, persistent binding of the Ku70/80 heterodimer at DSBs during the G_1 phase also limits end resection and promotes DSB repair by NHEJ [35, 36], although single molecule analyses have suggested MRN on its own is capable of removing Ku70/80 via an MRE11-dependent endonucleolytic cleavage reaction at DSBs [37].

Following end resection, the heterotrimeric ssDNA-binding protein RPA rapidly binds to the ssDNA overhangs to prevent secondary structure formation and ssDNA degradation [38, 39].

1.1.2 RAD51 Recruitment

A variety of protein co-factors, including BRCA2, PALB2, RAD52 and RAD51 paralogs, facilitate the removal of RPA and the polymerisation of RAD51 onto the exposed ssDNA overhangs (Fig. 1.2B). In particular, it has been suggested that BRCA2 in complex with the DSS1 peptide mediates removal of RPA [40, 7]. BRCA2 has been proposed to facilitate the recruitment of RAD51 to ssDNA [41] by stabilising RAD51-ssDNA polymers via its C-terminal TR2 domain [3, 4], and nucleating and stabilising RAD51-ssDNA polymers via its BRC motifs by inhibiting RAD51-dependent ATP hydrolysis [42, 5, 6]. RAD51 has been shown to polymerise on DNA in the 3' - 5' direction [43], and BRCA2 has been proposed to enhance RAD51 polymer elongation in the 3' - 5' direction from individual nucleation sites on ssDNA [44]. Additionally, it was shown that RAD51 filaments on ssDNA at stalled replication forks protect the underlying DNA from MRE11-dependent nucleolytic degradation [45, 46], and BRCA2 was shown to enhance fork protection by stabilising RAD51 filaments via its TR2 domain [47, 48]. PALB2, a key BRCA2 binding partner, and RAD52 have also been suggested to promote RPA removal and RAD51

filament formation [8, 49, 10].

RAD51 paralogs are proteins that structurally resemble the RAD51 ATPase domain and are conserved throughout eukaryotes. In human cells, RAD51 paralogs form three functional complexes: the BCDX2 complex consisting of RAD51B, RAD51C, RAD51D and XRCC2 [50], the CX3 complex consisting of RAD51C and XRCC3 [51], and the Shu complex consisting of SWS1 and SWSAP1 [52]. RAD51 paralog complexes are predicted to promote RAD51 filament formation on ssDNA overhangs, although their mechanisms of action are still debated. Interestingly, the *Caenorhabtidis elegans* Shu complex composed of the RAD51 paralogs SWS1, RFS-1 and RIP-1 has been shown to stabilise RAD51 presynaptic filaments by binding to the 5' filament end and decreasing RAD51 dissociation from ssDNA [53, 54, 55].

1.1.3 Strand Invasion and Homology Search

RAD51 polymerises on the exposed ssDNA overhangs to form a nucleoprotein filament, also known as the presynaptic filament [56] (Fig. 1.2B). Importantly, the ATP-dependent helicases RECQ5 and FBH1 catalyse presynaptic filament disassembly and therefore suppress HR [57, 58]. During strand invasion, RAD51 is then thought to form a three-strand intermediate that includes the invading strand and the two complementary sister chromatid DNA strands, also known as the synaptic complex [59, 60] (Fig. 1.2B). The BRCA1-BARD1 complex promotes synaptic complex formation and is essential for RAD51-mediated DSB repair [61]. RAD51 then promotes annealing of the invading strand with the complementary sister chromatid strand by excluding the opposite sister chromatid strand to form the postsynaptic complex. This process is mediated by the formation of a ssDNA D-loop (Fig. 1.2A-B). The ATP-dependent helicase RTEL1 specifically disrupts RAD51-mediated D-loops and downregulates HR [62]. RAD54 interacts with RAD51 and is thought

to enhance postsynaptic complex formation [63], while NEK1-dependent RAD54 phosphorylation was also shown to promote RAD51 dissassembly [64].

1.1.4 DNA Synthesis and Ligation

It has been suggested that 1st end DNA synthesis is primarily carried out by the B Family DNA polymerase δ [65] (Fig. 1.2A). DNA polymerases α [66, 67], ϵ [66], ζ [68] of the B Family, DNA polymerases γ [67], θ [69, 70] and ν [71] of the A Family, and DNA polymerase η [72] of the Y Family have also been implicated in HR-mediated DNA synthesis, although their specific roles are still disputed. Interestingly, DNA polymerase θ has also been proposed to promote annealing of complementary 3' ssDNA overhangs prior to RAD51 recruitment and stimulate MMEJ [73]. RPA and the loading clamp PCNA have been proposed to prevent DNA polymerase δ stalling due to DNA topological constraints and maintain high DNA polymerase δ processivity, respectively [65, 68]. However, DNA polymerase η has been shown to catalyse 1st end DNA synthesis more effectively compared to DNA polymerase δ *in vitro* [74]. This is then followed by re-annealing of the invading strand to its original complementary sequence and re-annealing of the two sister chromatid strands. Faithful DSB repair is then completed via 2nd end DNA synthesis, which also appears to be catalysed by DNA polymerase δ [65] (Fig. 1.2A). Ligation of DNA ends has been associated with DNA ligase I, although other ligases may also be involved [75].

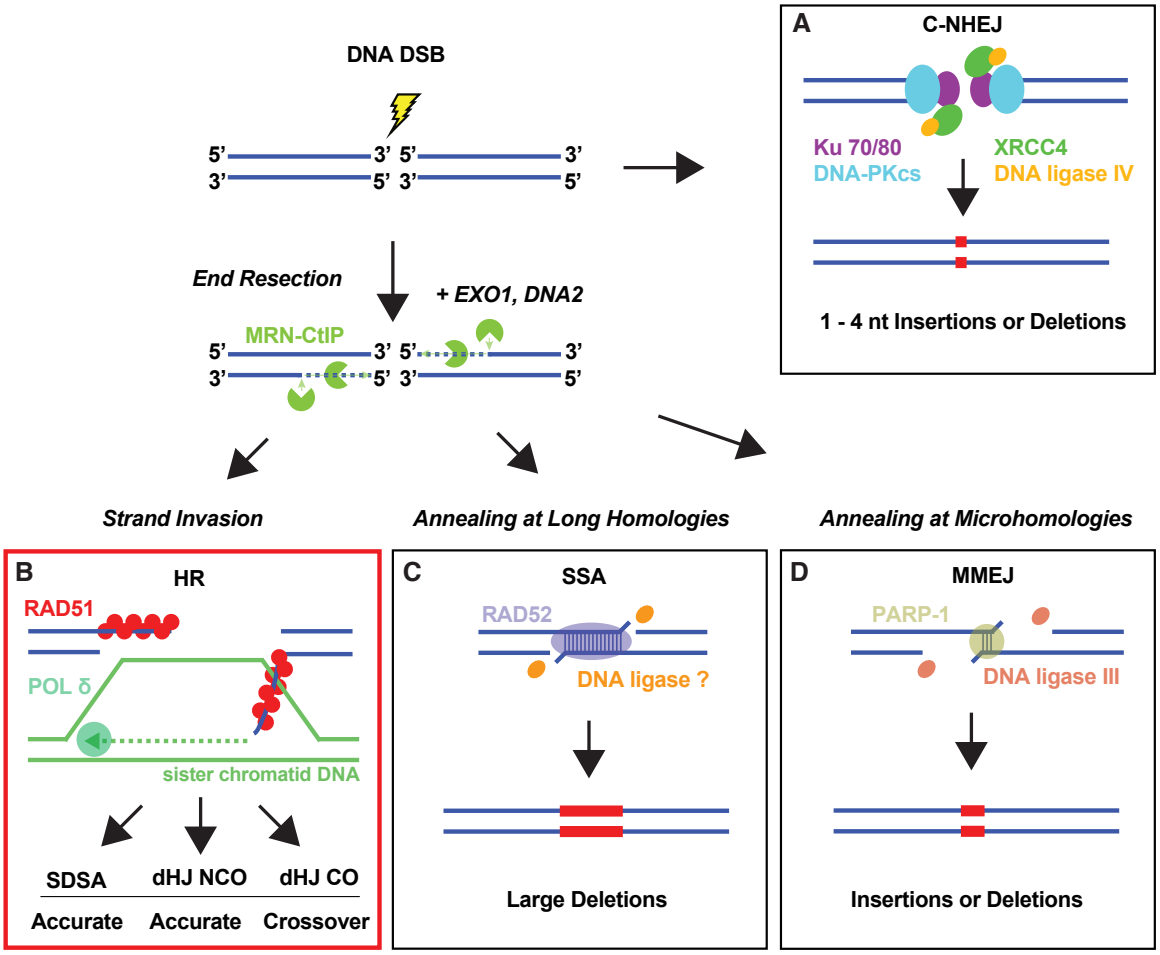


Figure 1.1: Homologous recombination (HR) and alternative DSB repair pathways. **A**. Canonical Non-Homologous End Joining (C-NHEJ) occurs throughout the cell cycle, involves the direct ligation of DNA ends with little or no homology [76] and is error-prone because it can lead to short insertions or deletions at the DNA ends [77]. C-NHEJ can also trigger chromosomal translocations when multiple DSBs arise within several chromosomes [78]. C-NHEJ requires the Ku70/80 heterodimer in complex with DNA-PKcs to protect the DNA ends from degradation. DNA ligase IV is then recruited to the DSB in complex with the scaffolding protein XRCC4 to ligate the DNA ends [79, 80, 77]. **B**. HR occurs during the S and G_2 cell cycle phases and is triggered via extensive DNA end resection. MRE11 within the MRN-CtIP complex catalyses the initial DNA-clipping step via an endonucleolytic cleavage followed by 3' - 5' exonuclease activity. Extensive 5' - 3' end resection is the catalysed by DNA2 (in complex with BLM or WRN) or EXO1. RAD51 polymerises on the ssDNA overhangs and catalyses strand invasion on the sister chromatid. This is followed by 1st end DNA capture and polymerase δ-mediated DNA synthesis. DSB repair can then follow three pathways: Synthesis-Dependent Strand Annealing (SDSA - accurate; see Fig. 1.2), double Holliday junction resolution leading to a non-crossover product (DHJ NCO - accurate), and double Holliday junction resolution leading to a crossover product (DHJ CO - usually leads to a crossover between homologous chromosomes during meiosis) [81]. **C**. Single-strand Annealing (SSA) occurs when complementary repeat sequences flanking a DSB anneal following extensive 5' - 3' end resection [82, 83]. Non-complementary 3' ssDNA tails are cleaved via endonuclease activity and DNA gaps are filled by DNA polymerases, leading to large deletions. RAD52 (Rad52 in *Saccharomyces cerevisiae*) is essential to mediate annealing of the long (> 30 base pairs) complementary sequences [84, 85], although the specific polymerases and ligases involved are still unknown. **D**. Microhomology-Mediated End Joining (MMEJ) is a highly mutagenic end-joining mechanism that involves direct ligation of DNA ends with short (< 30 base pairs) complementary sequences [2]. It leads to small insertions or deletions at the DSB site and promotes chromosomal translocations by joining DSBs on different chromosomes [86, 87]. MMEJ initiation involves MRE11-CtIP-dependent DNA clipping [87]. The processes of damage recognition and repair involve the concerted action of PARP1 and the XRCC1/ligase III complex [88]. This Figure is based on Figures 1 and 4 from [2].

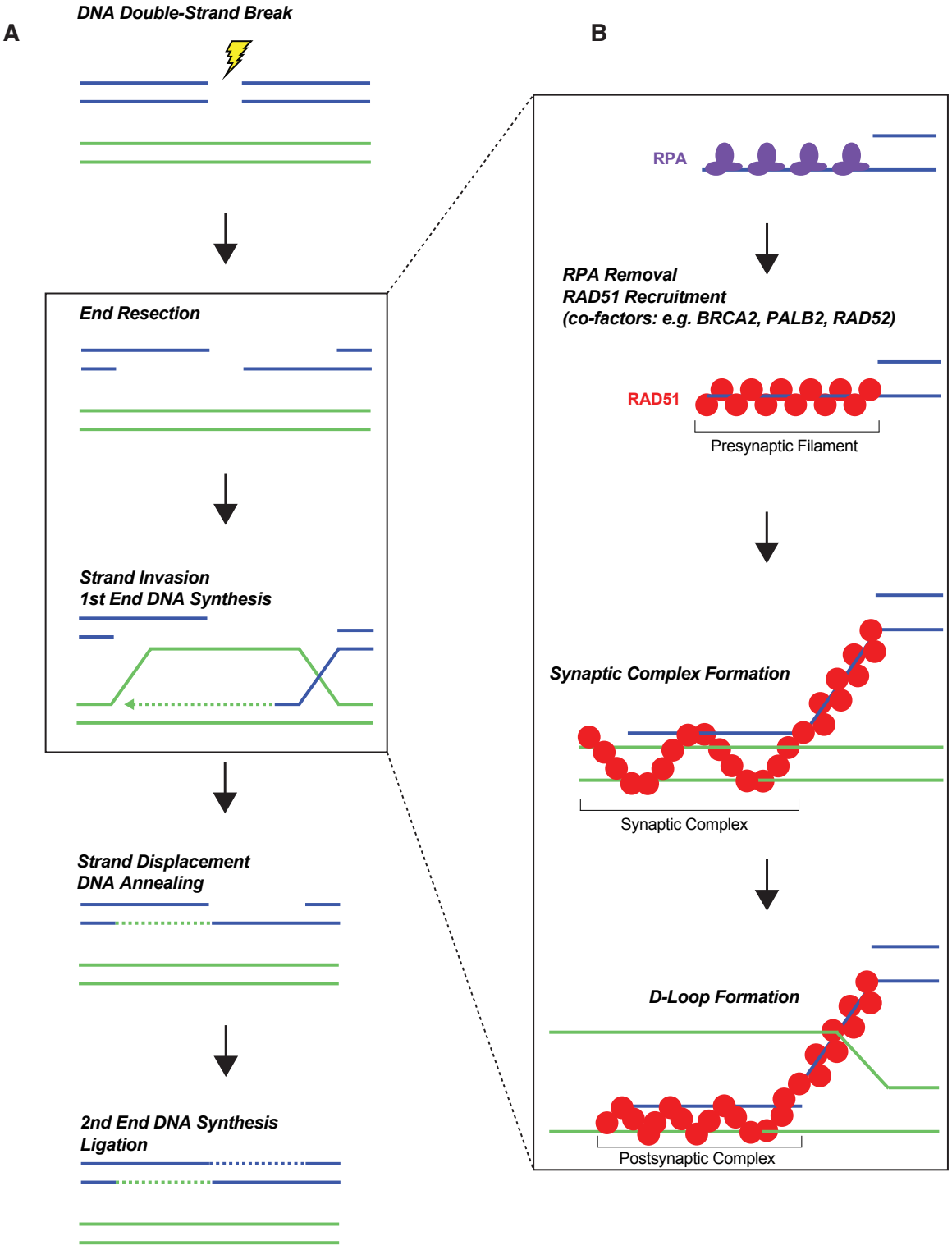


Figure 1.2: HR and the Function of RAD51. **A**. The most common HR pathway in somatic cells, also known as Synthesis-Dependent Strand Annealing (SDSA). This pathway does not lead to crossover products. Holliday-junction-dependent HR pathways leading to crossover products are reviewed elsewhere [81, 89]. **B**. The different RAD51-dependent steps in HR. This Figure is based on Figures 2 and 3 from [81].

1.2 The RAD51 Nucleoprotein Filament

RAD51 forms right-handed filaments on DNA to guide strand invasion and catalyse HR. The protein sequence of RAD51 is evolutionarily conserved (Fig. 1.4A), and its tertiary structure is similar to that of *S. cerevisiae* Rad51 [90] and archaeal RadA [91, 92]. In recent years, key biochemical and structural features of RAD51 pre-synaptic (RAD51-ssDNA) and post-synaptic (RAD51-dsDNA) filaments have been elucidated via x-ray crystallography, electron microscopy (EM) and single molecules studies.

1.2.1 The Roles of Nucleotide and Cation Binding

The nucleotide state of RAD51 plays a critical role in determining the biochemical and structural features of RAD51 filaments. Pre-synaptic filament assembly, disassembly and strand-exchange activities are all known to be coupled to RAD51-dependent ATP binding and hydrolysis [93, 56, 94, 95]. For human RAD51, *S. cerevisiae* Rad51 and *E. coli* RecA, ATP binding is associated with a high-affinity DNA-bound state required to form catalytically competent nucleoprotein filaments for homologous DNA pairing, while hydrolysis of ATP to ADP weakens interactions with DNA and promotes filament disassembly [96, 97, 95, 98, 99, 100]. However, on dsDNA human RAD51 and *S. cerevisiae* Rad51 hydrolyse ATP at the rates of 0.1 and 0.7 ATP molecules / min respectively [101, 56, 96], while RecA hydrolyses ATP at a much higher rate of 25 - 30 ATP molecules / min [98, 102]. Due to its low ATP hydrolysis rate, human RAD51 requires the activity of additional protein co-factors such as RAD54 to facilitate dissociation from homology search intermediates [103]. Analyses using DNA curtain assays have instead proposed the high ATP hydrolysis of RecA enables fast turnover of dsDNA during homology search [104], although the protein co-factor RecX is known to promote RecA dissassembly from ssDNA [105]. Additionally, Mg^{2+} and Ca^{2+} cations coordinate the oxygen atoms of nucleotide phosphates within the RAD51 nucleotide binding pocket, and

are known to promote and inhibit RAD51-dependent ATP hydrolysis, respectively [101, 106]. In particular low concentrations ($\approx$ 0.1 mM) of Mg^{2+} were shown to enhance ATP hydrolysis by RAD51, but high concentrations ($\geq$ 3 mM) were shown to inhibit ATP hydrolysis by decreasing the rate of ADP release [107]. Structural analyses have also suggested Ca^{2+} induces a more ordered nucleoprotein filament conformation compared to Mg^{2+}, with preferentially perpendicular orientation of nucleobases relative to the filament axis due to an additional Ca^{2+} cation stabilising the C-terminal region of the RAD51 L2 DNA-binding loop (residues 283 - 292) [108].

1.2.2 Active Versus Inactive Filaments

Importantly, RAD51 high-affinity binding to DNA in the presence of ATP or non-hydrolysable ATP analogs is associated with the formation of filaments with high pitch (i.e. the height of a single helical turn along the helical axis), also known as active filaments. In these high pitch RAD51-ssDNA and RAD51-dsDNA filaments, RAD51 stretches the underlying DNA to $\approx$ 150% the length of free B-DNA [109, 110] (Fig. 1.3). Upon ATP hydrolysis and prior to filament disassembly, RAD51 transitions into a low affinity DNA-binding state and forms compressed filaments with lower pitch [111, 112]. Compressed pre-synaptic filaments are referred to as inactive filaments due to their inability to effectively catalyse D-loop formation [94] (Fig. 1.3). In particular, both cryo-EM and negative stain EM analyses have suggested RAD51 forms pre-synaptic filaments with an average $\approx$ 10 nm pitch in the presence of the non-hydroysable ATP analogs AMP-PNP [113, 60] or ADP-AlF_4^- [112] (Table 1.1). Similarly, RAD51 post-synaptic filaments in the presence of AMP-PNP also display an average pitch of $\approx$ 10 nm [93, 60]. In contrast, it was suggested that RAD51-ADP forms compressed pre-synaptic filaments, with an average pitch of $\approx$ 8 nm, although this was proposed due to the similarity of the RAD51-ADP filament pitch (data not reported) to that of the RAD51-ATPγs filament pitch [112].

Both RAD51 pre-synaptic and post-synaptic filaments display 6.3 - 6.4 RAD51 protomers per helical turn [112, 113, 60] (Table 1.1). *S. cerevisiae* Rad51 [114, 112, 97], *E. coli* RecA [115, 116, 117, 118] and *M. voltae* RadA [119] also form nucleoprotein filaments with similar pitch and protomers per helical turn compared to their human RAD51 counterparts (Table 1.1).

However, a study combining optical tweezers, single molecule fluorescence microscopy and x-ray crystallography recently suggested ATP-bound RAD51 forms compact and extended filaments on ssDNA via two distinct protomer-protomer orientations [120], suggesting a structural similarity with yeast Rad51 [90]. Whether RAD51 displays two distinct ATP-bound states also on dsDNA remains unanswered.

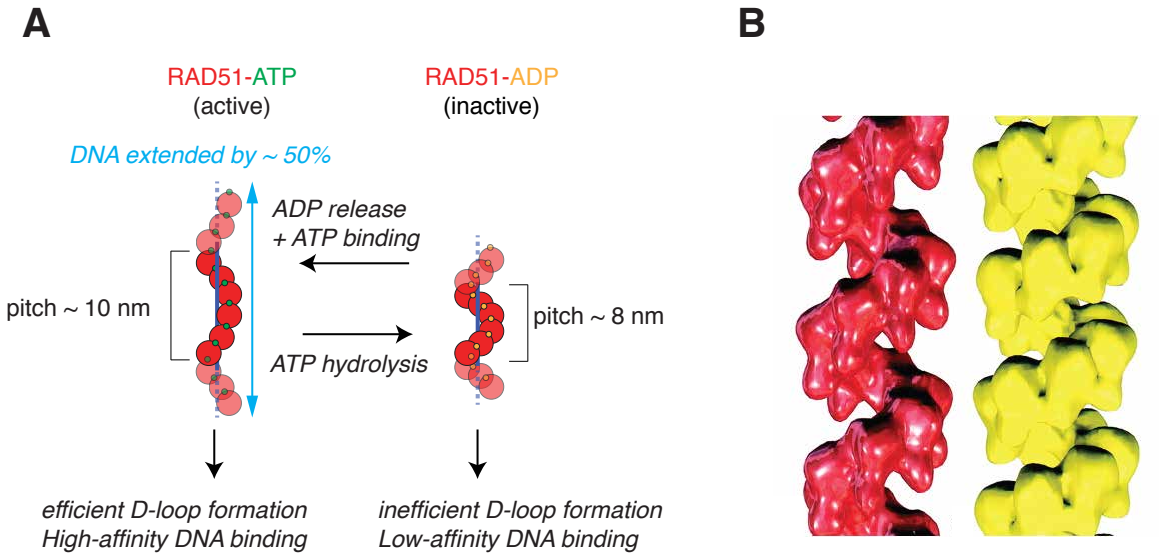


Figure 1.3: **A**. Cartoon representation of the reversible interconversion [121] between RAD51-ATP (active) and RAD51-ADP (inactive) filaments. RAD51-ATP forms filaments with an extended pitch ($\approx$ 10 nm [112]) that stretch the underlying DNA to 150 % of the length of B-DNA, while ATP hydrolysis to ADP promotes filament contraction to a $\approx$ 8 nm pitch [112] and abolishes extension of DNA [121]. Active filaments catalyse efficient D-loop formation and show high affinity DNA-binding, while inactive filaments are less proficient at D-loop formation and display lower affinity DNA-binding [96, 97, 95, 98, 99, 100]. **B**. Transmission electron microscopy image reconstruction of the RAD51-ssDNA active filament (left: RAD51 bound to the ATP analog ADP-AlF_4^- [112]) and inactive filament (right: RAD51 bound to the ATP analog ATPγs. Importantly, the authors stated that the structure of RAD51-ATPγs-ssDNA filament was very similar to that of RAD51-ADP, but did not report the structure for the RAD51-ADP-ssDNA filament [112]). Reproduced from [112].

Table 1.1: Helical pitch and protomers-per-turn for of *H. sapiens* RAD51 (HsRAD51), *S. cerevisiae* Rad51 (ScRad51), *M. voltae* RadA (MvRadA) and *E. coli* RecA (EcRecA) filament structures. For Cryo-EM and NS EM, helical pitch and protomers-per-turn values represent distribution averages. The compressed HsRAD51-ssDNA (ATPγS) structure is proposed to mimic the compressed structure of HsRAD51-ssDNA (ADP) (inactive filament) [112]. For ScRad51 (no DNA) (ATPγS), ssDNA was included in the crystallisation sollution but no ssDNA electron density was detected. Its $\approx$ 13 nm helical pitch is higher than the expected average $\approx$ 10 nm as measured for ScRad51-dsDNA (ATPγS) [112]. nucl: nucleotide or nucleotide analog. Cryo-EM: cryo-electron microscopy. NS EM: negative stain electron microscopy. Cryst: X-ray crystallography.

Structure (nucl)	Method (resolution)	Pitch (nm)	Protomers / Turn
HsRAD51-ssDNA (AMP-PNP) [113]	Cryo-EM (4.2 Å)	10.3	6.4
HsRAD51-ssDNA (AMP-PNP) [60]	Cryo-EM (4.4 Å)	10	6.3
HsRAD51-ssDNA (ADP-AlF_4^-) [112]	NS EM	9.9	6.39
HsRAD51-ssDNA (ATPγS) [112]	NS EM	7.6	6.43
HsRAD51-dsDNA (AMP-PNP) [60]	Cryo-EM (4.5 Å)	10	6.3
ScRad51 (no DNA) (ATPγS) [90]	Cryst (3.25 Å)	13	6
ScRad51-dsDNA (ATPγS) [112]	NS EM	9.4	6.28
ScRad51-dsDNA (ATP) [114]	NS EM	9.9	6.18
MvRadA (no DNA) (AMP-PNP) [119]	Cryst (2.0 Å)	10.7	6
EcRecA-ssDNA (ADP-AlF_4^-) [117]	Cryst (2.8 Å)	9.4	6.16
EcRecA-dsDNA (ADP-AlF_4^-) [117]	Cryst (3.15 Å)	9.4	6.2
EcRecA-dsDNA (ATPγS) [112]	NS EM	9.1	6.16
EcRecA (no DNA) (ADP) [115]	Cryst (2.3 Å)	8.27	6.1

Single molecule fluorescence microscopy analyses have also shown that compressed RAD51 post-synaptic filaments can re-extend into the high-pitch form when presented with ATP or AMP-PNP, and upon ATP hydrolysis can undergo reversible cycles of compression and extension [121] (Fig. 1.3), similarly to RecA [122]. Furthermore, it was proposed that allosteric interactions within filaments allow adjacent RAD51 protomers to sense and respond to the nucleotide-bound state of their immediate neighbours, therefore triggering the transition from the RAD51-ATP extended state to the RAD51-ADP compressed state, and viceversa [121]. Residues F129 within the N-terminal Walker A motif (involved in nucleotide binding) and H294 within the L2 DNA-binding loop of an adjacent protomer have been proposed to regulate this allosteric conformational change [123, 124]. Interestingly, another study using magnetic tweezers suggested that the inter-conversion between compressed and extended post-synaptic filaments is dependent on the ability of RAD51 to unwind dsDNA from $\approx$ 10.5 bp/turn ($\approx$ 3.6 nm/turn) [125] to $\approx$ 18.6 bp/turn ($\approx$ 6.4 nm/turn) [126].

The elastic nature of RAD51 nucleoprotein filaments is considered to play a role in HR regulation. Importantly, the RAD51 co-factor BRCA2 is proposed to stabilise RAD51 filaments in an extended state via its TR2 domain and inhibits filament dissociation via its BRC RAD51-binding motifs, therefore promoting strand-exchange catalysis during the initiation of HR [127, 4, 3]. Indeed, a single-molecule study using magnetic tweezers suggested the BRCA2 TR2 domain peptide on its own does not decrease RAD51 dissociation speed, suggesting other BRCA2 domains, such as the BRC motifs, may also be involved in RAD51 filament stabilisation [128]. However, the regulation and biological significance of RAD51 nucleoprotein filament elasticity is not fully understood. Several other RAD51 co-factors, such as RAD52 [10], RAD54 [63] and RAD51 paralogs [50, 51, 52] may also regulate RAD51 assembly on ssDNA and dsDNA during HR by modifying RAD51 filament

elasticity.

1.2.3 RAD51-DNA Interactions

Cryo-EM analysis has shown how RAD51 uses its C-terminal domain to form polar interactions with the ssDNA and dsDNA backbones [113, 60]. In particular, RAD51 has been shown to bind to ssDNA and dsDNA with a RAD51 molecule : nucleotide / base pair stoichiometry of 1:3 [113]. On ssDNA, protomer residues N290, G289 and R229 of the L2 loop (residues 268 - 292) within the ATPase domain make key contacts with the nucleotide triplet phosphates [60] (Fig. 1.4B, see the central, cyan-coloured RAD51 protomer, labelled RAD51 - 0). On dsDNA, L2 loop residues N290, G288, G289 form similar interactions only with the proximal DNA strand (Fig. 1.4C, RAD51 - 0). Each RAD51 protomer also stabilises the inter-nucleotide triplet separation on ssDNA via V273 (Fig. 1.4B, RAD51 - 0 and green-coloured RAD51 protomer, RAD51 - 5'), and the inter-base pair triplet separation on dsDNA via R235, A271, and V273 [60] (Fig. 1.4C, RAD51 - 0 and RAD51 - 5').

1.2.4 RAD51 Protomer-Protomer Interactions

RAD51 forms protomer-protomer interactions mainly via three interfaces. The first interface consists of a short β-strand of the linker region (residues 85 - GFTTA - 89) between the N-terminal domain and the ATPase domain of one protomer (Fig. 1.4D, RAD51 - 0) binding to the central β-sheet in the ATPase domain of the adjacent protomer (Fig. 1.4D, RAD51 - 5') [60]. Specifically, **F86** forms key hydrophobic interactions with the ATPase domain of the adjacent protomer, and mutations to this residue strongly decrease the RAD51 protomer-protomer affinity [129, 92, 3] (Fig. 1.4D). The second interface is mediated by a sandwiched ATP molecule. In particular, the Walker A motif residues K133 and T134 on one protomer (Fig. 1.4E, RAD51 - 5') and the Walker B motif residues D316 and P321 on the adjacent protomer stabilise the ATP molecule (Fig. 1.4E, RAD51- 0) [60]. The third interface is

mediated by a hydrophobic interaction between Y54 in one protomer (RAD51 - 0), and F195 in the adjacent protomer (RAD51 - 5') [60].

Despite the development of detailed molecular structures of RAD51-DNA filaments, the mechanism of RAD51 filament formation on ssDNA and dsDNA remains unclear and highly disputed.

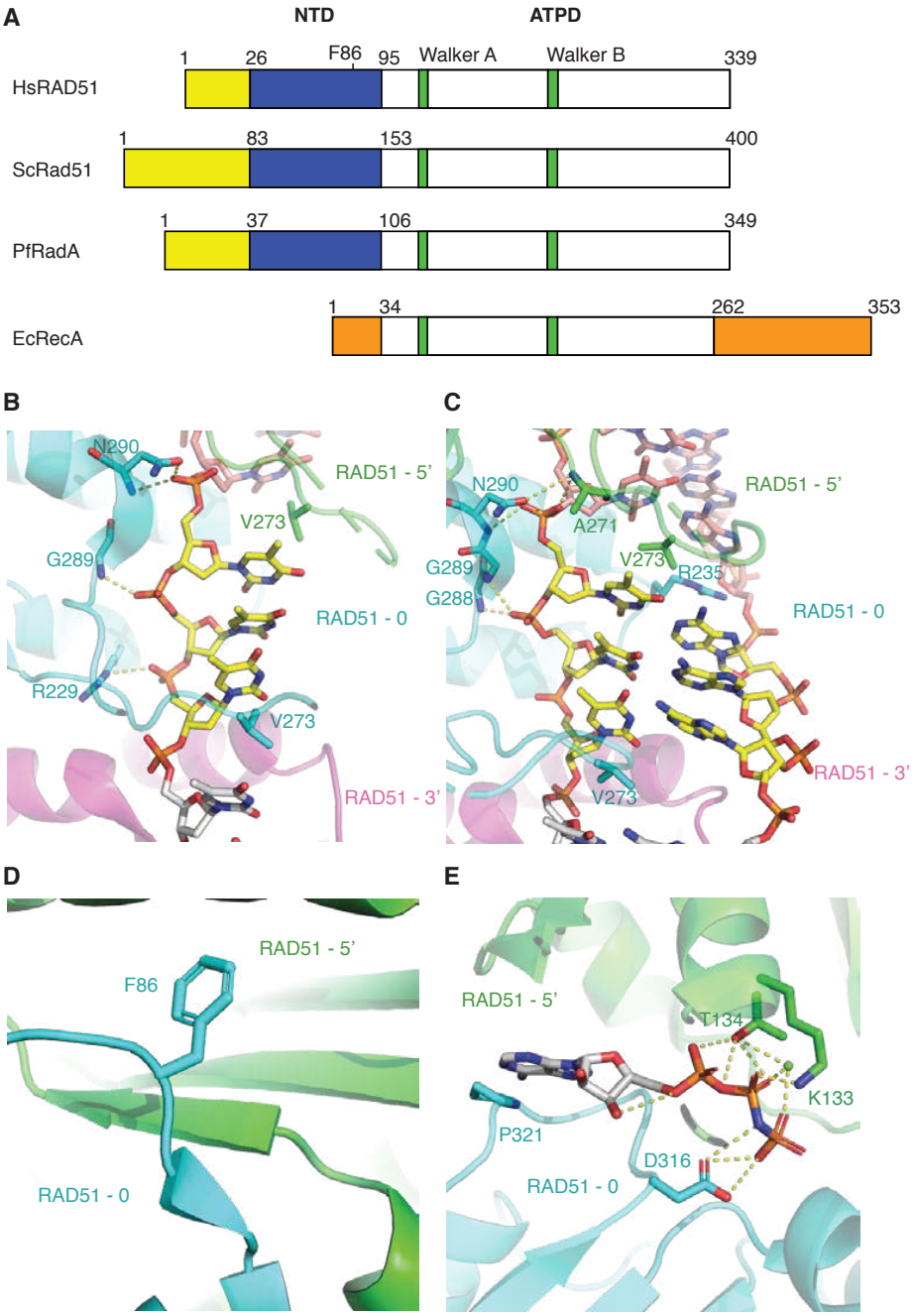


Figure 1.4: The sequence homology of RAD51 orthologs and the structure of RAD51-DNA filaments. **A**. Cartoon comparison of four RAD51 orthologs: *Homo sapiens* RAD51, (HsRAD51), *S. cerevisiae* Rad51, (ScRad51), *Pyrococcus furiosus* RadA (PfRadA) and *Escherichia coli* RecA, (EcRecA). Regions of homology between the N-terminal domains (NTD) and the ATPase domains (ATPD) are coloured in blue and white, respectively. Walker A and B motifs are denoted in green. Non-homologous regions are coloured in yellow and orange. The HsRAD51 F86 residue position, as well as the first and last residue positions of each ortholog region, are denoted. This Figure is based on Figure 1F from [91]. **B**. RAD51 interactions with ssDNA. **C**. RAD51 interactions with dsDNA. **D**. RAD51 protomer-protomer interaction via the F86 linker region residue (RAD51 - 0) and the ATPase domain central β-sheet (RAD51-5'). **E**. Nucleotide-(AMP-PNP) and cation-(Mg^{2+}: green ball) mediated RAD51 protomer-protomer interaction. AMP-PNP is a non-hydrolysable ATP analog used to crystalise RAD51 in a nucleotide-bound state. **B-E**. The three RAD51 protomers are coloured in green (RAD51 - 5'), cyan (RAD51 - 0) and magenta (RAD51 - 3'). Carbon atoms of nucleotide / base pair triplets are coloured light red (5'), yellow and white (3'). DNA, AMP-PNP and key RAD51 residue side chains are highlighted as ball-and-stick. Main chains of N290 and G289 are also represented. Dotted lines represent polar interactions. Images were generated using PyMOL (PDB 5H1B, 5H1C [60]).

1.3 RAD51 Polymerisation

During HR initiation, RAD51 polymerises specifically on ssDNA at DSBs and not on undamaged dsDNA. How is this accomplished? The mechanism of HR is evolutionarily conserved and the bacterial ortholog of RAD51, RecA, preferentially polymerises on ssDNA compared to dsDNA [93, 130]. RAD51 has lower affinity to ssDNA compared to RecA [93], and the mechanism by which it polymerises on ssDNA over dsDNA is unknown. Early biochemical studies and recent single-molecule studies have provided limited insight into the mechanisms of RAD51 polymerisation on ssDNA and dsDNA.

1.3.1 Static Analyses

Early studies analysing RAD51 association to DNA were based on the electrophoretic mobility shift assay (EMSA) (i.e. a qualitative end-point assay that measures protein-DNA binding as a band-shift on an agarose gel), and negative stain EM (Fig. 1.5A).

In low salt conditions ($\approx$ 40 mM KCl), RAD51 was shown to display higher affinity to dsDNA compared to ssDNA [56]. Interestingly, EMSA analyses also highlighted that under moderate salt conditions ($\approx$ 150 mM NaCl) and in the presence of ADP, RAD51 is more prone to aggregation compared to RecA and suggested RAD51-ADP is unable to effectively form nucleoprotein filaments on ssDNA [131]. However, two studies later suggested RAD51 had higher affinity for ssDNA compared to dsDNA in the presence of ammonium sulphate [132, 111]. Indeed, a systematic analysis of neutral salts (i.e. salts with no acidic or basic character, especially when dissolved in water) found that large cations such as K^+ and NH_4^+ increase the relative affinity of RAD51 for ssDNA compared to dsDNA and promote RAD51-dependent strand-exchange reactions *in vitro* [133]. In accordance with this, it was suggested that two K^+ or NH_4^+ cations bind to F129 within the Walker A motif and H294 within the L2 ssDNA-binding loop and trigger a structural rearrangement of

the L2 loop that promotes effective pre-synaptic filament assembly and strand exchange [124]. This was largely inferred based on the structural similarity of human RAD51 with the *Methanococcus voltae* RadA, in which two K^+ were shown to alter the spatial orientation of F107 within the Walker A motif and H280 within the L2 loop [134] (Fig. 1.5A-C). Importantly, it was later suggested using negative stain EM that RAD51 on ssDNA in the presence of ATP and Mg^{2+} forms $\approx$ 8 nm (i.e. inactive) filaments in the absence of monovalent salts, but forms $\approx$ 10 nm pitch (i.e. active) filaments in the presence of 200 mM K^+ or NH_4^+ [135], the former being more physiologically relevant (Fig. 1.5D). This further supports the idea that monovalent cations are required for catalysis of strand exchange. However, it is unclear whether K^+ availability regulates RAD51 filament formation *in vivo*.

Overall, static analyses have provided limited, qualitative observations on the affinity of RAD51 for ssDNA and dsDNA.

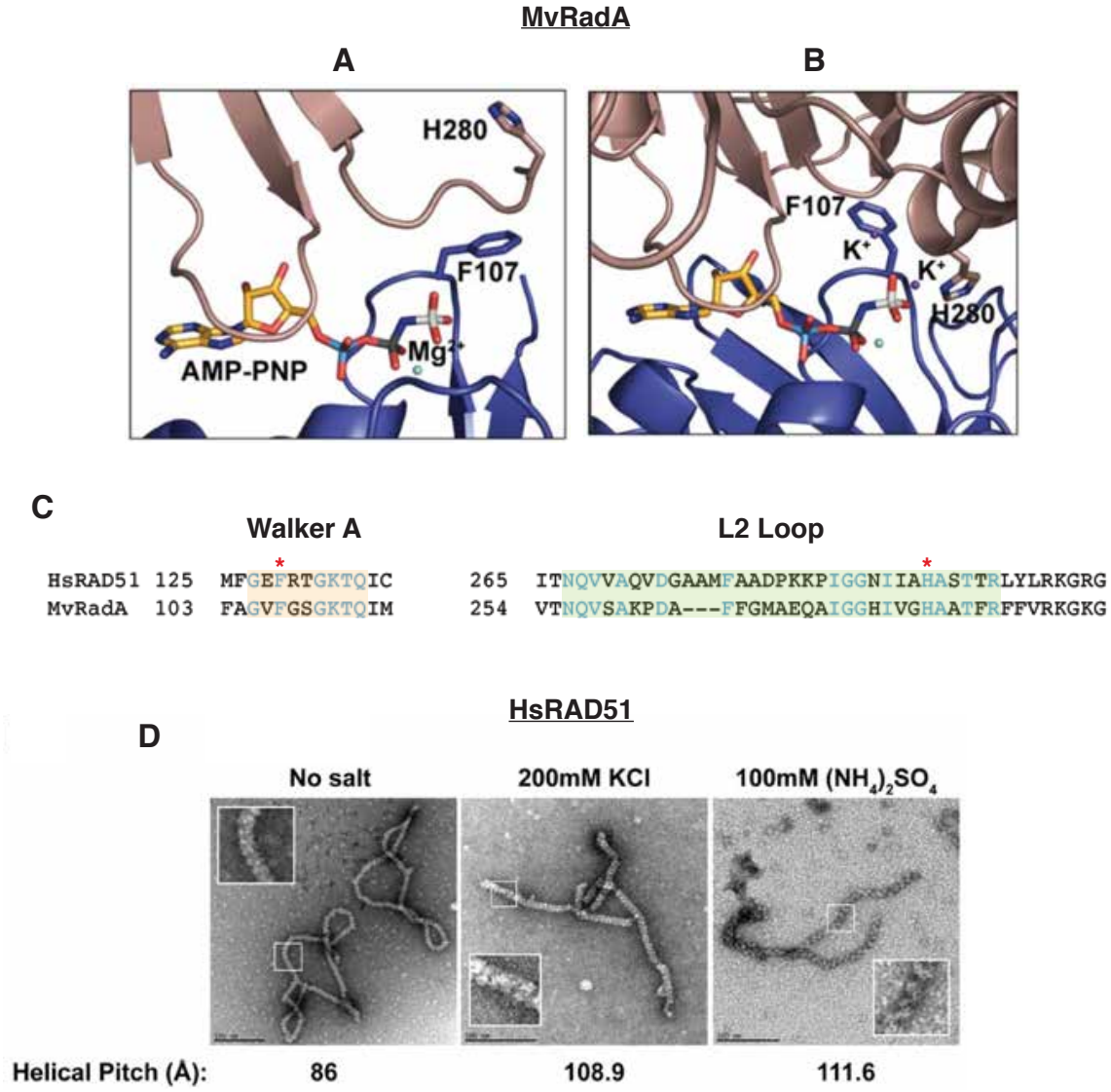


Figure 1.5: The impact of monovalent cations on presynaptic filament formation. **A**. Structure of the MvRadA protomer-protomer interface in the absence of K^+ cations (PDB 1T4G). **B**. Structure of the MvRadA protomer-protomer interface in the presence of K^+ cations (PDB 1XU4). **A-B**. The two RAD51 protomers are coloured in blue and salmon. AMP-PNP and F107 and H280 side chains are depicted in ball-and-stick form. Mg^{2+} and K^+ cations are depicted as cyan and purple balls, respectively **C**. Amino acid sequence alignment of HsRAD51 and MvRadA Walker A and L2 loop. The Walker A motif is highlighted in salmon, and the L2 loop is highlighted in light green. Conserved residues are indicated in light blue. HsRAD51 residues F129 and H294 (MvRadA residues F107 and H280) are indicated with red asterisks. **A-B** are reproduced from [124]. **C** is based on Fig. 1C from [124] and was generated using the hsRAD51 and MvRadA UniProt amino acid sequences and the UniProt Align application (https://www.uniprot.org/align/). **D**. Negative-stain EM analysis of HsRAD51 pre-synaptic filaments (+ ATP + Mg^{2+}) in the presence of no salt (left), 200 mM KCl (center), or 100 mM ammonium sulphate (right). Average helical pitches of filaments in each condition are depicted below (pitch standard deviations were not reported). Insets depict enlarged images of the filaments. Reproduced from [135].

1.3.2 Kinetic Analyses

More recent kinetic studies used either magnetic tweezers, optical tweezers together with single-molecule fluorescence microscopy, or surface plasmon resonance (SPR) to gain a more mechanistic understanding of RAD51 polymerisation on ssDNA and dsDNA.

Two studies using magnetic tweezers suggested that the rate-limiting step of RAD51 polymerisation is the formation of a $\approx$ 5 RAD51 protomer nucleus on DNA [109, 11]. Importantly, multiple RAD51 nucleation events were shown to occur on single dsDNA molecules [136, 137], with an average of one event every 500 bps when RAD51 is incubated with dsDNA at a molar ratio of 1 protomer per 3 bps [137]. A study using single-molecule fluorescence microscopy and optical tweezers suggested instead nucleation involves the binding of 2 - 3 RAD51 protomers and that RAD51 binding on both ssDNA and dsDNA can be stabilised via inhibition of ATP hydrolysis [110].

A seminal study combining single molecule fluorescence microscopy, optical tweezers and microfluidics proposed that RAD51-dependent ATP hydrolysis occurs uniformly throughout the post-synaptic filament while RAD51 disassembly only occurs when terminal RAD51 protomers have hydrolysed ATP, leading to RAD51 disassembling in bursts interspersed with pauses [138] (Fig. 1.6). In addition, the study suggested the rate of ATP hydrolysis is independent of DNA tension, while the rate of dissociation of RAD51-ADP from filament termini exponentially decreases with applied DNA tension. These observations support the idea that RAD51 dissociation from filament ends can only occur via ATP hydrolysis followed by the release of RAD51-induced tension within the nucleoprotein filament [138]. However, it is still unclear whether ATP hydrolysis occurs uniformly within the RAD51 filament as previously assumed for RecA [102], or whether it may occur in waves as more re-

cently proposed for RecA [139].

Magnetic tweezer studies have suggested RAD51 polymerises faster on ssDNA but depolymerises more slowly on dsDNA, suggesting the RAD51-dsDNA complex is stable once formed [11]. Finally, another single molecule study suggested RAD51 nucleation under physiological salt conditions (100 - 150 mM KCl) is favoured on ssDNA compared to dsDNA. The study also highlighted that pre-formed RAD51 polymers can bind directly from solution onto DNA and are heterogeneous in size, ranging from monomers to 10-mers [12].

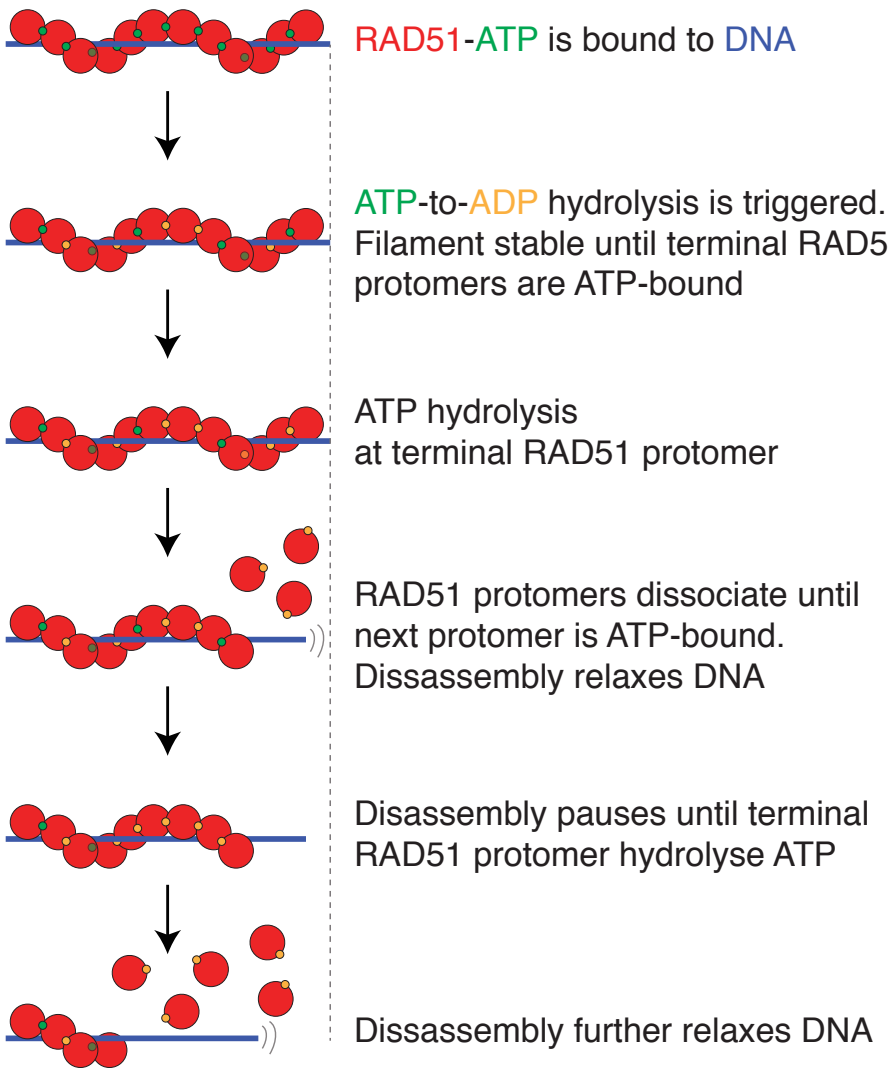


Figure 1.6: The model proposed by van Mameren et al. describing RAD51 filament dissassembly on DNA at the single molecule level [138]. The model proposes RAD51 dissociation, and not ATP hydrolysis to ADP, promotes the transition of DNA from the extended state to the relaxed state.

Three studies have used SPR to quantitatively analyse RAD51 binding kinetics to DNA using SPR [140, 124, 135]. The first used SPR to analyse how nucleotide state affects RAD51 cooperative binding to ssDNA [140], and by analysing the SPR data using a cooperative binding model [141, 142], it was proposed that ATP binding reduces the monomeric RAD51 affinity to ssDNA and consequently promotes efficient RAD51 polymerisation on ssDNA. However, these analyses relied on the

assumption that RAD51 polymerisation on DNA occurs via the sequential addition of monomers, an assumption which currently is not experimentally supported [12]. Faster polymerisation on ssDNA in the presence of ATP compared to ADP was also qualitatively observed in another study using SPR [131]. The two other studies (previously mentioned in section 1.3.1) used SPR to support their findings that residues F129 (Walker A motif) and H294 (L2 loop of adjacent protomer) are key for efficient RAD51 polymerisation on both ssDNA and dsDNA [124], and that D316 promotes ATP hydrolysis and RAD51 filament dissassembly [135]. Interestingly, both studies also showed faster RAD51 association to ssDNA compared to dsDNA, although neither study analysed this any further.

Although kinetic analyses have suggested RAD51 polymerisation is faster on ssDNA compared to dsDNA, the molecular mechanism driving this difference remains unclear.

1.4 The Impact of DNA Flexibility on Protein-DNA Interactions

DNA is a highly dynamic structure that can both bend and twist. DNA flexibility can be described by two parameters: torsional flexibility, characterising variations in the twist angle between adjacent nucleotides or base pairs, and bending flexibility, characterising deviations of the ssDNA or dsDNA long axis from a straight trajectory [143]. DNA flexibility depends on the stacking energy of base steps, the number of hydrogen bonds within base pairs and the presence of exocyclic base groups (i.e. the chemical groups outside the base ring structure) in the major and minor grooves [143].

DNA torsional flexibility plays a major role in DNA transcription and replication. Both processes require the untwisting and melting of dsDNA for initiation, and this was

shown to occur primarily at TA-rich sequences due to the lower energy barrier required for melting compared to GC-rich sequences [144]. Bending flexibility instead is known to play a major role in DNA packaging within eukaryotic cell nuclei, the key example being the packaging of 147 DNA base pairs around a histone octamer to form a nucleosome [145].

DNA bending flexibility (from now on referred to as flexibility) can be described via the persistence length (L_p), a mechanical parameter quantifying polymer stiffness. The persistence length approximately corresponds to the length scale on which a polymer is stiff. ssDNA is more flexible than dsDNA and has an L_p of $\approx$ 1.5 - 3 nm [146, 147, 148], similar to the length of a single nucleotide triplet RAD51 binding site ($\approx$ 2 nm [147]), while dsDNA has an L_p of $\approx$ 30 - 55 nm [149, 150] (Fig. 1.7A). However, the impact of DNA flexibility on RAD51 polymerisation, and on protein-DNA interactions in general, remains understudied.

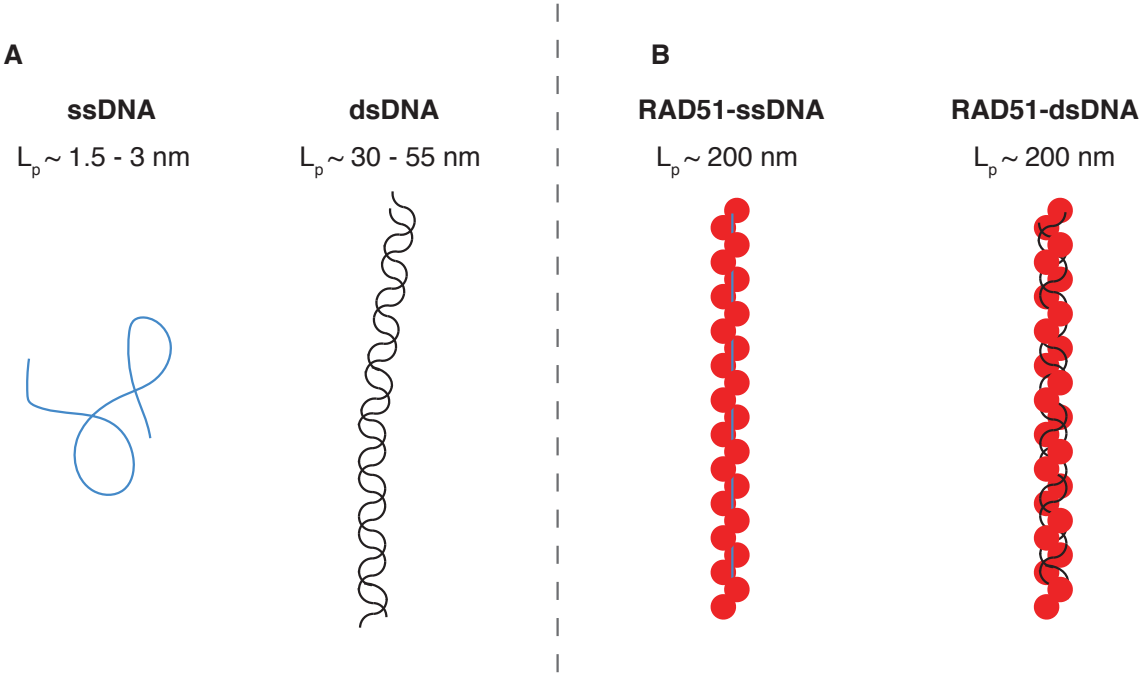


Figure 1.7: Schematics of the flexibilities of ssDNA and dsDNA with and without RAD51. **A**. Flexibilities of ssDNA and dsDNA [149, 147, 148]. **B**. Flexibilities of RAD51-ssDNA and RAD51-dsDNA filaments [11, 109].

1.4.1 Higher DNA Flexibility - Faster Recruitment of Repair Factors?

Early studies analysed the impact of DNA conformational changes on the repair efficiency of damaged or misincorporated DNA bases. In particular, a study using molecular dynamics simulations suggested that the O^6-alkylguanine-DNA alkyltransferase enzyme involved in mismatch repair preferentially binds to DNA regions with increased local flexibility due to the presence of the modified base O^4methylthymine [151]. Another study suggested that increased local DNA flexibility due to misincorporation of uracil enhances binding of the base excision repair (BER) enzyme uracil DNA glycolase, which removes uracil from ssDNA and dsDNA [152]. A later study also suggested this BER enzyme recognises uracil in DNA because it spends more time in an extrahelical conformation compared to a normal DNA base [153].

More recent studies have eluclidated the impact of DNA duplex conformation in nucleotide excision repair (NER). Importantly, NER enzymes recognise their substrates because they are bulky and they destabilise the DNA duplex conformation [154]. In particular, efficient DNA damage recognition by NER enzymes such as XPC-RAD23 has been correlated with disrupted base pairing, local DNA bending and changes in local DNA flexibility at, for example, acetylaminofluorine and benzo[a]pyrene DNA adducts [155, 156, 157, 158].

Furthermore, previous studies on the Eukaryotic High-Mobility Group B (HMGB) proteins have also suggested that DNA flexibility plays a role in regulating DNA repair, as well as chromatin remodelling and transcriptional activation [159, 160]. HMGB proteins are abundant, ubiquitously expressed, sequence non-specific DNA-binding proteins that unwind and bend DNA to increase local flexibility [161, 162]. In particular, single molecule studies using optical tweezers and atomic force microscopy have shown that HMGB-2 [163] and HMO1 [164] can decrease the L_p

of naked dsDNA $\approx$ 10-fold. *In vitro*, HMGB-1 was previously shown to inhibit the repair of cisplatin-induced DNA adducts via NER [165, 166], while it has also been shown to enhance the repair of psoralen-induced interstrand crosslinks by directly interacting with damaged DNA, RPA and NER proteins [167].

Evidence of the impact of DNA flexiblity on DSB repair is scarce. Although HMGB-1 and HMGB-2 have also been implicated in facilitating the generation of DSBs required for V(D)J recombination in B cells and T cells via direct interaction with the lymphoid-specific RAG1/2 proteins [168], it is unclear whether they also play a role in DSB repair. Furthermore, a recent study using optical tweezers suggested that human RAD52 enhances dsDNA flexibility and prevents DNA melting, potentially facilitating the second-end capture or strand-annealing steps during RAD51-dependent HR reactions [169]. Nonetheless, it remains unclear whether DNA flexibility directly regulates the binding kinetics of RAD51 at resected ssDNA during HR, and whether it contributes to RAD51 recruitment to DSBs *in vivo*.

1.4.2 Overcoming the Entropic Penalty to Binding DNA

Entropy is a thermodynamic measure of randomness within a system and the second law of thermodynamics states that without energy input, systems will evolve towards more randomness or higher entropy. In the case of linear polymers such as DNA, entropy characterises the probability of a polymer having its opposite ends at close distance [170]. In the absence of polymer stretching, a straight linear polymer will have its ends at a distance approximately equal to the contour length of the polymer, and hence have low entropy. The lower the average distance between the ends of the polymer, the higher the polymer entropy and the higher the flexibility [170].

RAD51 polymerisation on DNA incurs an entropic penalty because polymerisation can reduce flexibility leading to a reduction in the number of spatial configurations a DNA molecular can explore. ssDNA is more flexible compared to dsDNA (Fig. 1.7A), and is expected to contribute a larger entropic penalty. This is further supported by the observation that RAD51 polymerisation stiffens ssDNA and dsDNA to a similar extent [11, 109] (Fig. 1.7B). So how does RAD51 overcome the entropic penalty to preferentially polymerise on ssDNA instead of dsDNA?

The Gibbs free energy equation describes how changes in enthalpy (i.e. the internal energy of a chemical system), entropy and temperature determine whether a protein-DNA interaction is energetically favourable [170]:

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

Where $\Delta G = G_{\text{bound}} - G_{\text{unbound}}$ is the change in free energy between the bound and unbound states, $\Delta H = H_{\text{bound}} - H_{\text{unbound}}$ is the change in enthalpy, T is the temperature (constant) and $\Delta S = S_{\text{bound}} - S_{\text{unbound}}$ is the change in entropy. In this framework, a protein-DNA interaction occurs spontaneously if $\Delta G < 0$, or alternatively, $\Delta H < T\Delta S$.

Site-specific proteins such as the bacteriophage λ cI repressor, the bacterial catabolite gene activator protein (CAP), the yeast TATA-binding protein (TBP) and general control protein 4 (GCN4) in complex with dsDNA have all been shown to adopt radically different structural properties to obtain favourable binding free energies ($\Delta G = G_{\text{bound}} - G_{\text{unbound}} < 0$) [171]. In particular, it was suggested that proteins that strongly bend dsDNA (CAP [172], TBP [173]) have a net unfavourable ΔH ($H_{\text{bound}} - H_{\text{unbound}}$) mainly due to base pair destacking, but this is counterbalanced by a favourable ΔS ($S_{\text{bound}} - S_{\text{unbound}}$) due to release of water molecules from non-polar and polar interaction surfaces [174, 175, 171]. In contrast, proteins that do

not bend dsDNA (bacteriophage λ cI repressor [176], GCN4 [177]) have a highly favourable ΔH, although this is counterbalanced by an unfavourable ΔS due to dsDNA stiffening [171].

Whether RAD51 also adopts similar strategies to stably polymerise on ssDNA in preference to dsDNA is still unknown. However, it is conceivable that the enthalpic contribution of RAD51 polymerisation may be finetuned to the entropic penalty of ssDNA to promote presynaptic filament formation during the initiation of HR.

1.5 Surface Plasmon Resonance

SPR biosensor technology enables real-time, label-free detection and monitoring of ensemble biomolecular binding events using small sample volumes and provides highly reproducible experimental results. Although first proposed as a biosensing technique in 1983 [178], it was not until the Swedish company Pharmacia AB introduced the first commercial SPR biosensor technology known as the Biacore [179, 180] that the technique became widely used to analyse biomolecular interactions due to its detection sensitivity and speed of data analysis. SPR is now a well established technique to analyse the kinetics and affinities of protein-protein interactions such as antibody-antigen interactions [181, 182, 183], interactions between peptide - major histocompatibility complexes (pMHCs) and T cell receptors [184, 185], and interactions between transmembrane proteins and ligands [186, 187, 188]. In the following section, I will describe the physical basis of SPR and how a typical SPR experiment is executed. I will then conclude by describing the potential experimental pitfalls associate with SPR.

1.5.1 The SPR Principle

SPR is a physical process that can occur when polarised light hits a thin metal film under total internal reflection conditions [189, 190, 191]. When polarised light hits

an interface of two transparent media, such as a glass prism and a buffer solution, from the side of the media with the highest refractive index (glass prism), the light is partly refracted and partly reflected towards the interface plane. Above a certain angle, all light is reflected and none of it is refracted. This phenomenon is known as total internal reflection (TIR) [191] (Fig. 1.8A).

In most SPR biosensors, the glass prism is coated with a gold film. The free conduction electrons in conducting metals such as gold form periodic oscillations called plasma waves [191]. The plasmon is the particle name for the plasma wave, similarly to how the photon is the particle name for a light wave. Importantly, surface plasmons are the plasmons that are confined to the surface of the gold film, occurring at the interface between the gold film and the buffer solution. These surface plasmons generate an electric field that, in the Biacore T200, extends $\approx$ 300 nm into the buffer solution [192]. This electrical field is known as an evanescent wave because it decays exponentially with distance [191].

Importantly, when the incoming polarised light exhibits the correct incidence angle within TIR, surface plasmon resonance occurs. At this resonance angle θ_1, the momentum of the photons in the polarised light is equal to the momentum of the surface plasmons, leading to the conversion of photons into plasmons [192]. This means that optical energy from the polarised light is coupled into the gold film, leading to a dip (i.e. minimum) in the intensity of the reflected light at the resonance angle θ_1 [192] (Fig. 1.8B). Importantly, the SPR angle changes in a manner dependent on the refractive index of the compartment on the gold film surface, often a dextran matrix [193]. The refractive index of the dextran matrix changes when, for example, proteins bind to ligands coupled to the dextran matrix. Consequently, the SPR angle shifts from θ_1 to θ_2, and this shift is proportional to the change in mass within the dextran matrix above the gold film (Fig. 1.8C). Binding of biomolecules is

followed in real-time by continuously monitoring the shift in the SPR angle. These angular shifts are measured in Response Units (RU), where 1 RU corresponds to a 0.0001° shift in the SPR angle, or the addition of 50 pg/mm^2 to the dextran matrix [194]. The curves describing the change in RU as a function of time are known as sensorgrams.

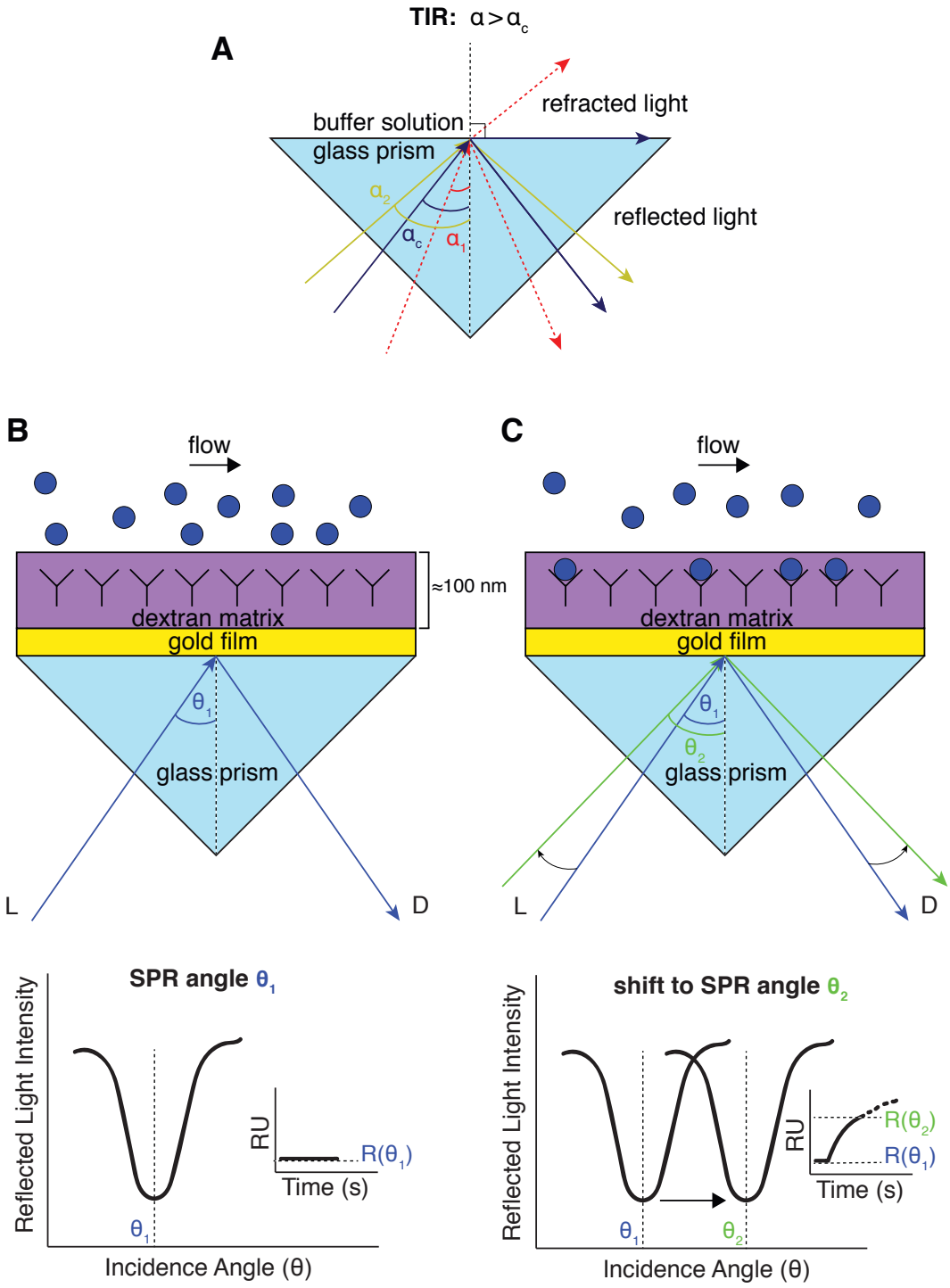


Figure 1.8: The physical basis of SPR. **A**. The principle of Total Internal Reflection (TIR). Polarised light at an incidence angle α_1 going from a dense medium (glass prism) to a less dense medium (buffer solution) is partially reflected and partially refracted. At the incidence angle α_c, the angle of refraction is 90°. At an incidence angle $\alpha_2 > \alpha_c$, TIR occurs and all light is reflected. **B**. Prior to analyte-ligand binding, the surface plasmons at the surface of the gold film resonate at the SPR angle θ_1. The angle is detected as the minimum in the plot of reflected light intensity versus incidence angle, and as response R(θ_1) on a sensorgram. **C**. Upon analyte-ligand binding, the refractive index change at the surface of the gold film leads to a shift in the SPR angle from θ_1 to θ_2 proportional to the addition of mass in the dextran matrix. This is recorded on the sensorgram as an increase in the RU from R(θ_1) to R(θ_2). **B-C**. The dextran matrix is $\approx$ 100 nm thick in the CM5 chips used throughout this thesis. L: light source. D: detector.

1.5.2 A Typical Experiment

As previously mentioned, SPR is an established technique to measure protein-protein interaction affinities and binding kinetics, the classical example being the interaction between an antibody and its cognate antigen [181].

In the simplest case, an antibody (ligand) is covalently coupled to the dextran matrix, and the cognate antigen (analyte) is injected over the antibody-coated matrix (Fig. 1.9A). Assuming the analyte is monovalent and homogenous, the ligand is homogenous, all binding events are independent and a one-to-one binding stoichiometry between the antibody and antigen is satisfied, then the Langmuir 1:1 binding model can be used to accurately fit the experimental sensorgram to describe the formation of the antibody-antigen complex as a function of time [195]:

$$\mathsf{L} + \mathsf{P} \underset{k_\mathsf{d}}{\overset{k_\mathsf{a}}{\rightleftharpoons}} \mathsf{LP} \tag{1.2}$$

Where L denotes unbound ligand, P denotes unbound peptide and LP denotes the ligand-analyte complex. The corresponding differential equation describing the formation of the ligand-analyte complex is as follows:

$$\frac{d[\mathsf{LP}]}{dt} = k_\mathsf{a}[\mathsf{L}][\mathsf{P}] - k_\mathsf{d}[\mathsf{LP}] \tag{1.3}$$

The concentration of P can be assumed to be constant as it is continuously replenished by the channel flow within the SPR machine ($[\mathsf{P}] = c$), while the total amount of free ligand sites can be described as the difference between the total number of ligand sites and the number of occupied ligand sites ($[\mathsf{L}] = \mathsf{L}_\mathsf{tot} - [\mathsf{LP}]$):

$$\frac{d[\mathsf{LP}]}{dt} = ck_\mathsf{a}(\mathsf{L}_\mathsf{tot} - [\mathsf{LP}]) - k_\mathsf{d}[\mathsf{LP}] \tag{1.4}$$

We can rewrite this to describe the real-time response change due to peptide binding:

$$\frac{d\mathsf{R}}{dt} = ck_{\mathsf{a}}(\mathsf{R}_{\mathsf{max}} - \mathsf{R}) - k_{\mathsf{d}}\mathsf{R} \tag{1.5}$$

where R is the response in RU ($\mathsf{R} \propto [\mathsf{LP}]$) and $\mathsf{R}_{\mathsf{max}}$ is the maximum theoretical achievable response ($\mathsf{R}_{\mathsf{max}} \propto \mathsf{L}_{\mathsf{tot}}$). This equation can be easily integrated to yield the association phase equation:

$$\mathsf{R}(t) = \frac{c\mathsf{R}_{\mathsf{max}}}{c + K_{\mathsf{D}}}(1 - e^{-(ck_{\mathsf{a}} + k_{\mathsf{d}})t}) \tag{1.6}$$

where the first term determines the total response at equilibrium ($K_{\mathsf{D}} = \frac{k_{\mathsf{d}}}{k_{\mathsf{a}}}$), while the second term determines the time to reach the equilibrium. Equation 1.6 includes both k_{a} and k_{d}, highlighting that both association and dissociation occur during the association phase.

The dissociation phase begins at the end of the association phase (i.e. at the end of the analyte injection). The dissociation phase equation can be derived from 1.5 by setting $c = 0$:

$$\mathsf{R}(t) = \mathsf{R}_0 e^{-k_{\mathsf{d}}t} \tag{1.7}$$

where R_0 is the response level at the end of the association phase. During the dissociation phase, only analyte dissociation occurs.

It is often useful to start out with equilibrium experiments to directly measure the K_{D}, as these experiments do not suffer from mass transport and analyte rebinding

effects (section 1.5.3) and enable to design kinetic experiments more effectively. When performing equilibrium experiments, only the first term in 1.6 is used:

$$\mathsf{R}_{\mathsf{eq}} = \frac{c\mathsf{R}_{\mathsf{max}}}{c + K_{\mathsf{D}}} \tag{1.8}$$

Where R_{eq} is the response at equilibrium. Importantly, $\mathsf{R}_{\mathsf{max}} = \mathsf{R}_{\mathsf{lig}}\mathsf{V}_{\mathsf{lig}}\frac{\mathsf{M}_{\mathsf{an}}}{\mathsf{M}_{\mathsf{lig}}}$, where $\mathsf{R}_{\mathsf{lig}}$ is the amount of immobilised ligand in response units, $\mathsf{V}_{\mathsf{lig}}$ is the ligand valency (in this case $\mathsf{V}_{\mathsf{lig}} = 1$), M_{an} is the molecular weight of the analyte and $\mathsf{M}_{\mathsf{lig}}$ is the molecular weight of the ligand. The experiment is repeated using a spectrum of analyte concentrations ($c = c_1, c_2, ..., c_n$), ideally spanning at least 2-3 orders of magnitude [191] (Fig. 1.9B). Equation 1.8 is then fit to R_{eq} data as a function of c to estimate the $K_{\mathsf{D}}^{c_n}$ for each sensorgram, and a mean K_{D} is then derived. Ideally, $K_{\mathsf{D}} \approx K_{\mathsf{D}}^{c_n}$, for $c = c_1, c_2, ..., c_n$.

Determination of binding kinetics also requires the experiment to be repeated using a spectrum of analyte concentrations, ideally between $0.1K_{\mathsf{D}}$ and 100 K_{D} [191]. The K_{D} value derived from equilibrium experiments can be used to determine the spectrum of analyte concentrations. When performing kinetic experiments, this allows for k_{d} to be fit to each dissociation phase using 1.7. The k_{a} is then fit individually to each association phase using 1.6 with the corresponding fit k_{d} value and the specific analyte injection concentration c. Assuming no mass transport or analyte rebinding effects are detected, mean k_{d} and k_{a} values are then calculated. Finally, an important validity test is to check that the K_{D} determined from equilibrium experiments is similar to that derived from the ratio of the k_{d} and k_{a} values determined from kinetic experiments ($K_{\mathsf{D}} = \frac{k_{\mathsf{d}}}{k_{\mathsf{a}}}$) [191].

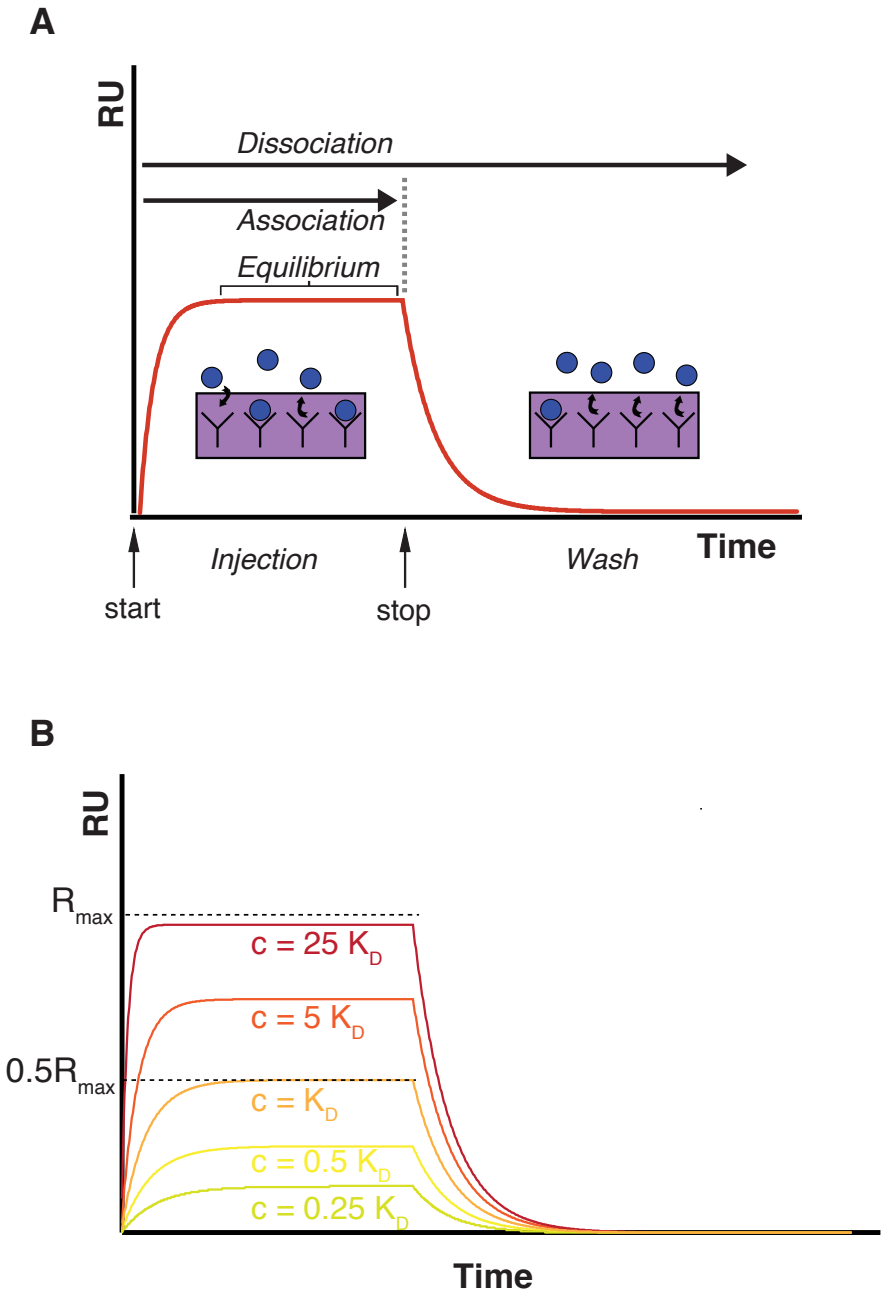


Figure 1.9: A typical SPR experiment. **A**. A ligand molecule is immobilised onto an SPR sensor chip. An analyte molecule is injected (start) over the ligand-coated SPR matrix to measure association kinetics (Injection). Equilibrium is reached when the rate of association is equal to the rate of dissociation (i.e. $ck_{\text{a}}(\text{R}_{\text{max}} - \text{R}) = k_{\text{d}}\text{R}$). Following analyte injection (stop), running buffer is injected to measure dissociation kinetics (Wash). During the injection, both RAD51 association and dissociation take place simultaneously. During the wash step, only dissociation occurs. **B**. In typical SPR experiments, the analyte concentration (c) is varied 2-3 orders of magnitude (2 orders of magnitude depicted). When $c \approx K_{\text{D}}$, the response reaches $\approx$ half of the maximum predicted response (0.5R_{max}).

1.5.3 Potential Pitfalls

The most common experimental pitfalls when measuring binding kinetics in SPR are the **mass-transport-limited binding** and **analyte rebinding** effects.

When a high amount of ligand is immobilised onto an SPR sensor matrix, the rate at which the analyte binds the ligand may exceed the rate at which it is delivered to the surface (referred to as the rate of mass transport) [191]. In this situation, ligand binding is considered mass-transport-limited, and fitting the k_{a} using 1.6 will lead to an apparent k_{a} that is lower than the true k_{a}. In particular, a binding experiment is transport limited if k_{a} is dependent on the flow rate through the matrix. Given that the analyte is transported to the sensor matrix via convection and diffusion, mass transport may be limited either by increasing the flow rate or by decreasing the amount of immobilised ligand. However, even at the maximum permissible flow rate (e.g. 100 μl/min) mass transport may still be limiting [196] due the presence of an unstirred "diffusion" layer near the sensor matrix through which transport occurs only via diffusion [197]. In these situations, mass transport limitation can only be reduced by reducing the amount of immobilised ligand. Nonetheless, mass transport prevents accurate measurement of k_{a} rates faster than the rate of diffusion ($\approx$ 100 /μM/s [198]).

A high amount of immobilised ligand may also lead to the analyte binding and dissociating multiple times from individual ligand molecules before being flushed out. This causes the apparent k_{d} derived from fitting the dissociation phase using 1.7 to be lower than the true k_{d} [191]. Importantly, a binding experiment suffers from rebinding effects if the k_{d} varies with analyte concentration. With increasing analyte concentrations, the apparent k_{d} will approach the true k_{d} because there are less free ligand binding sites for the analyte to re-bind to. Similarly, analyte rebinding may be limited by decreasing the amount of immobilised ligand, or by injecting a

competitor molecule at the beginning of the dissociation phase that sequesters the analyte after ligand dissociation and prevents rebinding [199]. However, accurate fitting of the k_{d} is difficult when the true analyte-ligand complex lifetime ($\frac{1}{k_{\mathrm{d}}}$) is higher than the timescale of the dissociation phase or when the true k_{d} is faster than the speed at which the SPR instrument can acquire data [191].

1.6 Fitting Unconventional SPR Data

Throughout this thesis, I address the following question: how does RAD51 discriminate between ssDNA and dsDNA? To tackle this question, I used SPR to measure real-time kinetics of RAD51 polymerisation on ssDNA and dsDNA oligos of different lengths and flexibilities.

SPR has been previously used to analyse protein-DNA interactions [200, 201, 202, 203] such as interactions between transcription factors and cognate DNA sequences [202, 203] and protein-protein interactions involving phosphatase reactions [204]. Importantly, SPR has also been used to study the binding kinetics of RecA [205, 140, 105] and RAD51 [140, 131, 124, 135] on DNA, although the quantitative analysis of SPR data describing protein polymerisation kinetics on DNA is not well established. Importantly, a mechanistic analysis of RAD51 polymerisation on DNA using SPR is challenging because RAD51 polymerises both in solution and on DNA, and pre-formed RAD51 polymers have been suggested to bind directly to DNA during nucleoprotein filament formation [12] (Fig. 1.10).

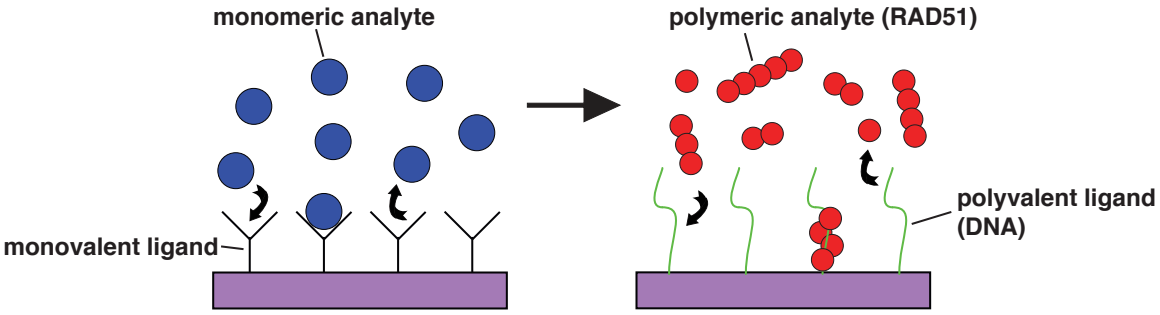


Figure 1.10: 1:1 interactions versus polymer-polymer interactions. Typical SPR experiments analyse binding kinetics between monovalent ligands (i.e. ligands that only have one analyte binding site) and monovalent, monomeric analytes (i.e. analytes that do not polymerise either in solution or when bound to the ligand molecule). These experiments are usually well described using the Langmuir model [195]. The analysis of binding kinetics between a polymeric analyte such as RAD51 and a polyvalent ligand such as DNA using SPR requires the development of a mathematical model that accounts for analyte polymerisation kinetics in solution and on the ligand molecule.

Under this assumption, the Langmuir model used to analyse 1:1 interactions [195] does not provide an accurate or mechanistic description of RAD51 polymerisation kinetics on DNA. To develop a mechanistic understanding of RAD51 polymerisation on DNA, I generated ordinary differential equation models that describe the assembly and disassembly of RAD51 polymers in solution, on ssDNA and on dsDNA as a function of time. These models enable me to infer the formation of RAD51 polymers of defined lengths on DNA based on the bulk RAD51-DNA binding curves from SPR. I then fit these models to the experimental data to analyse and interpret the differences in RAD51 polymerisation kinetics on ssDNA and dsDNA.

1.7 Summary of Thesis

In chapter 2, I describe the experimental reagents and methods used throughout the study. I also describe the mathematical modelling approach, and provide details of the model assumptions and equations.

In chapter 3, I present SPR data describing the RAD51 binding kinetics to ssDNA and dsDNA oligos of variable length. I then develop ODE models to fit the experimental data and derive kinetic parameters to describe the RAD51 polymeri-

sation steps on ssDNA and dsDNA. I suggest RAD51 polymer adsorption and/or elongation is faster on ssDNA compared to dsDNA, despite RAD51 requiring five molecules to form a stable nucleus on ssDNA versus two molecules on dsDNA.

In chapter 4, I present the Bend-To-Capture mechanism to describe how RAD51 preferentially associates on ssDNA compared to dsDNA. The mechanism suggests that the higher flexibility of ssDNA compared to dsDNA promotes faster polymerisation on ssDNA. I tested this by analysing the RAD51 binding kinetics to ssDNA molecules of variable flexibility using SPR and fitting the models from chapter 3 to these data. I conclude that RAD51 polymerises faster on ssDNA compared to dsDNA due to the higher flexibility of ssDNA.

In chapter 5, I present the Entropic Penalty Compensation mechanism to describe how RAD51 stably binds to flexible DNA. I hypothesised that the enthalpic contribution of RAD51 polymerisation enables RAD51 to overcome any entropic penalty to stably bind flexible DNA. To test this, I used SPR to analyse the RAD51 binding kinetics of a RAD51 mutant (F86E) with a reduced enthalpic contribution due to a reduced protomer-protomer affinity, and fit the models from chapter 3 to these data. I observed shorter F86E RAD51 lifetimes on flexible DNA compared to stiff DNA, and conclude that the enthalpic contribution of WT RAD51 polymerisation enables WT RAD51 to have long lifetimes on both flexible and stiff DNA.

In chapter 6, I summarise the results and discuss conclusions, alternative explanations of the SPR data and experimental caveats. I then compare and contrast the mathematical models developed in chapter 3 with other models in the literature. Finally, I provide ideas on how to expand the current project to better understand the impact of DNA flexibility on RAD51 polymerisation.

Chapter 2

Materials and Methods

2.1 RAD51 Molecular Cloning

The bacterial expression vector pET11d carrying the human WT RAD51 (WT-pET11d) was used as a template for the PCR-mediated QuikChange site-directed mutagenesis to introduce the F86E substitution (F86E-pET11d) with a forward primer 5'- GCTAAATTAGTTCCAAT GGGTGAGACCACTGCAACTGAATTCCACC - 3' and a reverse primer 5'- GGTGGAATT CAGTTGCAGTGGTCTCACCCATTGGAACTAATT-TAGC - 3'.

2.2 RAD51 Protein Purification

To express RAD51, Rosetta 2 (DE3) pLysS cells (Novagen) carrying the WT-pET11d or F86E-pET11d were grown in LB media containing 100 µg/ml ampicillin and 25 µg/ml chloroamphenicol, and by adding 0.5 mM IPTG at OD595 = 0.6, protein expression was induced. Cell pellets were resuspended in PBS and mixed with an equal volume of lysis buffer (3 M NaCl, 100 mM Tris-HCl pH 7.5, 4 mM EDTA pH 8, 20 mM ß-mercaptoethanol, Protease Inhibitor Cocktail (Sigma)). The suspension was sonicated and spun at 20 k rpm with a 45 Ti rotor (Beckman) for 30 minutes. The supernatant was slowly mixed with 0.1 % polyethylenimine at 4°C for 1 hour, and spun at 20 k rpm using a 45 Ti rotor for 30 minutes to remove DNA. The supernatant was then slowly mixed with an equal volume of 4 M ammonium sulphate (2

M final concentration) at 4°C for 1 hour, and spun at 10 k rpm using a JA-17 rotor (Beckman) for 30 minutes. Pellets containing RAD51 were suspended in 25 ml of resuspension buffer (0.5 M KCl, 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 mM DTT, 10 % glycerol) and spun again at 20 k rpm for 30 minutes to remove residual DNA. The supernatant was then dialysed overnight at 4°C in TEG buffer (50 mM Tris pH 7.5, 1 mM EDTA, 2 mM DTT, 10 % glycerol) containing 200 mM KCl (TEG200). For F86E RAD51, cell lysate was prepared as WT RAD51, and dialysed overnight at 4°C in TEG buffer containing 50 mM KCl (TEG75).

RAD51 purification was carried out by chromatography at 4°C using the AKTA Pure Protein Purification System (GE Healthcare). For WT RAD51, the dialysed WT RAD51 containing sample was loaded onto a 5 ml HiTrap Heparin column (GE Healthcare), and eluted via a linear gradient of 200 mM – 600 mM KCl. The peak fractions were pooled and dialysed in TEG200, and concentrated in 1 ml HiTrap Q column (GE Healthcare) followed by an isocratic elution with TEG buffer containing 600 mM KCl (TEG600). No size-exclusion chromatography was carried out on WT RAD51. This simplification of the purification protocol has been widely accepted for biochemical analysis of RAD51 (e.g. [140, 137, 206]) and, importantly, non-specific RAD51 binding to the SPR chip flow cells (which is often caused by protein aggregates) was not detected. Peak fractions containing WT RAD51 were dialysed in SPR buffer (150 mM KCl, 20 mM Hepes pH 7.5, 2 mM DTT, 10 % glycerol), aliquoted, snap frozen and stored at -80°C (final yield: $\approx$ 4.5 mg, $\approx$ 3 mg/ml). The dialysed F86E RAD51 containing cell lysate was fractionated via a 5 ml HiTrap Heparin column, but with a linear gradient of 75 mM – 600 mM KCl. The flow-through and F86E RAD51 peak fractions were pooled and dialysed in TEG buffer containing 50 mM KCl (TEG50). The column was flushed with 1 M KCl to remove high molecular weight contaminant proteins bound to the column, and the F86E-containing solution was reloaded onto the 5 ml HiTrap Heparin column. This

enabled to improve final protein purity. Following a 50 mM – 600 mM KCl linear gradient elution, F86E RAD51 peak fractions were pooled and dialysed in TEG buffer containing 100 mM KCl (TEG100). The sample was concentrated in a 1 ml HiTrap Q column followed by a TEG600 isocratic elution. Peak fractions were applied on a 24 ml Superdex200 10/300 GL size-exclusion column (GE Healthcare), and fractions containing monomeric F86E RAD51 were pooled and dialysed in TEG100. F86E RAD51 was then reconcentrated via a TEG600 isocratic elution using a 1 ml HiTrap Q column. Peak fractions were dialysed in SPR buffer, aliquoted, snap frozen and stored at -80°C (final yield: $\approx$ 0.9 mg, $\approx$ 1.4 mg/ml).

2.3 Multi-Angle Light Scattering

To confirm the monomeric status of F86E RAD51, the peak size-exclusion chromatography F86E RAD51 elution fraction was serially diluted (5, 1:1 serial dilutions) and loaded onto a Superdex200 10/300 GL size exclusion column (GE Healthcare) equilibrated with MALS Buffer (150 mM KCl, 50 mM Tris pH 7.5, 1 mM EDTA, 2 mM DTT). Each elution was analysed using a Wyatt Heleos8+ 8-angle light scatterer linked to a Shimadzu HPLC system comprising LC-20AD pump, SIL-20A Autosampler and SPD20A UV/Vis detector. Data collection was carried out by Dr David Staunton (Biophysics Facility Manager - Department of Biochemistry, University of Oxford). Data analysis was carried out using the ASTRA software (Wyatt).

2.4 DNA Oligo and Duplex Synthesis

The dsDNA dN-5p, dN-8p and dN-11p sequences were designed as two complimentary ssDNA sequences connected via two units of hexaethylene glycol (flexible linker) (Fig. 2.1) to minimise DNA melting at 37°C (i.e. the temperature at which the SPR experiments were carried out - section 2.6) [207, 208]. The ssDNA dN-5, dN-8, dN-11 molecules were designed using one of the two corresponding dsDNA

annealing sequences. The ssDNA dN-14 and dN-17 molecules were designed by extending the dN-11 sequence. Oligonucleotide purification and characterisation were carried out by Dr Afaf El-Sagheer and Professor Tom Brown (Department of Physical and Theoretical Chemistry, University of Oxford). Oligonucleotides were freeze dried and resuspended in water without the need for desalting.

The biotinylated ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p molecules were ordered as lyophilised samples from Integrated DNA Technologies (IDT) and re-suspended in TE buffer (20 mM Tris pH 8.0, 2 mM EDTA). The two complementary strands of the dsDNA molecule were annealed by IDT before lyophilisation. The DNA molecules used in this study are summarised in Table 2.1.

Table 2.1: Names, masses and sequences of the DNA molecules used throughout the study. **L**: hexaethylene glycol flexible linker. Biotin TEG: Biotin linked to a flexible triethylene glycol spacer. Biotin: Biotin linked to a standard C6 spacer. dN: mixed base composition.

Molecule Name	Mass (kDa)	Sequence
ssDNA dN-5	2.057	5'-CGGAC-Biotin TEG-3'
dsDNA dN-5p	4.286	5'-CGGAC **LL** GTCCG-Biotin TEG-3'
ssDNA dN-8	2.594	5'-CTGACTGC- Biotin TEG-3'
dsDNA dN-8p	6.139	5'-CTGACTGC **LL** GCAGTCAG-Biotin TEG-3'
ssDNA dN-11	3.926	5'-CGTCGATAGGC-Biotin TEG-3'
dsDNA dN-11p	7.993	5'-CGTCGATAGGC **LL** GCCTATCGACG-Biotin TEG-3'
ssDNA dN-14	4.872	5'-CGTCGATGAGCAGT-Biotin TEG-3'
ssDNA dN-17	5.795	5'-CGTCGATGAGCAGTGTC-Biotin TEG-3'
ssDNA dN-50	15.682	5'-Biotin-TTGAGAGAGCAGACCACAATTATCA CCTACACGACATCATTTTATATCAA-3'
dsDNA dN-50p F	31.157	5'-Biotin-TCGAGAGGGTAAACCACAATTATCG CCTACCCAAAACTATTTTATATCAA-3'
dsDNA dN-50p R	(Reverse)	5'-TTGATATAAAATAGTTTTGGGTAGG CGATAATTGTGGTTTACCCTCTCGA-3'
ssDNA dT-50	15.541	5'-Biotin-TTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTT-3'
ssDNA dA-50	15.992	5'-Biotin-AAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAA-3'

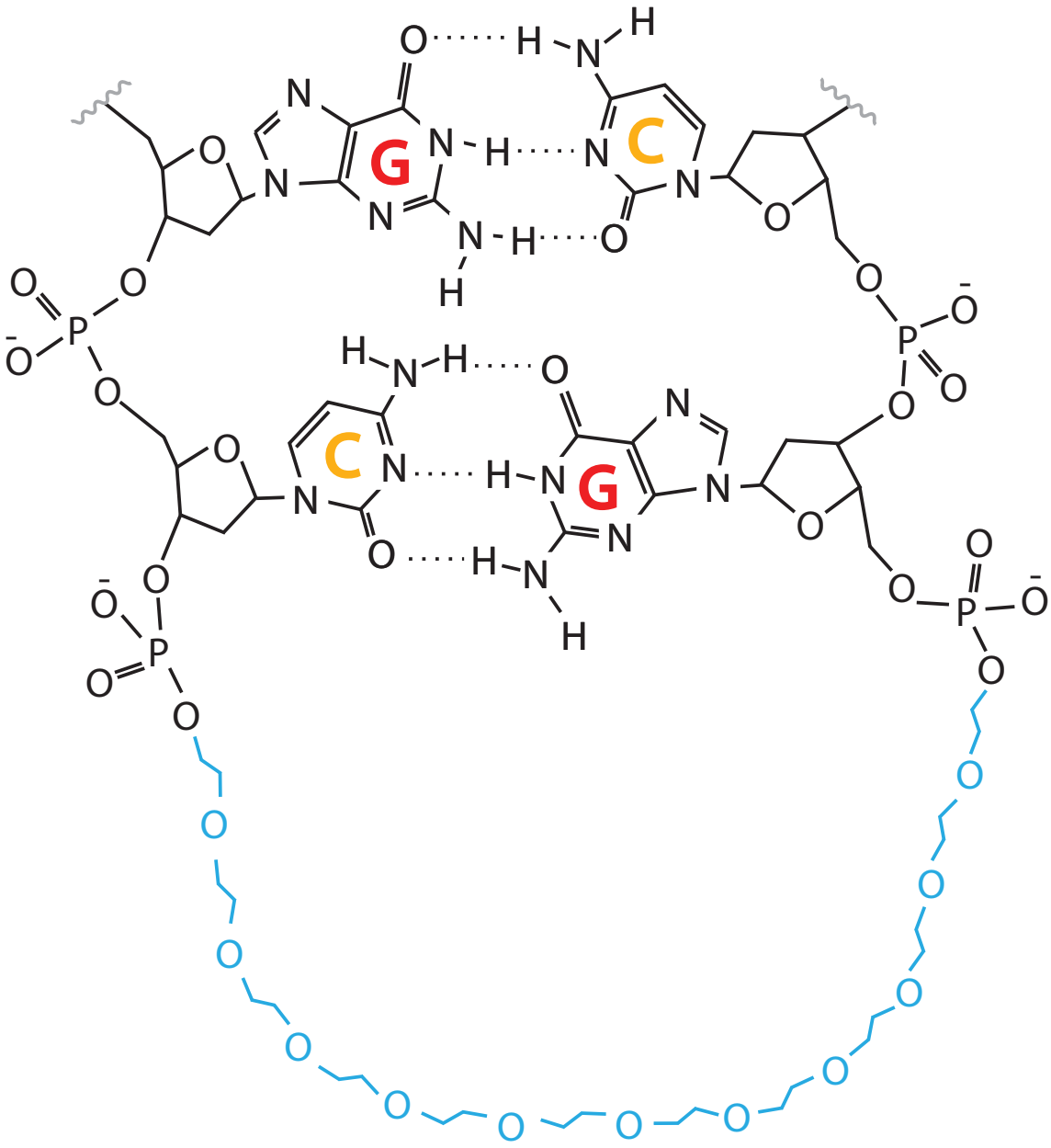


Figure 2.1: Chemical diagram of the dN-11p double-hexaethylene-glycol-linked oligo. The double hexaethylene glycol (HEG) linker (highlighted in cyan) connects the opposite strands to form a hairpin structure. The first two base pairs (Cytosine: orange; Guanine: red) starting from the double HEG linker are shown for simplicity. The structure of the double HEG linker in dN-11p is analogous to that of dN-5p and dN-8p.

2.5 Small-Angle X-Ray Scattering

The flexibilities of the ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p DNA molecules were assessed using small-angle x-ray scattering (SAXS) at the Diamond Light Source (Harwell, UK). SAXS data was collected using the size exclusion KW 402.5 (2.4 ml) column (Shodex). 50 µl of DNA sample was injected and elution was carried out at 37°C at 75 µl/min using SAXS buffer (150 mM KCl, 20 mM Hepes pH 7.5, 2 mM DTT, 2.5 mM ATP pH 7.5, 10 mM $CaCl_2$). Data collection was carried out by Dr Robert Rambo (principal beamline scientist - Diamond Light Source). The flexible cylinder model was fit to the four scattering data sets within the 0.0037 - 0.27 [1/Å] range to derive the persistence lengths. Data plotting and fitting was carried out using the SasView software.

2.6 Surface Plasmon Resonance

The binding kinetics of WT and F86E RAD51 on ssDNA and dsDNA were assessed via SPR using a Biacore T200 (GE Healthcare). CM5 SPR chip flow cells were activated by injecting 100 µl of a 1:1 N-Hydroxysuccinimide (NHS), ethyl (dimethylaminopropyl) carbodiimide (EDC) mix at 10 µl/min. 100 µl of purified streptavidin (1 mg/ml) was then injected over the flow cells at 10 µl/min. Unbound amine groups were then deactivated by injecting 50 µl of 1 M ethanolamine-HCl. 50 µl of 10 mM glycine pH 2.5 was injected to remove uncoupled, sterically bound streptavidin. For all experiments, $\approx$ 8 - 15 RU of biotinylated DNA substrates were immobilised onto either flow cells 2, 3 or 4. After DNA immobilisation, 10 µl of purified biotin was injected over all four flow cells at 10 µl/min to block the remaining free streptavidin binding sites. Chip activation and ligand immobilisation steps were carried out at 25°C. Using high performance injections, the CM5 chip surfaces were then primed via 10 injections of 10 µl SPR running buffer (150 mM KCl, 20 mM Hepes pH 7.5, 2 mM DTT, 5 mg/ml BSA, 2.5 mM ATP pH 7.5, 10 mM $CaCl_2$, 10 % glycerol) at

30 µl/min at 37°C. RAD51 binds to 3 nucleotides of ssDNA [113] and the predicted maximum number of RAD51 molecules bound to a ssDNA 50mer is 16. However, with a sufficiently long SPR injection RAD51 can form polymers longer than length 16 on a ssDNA 50mer (Fig. 2.2), suggesting the possibility of additional RAD51 polymer elongation off DNA. To avoid this, the RAD51 SPR injection time was limited to 60 seconds (Fig. 2.2). Hence, WT or F86E RAD51 diluted in SPR running buffer to the specified concentration was injected over the DNA-bound and DNA-free flow cells (negative control flow cells) at 30 µl/min for 60 seconds (association phase), followed by the injection of SPR running buffer for 20 minutes at 30 µl/min (dissociation phase) at 37°C. To maximise data reproducibility, CM5 chip surfaces were not regenerated and fresh CM5 chips were used for each experiment.

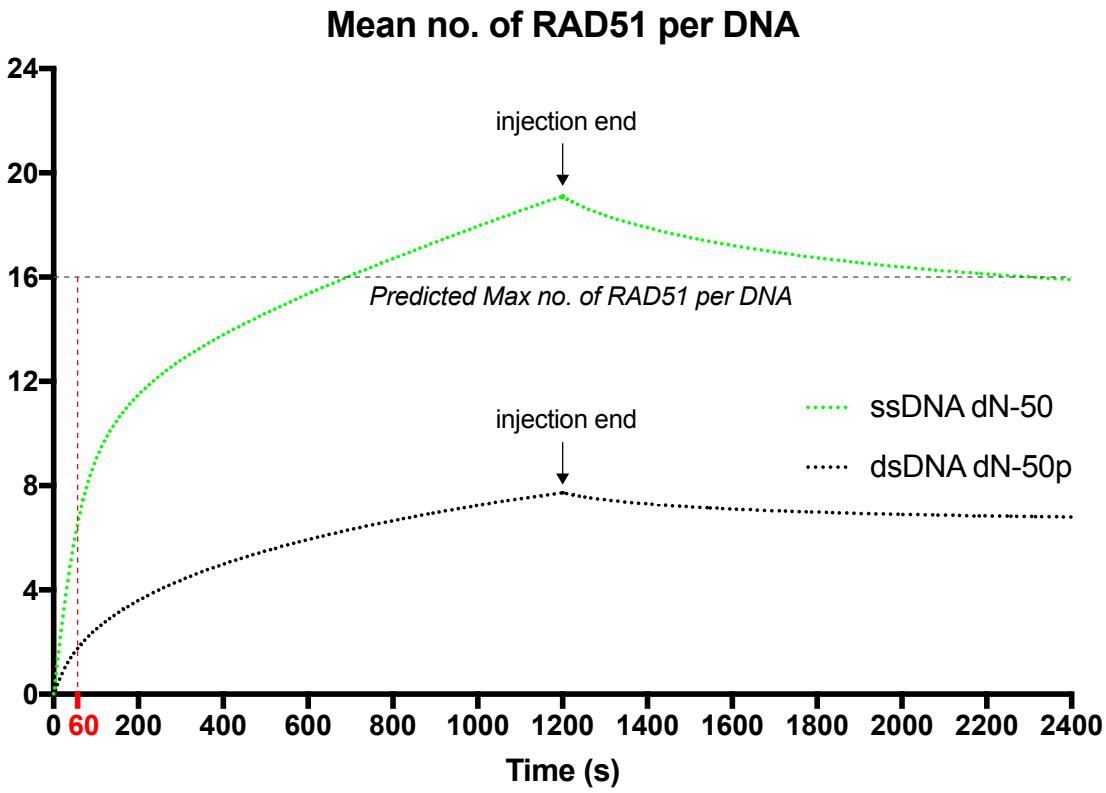


Figure 2.2: RAD51 polymers can potentially elongate off DNA. WT RAD51 SPR curves for ssDNA dN-50 and dsDNA dN-50p. RAD51 was injected at 150 nM for 20 minutes in the presence of 2.5 mM ATP and 10 mM $CaCl_2$. The black dotted line indicates the predicted maximum number of RAD51 molecules per ssDNA dN-50 or dsDNA dN-50p (16), while the red dotted line indicates the RAD51 SPR injection time chosen for all successive experiments (60 seconds). All curves were processed and normalised according to section 2.7.

2.7 SPR Data Processing

SPR traces were processed using the BIAEvaluation software (GE) as following. First, the negative control flow cell trace (Fc1 or Fc3) was subtracted from the experimental flow cell trace (Fc2 or Fc4). The trace was then vertically and horizontally aligned, such that the start of the protein injection occurs at time = 0 s, Response = 0 RU. Subsequently, the drift trace between the negative control flow cell and the experimental flow cell during the final priming injection was subtracted from the experimental flow cell trace (double-referencing) [209]. All experimental curves were normalised using the following equation:

$$\mathbf{N} = \frac{\mathbf{S}}{L(\frac{M_{\text{RAD51}}}{M_{\text{DNA}}})} \tag{2.1}$$

Where **S** is the signal in RU units (1 RU $\approx 50\frac{\text{pg}}{\text{mm}^2}$), **N** is the normalised signal, L is the amount of DNA ligand immobilised onto the experimental flow cell (RU), M_{RAD51} is the molecular weight (kDa) of RAD51 ($\approx$ 37 kDa) and M_{DNA} is the molecular weight (kDa) of the immobilised DNA molecule. The normalised signal represents the mean number of RAD51 molecules bound to an immobilised DNA molecule, and has a predicted maximum at 16 (i.e. the predicted maximum number of RAD51 molecules bound to a ssDNA or dsDNA 50mer).

2.8 ABC-SMC

Approximate Bayesian Computation (ABC) methods are used in statistical inference to estimate posterior distributions of model parameters without the requirement of a likelihood function. In particular, they replace the calculation of the likelihood function with a comparison between observed and simulated data [210].

The simplest ABC method is known as the **ABC rejection sampler**. Let β be a vector of k parameters (i.e. a particle) to be estimated, and $\pi(\beta)$ be its prior distribution. The objective is to estimate the posterior distribution $\pi(\beta|x) \propto f(x|\beta)\pi(\beta)$, where $f(x|\beta)$ is the likelihood of β given the data x. The algorithm proceeds as follows:

1. sample a paricle β^* from $\pi(\beta)$
2. simulate a data set x^* via a proposed model
3. compare the simulated data set x^* and the experimental data set x_0 using a distance function d and a tolerance threshold ϵ. If $d(x_0, x^*) \leq \epsilon$, accept β^*, otherwise reject.
4. repeat 1 - 3 until N particles have been accepted

A common distance function is the sum of squared residuals (SSR), although distance functions based on other summary statistics are also used [210].

The output of the ABC rejection sampler is a sample of particles from the distribution $\pi(\beta|d(x_0, x^* \leq \epsilon)$. If ϵ is sufficiently small, then $\pi(\beta|d(x_0, x^* \leq \epsilon)$ will be a good approximation of the posterior distribution.

The main problem with the ABC rejection sampler is that the particle acceptance rate is low when the prior distribution is very different to the posterior distribution.

This leads to a high proportion of sampled particles being rejected, for which a lot of computational effort has been spent for simulation. ABC coupled to Sequential Monte Carlo (ABC-SMC) can be used to reduce this computational cost.

In **ABC-SMC**, a population of particles ($\beta^1, ..., \beta^i, ...\beta^N$) sampled from the prior distribution $\pi(\beta)$ are propagated through a series of intermediate distributions $\pi(\beta|d(x_0, x^* \leq \epsilon_t)$, $t = 1, ..., T-1$, until that population represents a sample from the target distribution $\pi(\beta|d(x_0, x^* \leq \epsilon_T)$ [210]. Importantly, the tolerances are chosen such that $\epsilon_1 > ..., > \epsilon_T$, so that the sampled distribution gradually evolves towards the target posterior distribution. ABC-SMC allows to both estimate posterior distributions of fit parameters in multi-parametric models as well as determine whether or not fit parameters are well determined [210].

In this thesis, I have used a quantile-based method to determine the tolerance ϵ_t for particle population t given the frequency distribution of the particle distances for particle population $t-1$ [211, 212]. This method determines ϵ_t at the beginning of the t-th iteration by sorting the distances $[d(x_0, x^{*(i,t-1)})]_{1 \leq i \leq N}$ and setting ϵ_t such that α percent of the distances are below ϵ_t, for a constant α. The ABC-SMC method is increasingly being used because it is less prone to local minima compared to more traditional methods, such as the Levenberg-Marquart algorithm as implemented in lsqcurvefit in MATLAB. However, it has been suggested that even this improved algorithm can overlook global minima in more complicated distance landscapes and cutting edge methods are continously being developed to further improve ABC-SMC [213].

The ABC-SMC algorithm I have used throughout this thesis proceeds as follows:

1. set $\alpha = 50\%$ (suggested in [213])

2. sample N particles from $\pi(\beta)$ to generate population $t = 0$, where $\pi(\beta)$ is a log-uniform distribution $\mathcal{LU}[a_k, b_k]$ for each component k of β, such that the lower bound is 10^{a_k} and the upper bound is 10^{b_k} for each component k of β. Calculate particle weights according to 2.2

3. calculate ϵ_1, sample particles from population $t = 0$ until N particles have been accepted based on $d(x_0, x^*) \leq \epsilon_1$. This is population $t = 1$. For all accepted particles, calculate updated weights according to equation 2.2 (below)

4. calculate ϵ_2, sample a particle β^* with associated weight w^* from population $t = 1$ and perturb each component k independently to obtain β^{**} using a component-wise Gaussian random walk perturbation kernel ($\beta_k^{**} \sim \mathcal{N}(\beta_k^*, \sigma_k^{(t=1)^2})$, where $\sigma_k^{(t=1)^2}$ is the variance of the k-th component of β^* for population $t = 1$ [214]). Simulate data set x^{**} using β^{**} and accept β^{**} if $d(x_0, x^{**}) \leq \epsilon_2$. Repeat sampling until N particles have been accepted. This is population $t = 2$. Update weights for all accepted particles according to equation 2.2

5. Repeat 4. for $\epsilon_3 ... \epsilon_T$ until N particles have been accepted from population $t = T$ and their weights have been calculated. Population T is a sample of particles that approximates the posterior distribution $\pi(\beta|x_0)$

The equation to calculate the weight $w^{(i,t)}$ for particle $\beta^{(i,t)}$ within population t is as follows [210]:

$$w^{(i,t)} = \begin{cases} 1, & t = 0 \\ \frac{\pi(\beta^{(i,t)})}{\sum_{j=1}^{N} w^{(j,t-1)} K_t(\beta^{(i,t)}|\beta^{(j,t-1)})}, & t > 0 \end{cases} \tag{2.2}$$

where $\pi(\beta^{(i,t)})$ is the prior probability of $\beta^{(i,t)}$ and K_t is the transition density for particles $\beta^{(i,t)}$ and $\beta^{(j,t-1)}$ calculated using a component-wise Gaussian random walk

perturbation kernel [215]. All particle weights are then normalised by dividing by $\sum_{i=1}^{N} w^{(i,t)}$.

The weighting of particles ensure that particles that had a low probability of sampling yet produced a good fit (small distance measure) in population $t-1$ receive a larger weight so that they are sampled more frequently in population t. This allows the ABC-SMC algorithm to hone in on regions of parameter space that are likely to produce an improved fit.

2.9 Dimer Models

During SPR injections, the analyte concentration (RAD51) remains constant because it is continually replenished by flow. This generally simplifies mathematical models of binding in SPR because it allows for binding to be described using 1st rather than 2nd order kinetics.

In the case of the dimer models, ordinary-differential-equation (ODE) models describing describing RAD51 polymerisation on ssDNA and dsDNA were formulated using first order mass action kinetics via the rule-based modelling language BioNetGen [216], and each consist of 17 linear ODEs. Each ODE describes the rate of change in mean concentration of a DNA-bound RAD51 n-mer on the SPR chip with respect to time, where $1 \leq n \leq b$ and b is the maximum number of RAD51 binding sites on the DNA molecule (i. e. $b = \frac{\text{number of nt or bp}}{3}$ [113]) (Table 2.2).

Table 2.2: DNA molecule names and lengths, with corresponding maximum RAD51 polymer lengths (b) and dimer model (DM) ODEs used for SPR data fitting. SPR data obtained using the ssDNA dN-5 and dN-11 oligos was not included in the data fitting because they were identical to the data obtained using the ssDNA dN-8 oligo and described no RAD51 binding to these oligos. NA: not applicable.

Molecule Name	Length (nt or bp)	b	DM ODEs
ssDNA dN-5	5	1	NA
dsDNA dN-5p	5	1	2.11 - 2.12
ssDNA dN-8	8	2	2.3 - 2.5
dsDNA dN-8p	8	2	2.11 - 2.13
ssDNA dN-11	11	3	NA
dsDNA dN-11p	11	3	2.11 - 2.14
ssDNA dN-14	14	4	2.3 - 2.7
ssDNA dN-17	17	5	2.3 - 2.8
ssDNA dN-50	50	16	2.3 - 2.10
ssDNA dT-50	50	16	2.3 - 2.10
ssDNA dA-50	50	16	2.3 - 2.10
dsDNA dN-50p	50	16	2.11 - 2.16

2.9.1 Assumptions

For both the ssDNA and dsDNA dimer models, I assume

- RAD51 polymerisation occurs mainly via the addition of dimers, and to a lesser extend via the addition of monomers [217] (i.e. the RAD51-RAD51 dissociation constant, $\mathbf{K_D}$, is much lower than the lowest RAD51 concentration in the SPR experiments). Specifically, I assume R_1 = 0.01(RAD51 injection concentration) and $R_2 = \frac{\text{(RAD51 injection concentration} - R_1)}{2}$, where R_1 is the concentration of RAD51 monomers in solution and R_2 is the concentrations of dimers in solution

- RAD51 dissociation occurs via the removal of single RAD51 protomers. This is a valid assumption when considering that in all the SPR experiments in this

study ATP hydrolysis is inhibited via the addition of Ca^{2+}, leading to very slow RAD51 filament disassembly which can be approximated as disassembly via monomer removal

- Each ssDNA or dsDNA molecule can accommodate a RAD51 polymer of maximum length b, where b is the maximum number of RAD51 binding sites on the DNA molecule (Table 2.2)

- RAD51 only nucleates once per DNA molecule (i.e. each DNA molecule only accommodates a single polymer of length n, where $1 \leq n \leq b$). In other words, polymer elongation is faster than nucleation. This is a valid assumption given that the longest ssDNA and dsDNA molecules used in this study are 50 nt and 50 bp long respectively, and RAD51 nucleation was shown to occur on average once every 500 bp in reactions with a 1 RAD51 : 3 bp molar ratio [137]. In the case of the experiments using WT RAD51 and the ssDNA dN-50 or dsDNA dN-50p, I used a 150 nM RAD51 concentration. Considering an average 10 RU of immobilised dsDNA and a dN-50p molecular weight of 31.157 kDa, the molar ratio in my experimental system is far lower compared to that in [137], and hence nucleation is expected on average to occur only once on a DNA molecule: [bps] $= 50(\frac{\mathsf{R_{lig}}}{100\mathsf{M_{lig}}})[218] = 50(\frac{10\mathsf{RU}}{(100)31157\mathsf{Da}}) \approx 160\ \mu$M. Hence $\frac{150\text{nM RAD51}}{160\mu\text{M bps}} \Rightarrow$ 1 RAD51 molecule per $\approx$ 1100 bps on the SPR chip.

2.9.2 Description of ODEs

The reaction network for the ssDNA dimer model was generated using BioNetGen [216] (section A.1). The resulting ODEs of the ssDNA dimer model are as follows:

$$\frac{dS_0}{dt} = S_1\mathbf{k_{ad.r}} - bS_0R_1\mathbf{k_{ad.f}} - 2(b-1)S_0R_2\mathbf{k_{ad.f}} \tag{2.3}$$

$$\frac{dS_1}{dt} = S_2\mathbf{k_{ad.r}} + bS_0R_1\mathbf{k_{ad.f}} - S_1\mathbf{k_{ad.r}} - S_1R_1\mathbf{k_{ad.f}} - S_1R_2\mathbf{k_{ad.f}} \tag{2.4}$$

$$\frac{dS_2}{dt} = S_3\mathbf{k_{ad.r}} + S_1R_1\mathbf{k_{ad.f}} + 2(b-1)S_0R_2\mathbf{k_{ad.f}} - S_2\mathbf{k_{ad.r}} - S_2R_1\mathbf{k_{ad.f}} - S_2R_2\mathbf{k_{el.f}} \tag{2.5}$$

$$\frac{dS_3}{dt} = S_4\mathbf{k_{el.r}} + S_2R_1\mathbf{k_{ad.f}} + S_1R_2\mathbf{k_{ad.f}} - S_3\mathbf{k_{ad.r}} - S_3R_1\mathbf{k_{el.f}} - S_3R_2\mathbf{k_{el.f}} \tag{2.6}$$

$$\frac{dS_4}{dt} = S_5\mathbf{k_s} + S_3R_1\mathbf{k_{el.f}} + S_2R_2\mathbf{k_{el.f}} - S_4\mathbf{k_{el.r}} - S_4R_1\mathbf{k_{el.f}} - S_4R_2\mathbf{k_{el.f}} \tag{2.7}$$

$$\frac{dS_5}{dt} = S_6\mathbf{k_s} + S_4R_1\mathbf{k_{el.f}} + S_3R_2\mathbf{k_{el.f}} - S_5\mathbf{k_s} - S_5R_1\mathbf{k_{el.f}} - S_5R_2\mathbf{k_{el.f}} \tag{2.8}$$

for $5 < n < 16$:

$$\frac{dS_n}{dt} = S_{n+1}\mathbf{k_s} + S_{n-1}R_1\mathbf{k_{el.f}} + S_{n-2}R_2\mathbf{k_{el.f}} - S_n\mathbf{k_s} - S_nR_1\mathbf{k_{el.f}} - S_nR_2\mathbf{k_{el.f}} \tag{2.9}$$

$$\vdots$$

$$\frac{dS_{16}}{dt} = S_{15}R_1\mathbf{k_{el.f}} + S_{14}R_2\mathbf{k_{el.f}} - S_{16}\mathbf{k_s} \tag{2.10}$$

with initial conditions $S_0(0) = 1, S_{1,\ldots,16}(0) = 0$.

The reaction network for the dsDNA dimer model was generated using BioNetGen [216] (section A.2). The resulting ODEs of the dsDNA dimer model are as follows:

$$\frac{dD_0}{dt} = D_1\mathbf{k_{ad.r}} - bD_0R_1\mathbf{k_{ad.f}} - 2(b-1)D_0R_2\mathbf{k_{ad.f}} \tag{2.11}$$

$$\frac{dD_1}{dt} = D_2\mathbf{k_s} + bD_0R_1\mathbf{k_{ad.f}} - D_1\mathbf{k_{ad.r}} - D_1R_1\mathbf{k_{el.f}} - D_1R_2\mathbf{k_{el.f}} \tag{2.12}$$

$$\frac{dD_2}{dt} = D_3\mathbf{k_s} + D_1R_1\mathbf{k_{el.f}} + 2(b-1)D_0R_2\mathbf{k_{ad.f}} - D_2\mathbf{k_s} - D_2R_1\mathbf{k_{el.f}} - D_2R_2\mathbf{k_{el.f}} \tag{2.13}$$

$$\frac{dD_3}{dt} = D_4\mathbf{k_s} + D_2R_1\mathbf{k_{el.f}} + D_1R_2\mathbf{k_{el.f}} - D_3\mathbf{k_s} - D_3R_1\mathbf{k_{el.f}} - D_3R_2\mathbf{k_{el.f}} \tag{2.14}$$

for $3 < n < 16$:

$$\frac{dD_n}{dt} = D_{n+1}\mathbf{k_s} + D_{n-1}R_1\mathbf{k_{el.f}} + D_{n-2}R_2\mathbf{k_{el.f}} - D_n\mathbf{k_s} - D_nR_1\mathbf{k_{el.f}} - D_nR_2\mathbf{k_{el.f}} \tag{2.15}$$

$$\vdots$$

$$\frac{dD_{16}}{dt} = D_{15}R_1\mathbf{k_{el.f}} + D_{14}R_2\mathbf{k_{el.f}} - D_{16}\mathbf{k_s} \tag{2.16}$$

with initial conditions $D_0(0) = 1, D_{1,...,16}(0) = 0$.

S_n and D_n ($n = 1, ..., 16$) represent the time-dependent concentration of ssDNA-bound and dsDNA-bound RAD51 n-mers on the SPR chip, respectively. S_0 and D_0 represent the concentration of unbound ssDNA oligos and dsDNA oligos on the SPR chip, respectively. The first RAD51 monomer adsorption reaction can occur on any of the DNA binding sites, and therefore the terms $S_0R_1\mathbf{k_{ad.f}}$ (equations 2.3, 2.4) and $D_0R_1\mathbf{k_{ad.f}}$ (equations 2.11, 2.12) are multiplied by b (Table 2.2). Similarly, the first dimer adsorption reaction can occur on any pair of DNA binding sites ($b-1$) and can initiate via either of the two protomers first binding to the DNA binding site, and therefore the terms $S_0R_2\mathbf{k_{ad.f}}$ (equations 2.3, 2.5) and $D_0R_2\mathbf{k_{ad.f}}$ (equations 2.11, 2.13) are multiplied by $2(b-1)$. R_1 and R_2 are constants that represent the concentration of RAD51 monomers and dimers, respectively. Given the assumption of RAD51 polymerisation mainly via dimer addition, I assume R_1

= 0.01(RAD51 injection concentration) and $R_2 = \frac{(\text{RAD51 injection concentration} - R_1)}{2}$. At the end of the RAD51 SPR injection, R_1 and R_2 are set to 0. $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_{el.r}}$ and $\mathbf{k_s}$ are the kinetic rate constants describing the dynamics of the ODE models. $\mathbf{k_{ad.f}}$ is the adsorption forward rate constant, $\mathbf{k_{ad.r}}$ is adsorption reverse rate constant, $\mathbf{k_{el.f}}$ is the elongation forward rate constant, $\mathbf{k_{el.r}}$ is the elongation reverse rate constant and $\mathbf{k_s}$ is the stabilisation rate constant. The values of these rate constants are obtained by fitting the ODE model outputs ($\sum_1^b iS_i$ and $\sum_1^b iD_i$) to their respective data sets (Table 2.2) either via ABC-SMC (Figs. 3.4A-B, 3.5A-B) or least-squares fitting (Fig. 4.5B-D). The initial conditions $S_0(0)$ and $D_0(0)$ were set to 1 to match the SPR data normalisation, such that $(\text{max} \sum_1^{16} iS_i) = (\text{max} \sum_1^{16} iD_i) = 16$ (section 2.7).

2.9.3 Model Selection

For both ssDNA and dsDNA, the sum-of-squared residuals (SSR) was used to select the best fit models (Figs. 3.3A, 3.3B). The ssDNA and dsDNA dimer models A (monomer addition only) were obtained by setting R_1 = RAD51 injection concentration and $R_2 = 0$. The ssDNA model B ($\mathbf{k_{el.r}} = \mathbf{k_s}$) was obtained by setting $\mathbf{k_{el.r}} = \mathbf{k_s}$. The dsDNA model C ($\mathbf{k_{el.r}} \neq \mathbf{k_s}$) was obtained by changing the rate constant $\mathbf{k_s}$ in the $D_2\mathbf{k_s}$ term (equations 2.12, 2.13) to $\mathbf{k_{el.r}}$ (Fig. 2.3A-B).

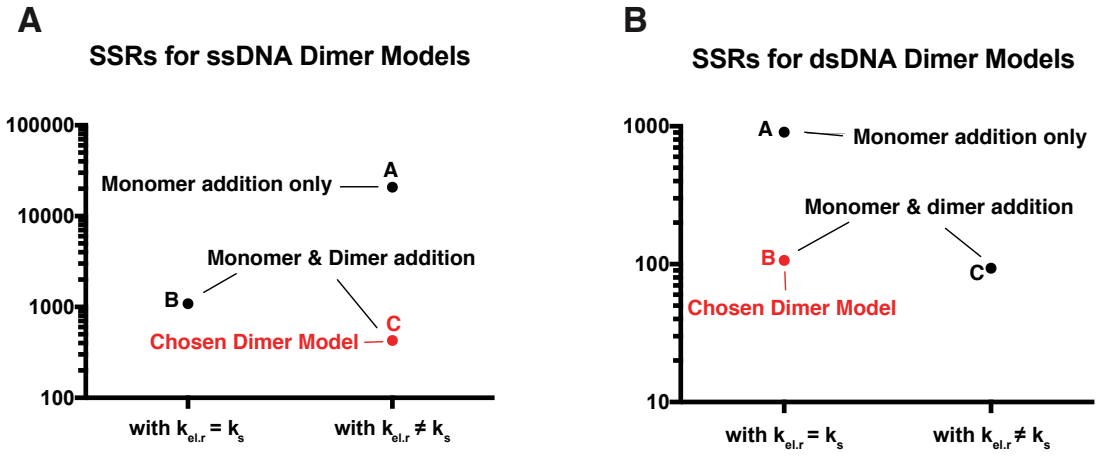


Figure 2.3: **A**. Sum-of-squared residuals of the ssDNA dimer models. Each model was simultaneously fit to one repeat of the ssDNA dN-X datasets via ABC-SMC. The chosen model (C) is indicated in red. **B**. Sum-of-squared residuals of the dsDNA dimer models. Each model was simultaneously fit to one repeat of the dsDNA dN-Xp datasets via ABC-SMC. The chosen model (B) is indicated in red.

ABC-SMC simulations were carried out in MATLAB 2016b to determine the SSRs for each model by determining the best fit parameter values (section 2.8) for the ssDNA dimer model to explain the ssDNA dN-X data (Figs. 2.3A, B.3A-B) and for the dsDNA dimer model to explain the dsDNA dN-Xp data (Fig. 2.3B, B.3C-D). For dsDNA, model B was chosen over model C because it displayed a similar SSR as model C but required only 4 fit parameters (Fig. 2.3B: model B $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_s}$) instead of 5 (Fig. 2.3B: model C $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_{el.r}}$, $\mathbf{k_s}$).

2.9.4 Interpretation of the ssDNA dT-50 $k_{el.f}$ Rate

Given the very high $\mathbf{k_{el.f}}$ value for the dT-50 oligo (Fig. 4.3B-C), it is important to highlight that under these experimental conditions RAD51 binding to the dT-50 oligo may be limited by the rate of diffusion of RAD51 within the SPR flow cell [198], and therefore the fitted $\mathbf{k_{el.f}}$ values may reflect the rate of RAD51 diffusion and not the actual bio-physical $\mathbf{k_{el.f}}$ rate.

2.10 n-mer Models

In order to relax the assumption of RAD51 polymerisation via dimer addition, I extended these ODE models describing RAD51 polymerisation on ssDNA and dsDNA to include RAD51 polymerisation in solution up to the maximum length of a 16-mer, collectively referred to as n-mer models. The ssDNA and dsDNA n-mer models each consist of an equilibrium sub-model and a kinetic sub-model. The equilibrium sub-model was formulated using second order mass-action kinetics via the rule-based modelling language BioNetGen [216], and consists of 16 non-linear ODEs describing the formation of RAD51 polymers in solution up to a maximum length of 16 (i.e. the predicted maximum number of RAD51 molecules binding to a DNA 50-mer [113]). Each ODE describes the rate of change in concentration of a RAD51 polymer in solution with respect to time. They are non-linear because monomeric RAD51 is consumed in the process of polymerisation to form RAD51 polymers. In this sub-model, any RAD51 n-mer ($1 \leq n < 16$) can bind to any other RAD51 m-mer ($1 \leq m < 16$) to form a RAD51 $(n+m)$-mer ($1 < n+m \leq 16$). In addition, any RAD51 k-mer can fall apart in every possible combination of m-mers and n-mers ($k = n + m$) (e.g. a pentamer can fall apart to form a monomer and a tetramer, or a dimer and a trimer). The single dissociation constant ($\mathbf{K_D}$) describing all of these pairwise interactions is a fit parameter derived from the SPR data. The model is evaluated at equilibrium to calculate the concentration of each RAD51 polymer in solution, prior to the SPR injection. The concentration distribution of RAD51 polymers in solution depends on the fit parameter $\mathbf{K_D}$ and on the RAD51 monomer concentration (i.e. the RAD51 SPR injection concentration). The predicted concentrations of RAD51 polymers are then inserted into the kinetic sub-models.

The ssDNA and dsDNA kinetic sub-models describe the formation of RAD51 polymers on ssDNA and dsDNA, respectively. Importantly, during SPR injections the

concentrations of RAD51 polymers in solution in the flow cell remains constant because they are continuously replenished by flow. For this reason, the kinetic sub-models were formulated using first order mass-action kinetics via the rule-based modelling language BioNetGen [216] and each consist of 17 linear ODEs. Each ODE describes the rate of change in mean concentration of a DNA-bound RAD51 n-mer on the SPR chip with respect to time, where $1 \leq n \leq b$ and b is the maximum number of RAD51 binding sites on the DNA molecule (Table 2.3).

Table 2.3: DNA molecule names and lengths, with corresponding maximum RAD51 polymer lengths (b) and n-mer model (NM) ODEs used for SPR data fitting. SPR data obtained using the ssDNA dN-5 and dN-11 oligos was not included in the data fitting because they were identical to the data obtained using the ssDNA dN-8 oligo and described no RAD51 binding to these oligos. NA: not applicable. ESM: equilibrium sub-model.

Molecule Name	Length (nt or bp)	b	NM ODEs
ssDNA dN-5	5	1	NA
dsDNA dN-5p	5	1	ESM, 2.26 - 2.27
ssDNA dN-8	8	2	ESM, 2.21 - 2.23
dsDNA dN-8p	8	2	ESM, 2.26 - 2.28
ssDNA dN-11	11	3	NA
dsDNA dN-11p	11	3	ESM, 2.26 - 2.29
ssDNA dN-14	14	4	ESM, 2.21 - 2.24
ssDNA dN-17	17	5	ESM, 2.21 - 2.24
ssDNA dN-50	50	16	ESM, 2.21 - 2.25
ssDNA dT-50	50	16	ESM, 2.21 - 2.25
ssDNA dA-50	50	16	ESM, 2.21 - 2.25
dsDNA dN-50p	50	16	ESM, 2.26 - 2.30

2.10.1 Assumptions

For the equilibrium sub-model, I assume

- any RAD51 n-mer ($1 \leq n < 16$) can bind to any other RAD51 m-mer ($1 \leq m < 16$) to form a RAD51 $(n+m)$-mer ($1 < n+m \leq 16$)
- any RAD51 k-mer can fall apart in every possible combination of m-mers and n-mers ($k = n + m$)
- the dissociation constant between RAD51 polymers $\mathbf{K_D}$ is independent of RAD51 polymer length: $\mathbf{K_D}(n,m) = \mathbf{K_D}(1,1) = \mathbf{K_D}$. This assumption was previously applied in [12]

For the ssDNA and dsDNA kinetic sub-models, I assume

- Each ssDNA or dsDNA molecule can accommodate a RAD51 polymer of maximum length b, where b is the maximum number of RAD51 binding sites on the DNA molecule (Table 2.3)
- RAD51 only nucleates once per DNA molecule (i.e. each DNA molecule only accommodates a single polymer of length n, where $1 \leq n \leq b$). In other words, polymer elongation is faster than nucleation. This is a valid assumption give that the longest ssDNA and dsDNA molecules used in this study are 50 nt and 50 bp long respectively, and RAD51 nucleation was shown to occur on average once every 500 bp in reactions with a 1 RAD51 : 3 bp molar ratio [137]. In the case of the experiments using WT RAD51 and the ssDNA dN-50 or dsDNA dN-50p, I used a 150 nM RAD51 concentration. Considering an average 10 RU of immobilised dsDNA and a dN-50p molecular weight of 31.157 kDa, the molar ratio in my experimental system is far lower compared to that in [137], and hence nucleation is expected on average to occur only once on a DNA molecule: [bps] $= 50(\frac{\mathsf{R_{lig}}}{100\mathsf{M_{lig}}})[218] = 50(\frac{10\mathsf{RU}}{(100)31157\mathsf{Da}}) \approx 160\ \mu$M. Hence $\frac{150\mathsf{nM\ RAD51}}{160\mu\mathsf{M\ bps}} \Rightarrow$ 1 RAD51 molecule per $\approx$ 1100 bps on the SPR chip.

- Any RAD51 n-mer ($1 \leq n \leq b$) can adsorb onto DNA from solution. If $n < b$, further RAD51 polymers in solution can elongate the DNA-bound n-mer to form a DNA-bound (n^*)-mer ($1 < n^* \leq b$)
- The rate of adsorption of RAD51 n-mers on DNA is the same as the rate of elongation of RAD51 n-mers on DNA. This rate is described as the polymerisation rate constant $\mathbf{k_p}$
- RAD51 dissociation occurs via the removal of single RAD51 protomers and of short, unstable RAD51 nuclei. This is a valid assumption when considering that in all the SPR experiments in this study ATP hydrolysis is inhibited via the addition of Ca^{2+}, leading to very slow RAD51 filament disassembly which can be approximated as disassembly via monomer removal or via the removal of short, unstable RAD51 nuclei

2.10.2 Description of ODEs - Equilibrium Sub-Model

A previous study applied a closed form expression to evaluate the concentration of RAD51 n-mers as a function of $\mathbf{K_D}$ and the RAD51 monomer concentration [12], but did not provide a derivation of the expression. For this reason, I preferred to develop a comprehensive ODE model using BioNetGen [216] that takes account of all n-mer + m-mer interactions. The BioNetGen code for the equilibrium sub-model is reported in section A.3. The resulting ODEs for the equilibrium sub-model in compact form are as follows:

for $n = 1$:

$$\frac{dR_1}{dt} = k_{\mathsf{d}}(\frac{1}{2}R_2 + \sum_{3}^{16} R_i) - k_{\mathsf{a}}R_1(2R_1 + \sum_{2}^{15} R_i) \tag{2.17}$$

for $2 \leq n \leq 8$:

$$\frac{dR_n}{dt} = k_{\mathsf{d}}(\sum_{n+1}^{2n-1} R_i + \frac{1}{2}R_{2n} + \sum_{2n+1}^{16} R_i - \left\lfloor \frac{n}{2} \right\rfloor R_n) - k_{\mathsf{a}}[R_n(\sum_{1}^{n-1} R_i + 2R_n + \sum_{n+1}^{16-n} R_i) - \sum_{1}^{\left\lfloor \frac{n}{2} \right\rfloor} R_i R_{n-i}] \tag{2.18}$$

for $9 \leq n \leq 15$:

$$\frac{dR_n}{dt} = k_{\mathsf{d}}(\sum_{n+1}^{16} R_i - \left\lfloor \frac{n}{2} \right\rfloor R_n) - k_{\mathsf{a}}(R_n \sum_{1}^{16-n} R_i - \sum_{1}^{\left\lfloor \frac{n}{2} \right\rfloor} R_i R_{n-i}) \tag{2.19}$$

for $n = 16$:

$$\frac{dR_{16}}{dt} = k_{\mathsf{a}} \sum_{1}^{8} R_i R_{16-i} - 8k_{\mathsf{d}}R_{16} \tag{2.20}$$

with initial conditions $R_1(0) =$ RAD51 SPR injection concentration, $R_{2,\ldots,16}(0) = 0$.

R_n is the time-dependent concentration of RAD51 n-mers in solution, k_{a} and k_{d} are the association and dissociation rate constants of the RAD51 protomer-protomer interaction, and $\lfloor . \rfloor$ is the floor function (e.g. $\lfloor 2.7 \rfloor = 2$). To fit the $\mathbf{K_D}$ of the RAD51 protomer-protomer interaction, k_{a} was set to 1 /μM/s and the k_{d} was fit by simulating the model at equilibrium. Hence, $\mathbf{K_D} = \frac{k_{\mathsf{d}}}{k_{\mathsf{a}}} = k_{\mathsf{d}}$.

2.10.3 Description of ODEs - Kinetic Sub-Models

The ODEs for the general ssDNA and dsDNA kinetic sub-models were both generated via the same BioNetGen code (section A.4). The resulting ODEs for the

general ssDNA kinetic sub-model in compact form are as follows:

$$\frac{dS_0}{dt} = \sum_{1}^{b} \mathbf{k_d^d} S_i - \mathbf{k_p} S_0 [\sum_{1}^{b} i(b+1-i) R_i] \tag{2.21}$$

$$\frac{dS_1}{dt} = -\mathbf{k_1^d} S_1 + \mathbf{k_p} b R_1 S_0 - \mathbf{k_p} S_1 \sum_{1}^{b-1} R_i + \mathbf{k_s} S_2 \tag{2.22}$$

$$\frac{dS_2}{dt} = -\mathbf{k_2^d} S_2 + \mathbf{k_p} [R_1 S_1 + 2(b-1) R_2 S_0] - \mathbf{k_p} S_2 \sum_{1}^{b-2} R_i + \mathbf{k_s} (S_3 - S_2) \tag{2.23}$$

for $3 \leq n \leq 15$:

$$\frac{dS_n}{dt} = -\mathbf{k_n^d} S_n + \mathbf{k_p} [\sum_{1}^{n-1} R_i S_{n-i} + n(b+1-n) R_n S_0] - \mathbf{k_p} S_n \sum_{1}^{b-n} R_i + \mathbf{k_s} (S_{n+1} - S_n) \tag{2.24}$$

$$\vdots$$

$$\frac{dS_{16}}{dt} = -\mathbf{k_{16}^d} S_{16} + \mathbf{k_p} [\sum_{1}^{15} R_i S_{16-i} + 16(b-15) R_{16} S_0] - \mathbf{k_s} S_{16} \tag{2.25}$$

With initial conditions $S_0(0) = 1, S_{1,\ldots,16}(0) = 0$.

Similarly, the ODEs for the general dsDNA kinetic sub-model in compact form are as follows:

$$\frac{dD_0}{dt} = \sum_1^b \mathbf{k_i^d} D_i - \mathbf{k_p} D_0 [\sum_1^b i(b+1-i)R_i] \tag{2.26}$$

$$\frac{dD_1}{dt} = -\mathbf{k_1^d} D_1 + \mathbf{k_p} b R_1 D_0 - \mathbf{k_p} D_1 \sum_1^{b-1} R_i + \mathbf{k_s} D_2 \tag{2.27}$$

$$\frac{dD_2}{dt} = -\mathbf{k_2^d} D_2 + \mathbf{k_p}[R_1 D_1 + 2(b-1)R_2 D_0] - \mathbf{k_p} D_2 \sum_1^{b-2} R_i + \mathbf{k_s}(D_3 - D_2) \tag{2.28}$$

for $3 \leq n \leq 15$:

$$\frac{dD_n}{dt} = -\mathbf{k_n^d} D_n + \mathbf{k_p}[\sum_1^{n-1} R_i D_{n-i} + n(b+1-n)R_n D_0] - \mathbf{k_p} D_n \sum_1^{b-n} R_i + \mathbf{k_s}(D_{n+1} - D_n) \tag{2.29}$$

$$\vdots$$

$$\frac{dD_{16}}{dt} = -\mathbf{k_{16}^d} D_{16} + \mathbf{k_p}[\sum_1^{15} R_i D_{16-i} + 16(b-15)R_{16} D_0] - \mathbf{k_s} D_{16} \tag{2.30}$$

With initial conditions $D_0(0) = 1, D_{1,...,16}(0) = 0$.

S_n and D_n ($n = 1, ..., 16$) represent the time-dependent concentration of ssDNA-bound and dsDNA-bound RAD51 n-mers on the SPR chip, respectively. S_0 and D_0 represent the concentration of unbound ssDNA oligos and dsDNA oligos on the SPR chip, respectively. The adsorption of a RAD51 i-mer onto a DNA molecule ($i \leq b$) can occur on any $b+1-i$ sequence of i DNA binding sites. In addition, the adsorption of a RAD51 i-mer can initiate via any of its i protomers first binding to its corresponding DNA binding site with any of the $b+1-i$ sequences, and therefore the terms $\mathbf{k_p} S_0 R_i$ (equations 2.21 - 2.25) and $\mathbf{k_p} D_0 R_i$ (equations 2.26 - 2.30) are multiplied by $i(b+1-i)$. The R_n values are constants that represent the concentration of RAD51 n-mers in solution and are derived from the equilibrium sub-model (equations 2.17 - 2.20). At the end of the RAD51 SPR injection, all R_n values are set to 0. $\mathbf{k_n^d}$ ($1 \leq n \leq 16$), $\mathbf{k_p}$, and $\mathbf{k_s}$ are the kinetic rate constants describing the dynamics of the ODE models. $\mathbf{k_n^d}$ represents the adsorption reverse rate constant for each n-mer on a DNA molecule with b DNA binding sites ($n \leq b$) (i.e. the one-step

dissociation of RAD51 n-mers from DNA), $\mathbf{k_p}$ is the polymerisation rate constant, and describes the rate of adsorption and elongation of each n-mer on DNA, and $\mathbf{k_s}$ is the rate of RAD51 single protomer dissociation from DNA. The values of these rate constants are obtained by fitting the ODE model outputs ($\sum_1^b iS_i$ and $\sum_1^b iD_i$) to their respective data sets (Table 2.3) either via ABC-SMC (Figs. 3.9, 5.7) or least-squares fitting (Fig. 4.5). The initial conditions $S_0(0)$ and $D_0(0)$ were set to 1 to match the SPR data normalisation, such that $(\max \sum_1^{16} iS_i) = (\max \sum_1^{16} iD_i) = 16$ (section 2.7).

2.10.4 Model Selection

The general ssDNA and dsDNA and kinetic sub-models were used as templates to generate a set of ssDNA and dsDNA kinetic sub-models. In particular, I started from the simplest possible ssDNA and dsDNA kinetic sub-models that describe RAD51 polymerisation on DNA via 3 fit parameters: $\mathbf{k_p}$, $\mathbf{K_D}$, and $\mathbf{k_s}$ acting as both the single RAD51 protomer dissociation rate constant and the reverse adsorption rate constant for all RAD51 n-mers (KSM A: Table 2.4). I then generated further kinetic sub-models by either 1) sequentially increased the number of different adsorption reverse rate constants, or 2) varying the reactions regulated by specific adsorption reverse rate constants (KSMs B-I: Table 2.4).

Table 2.4: The set of ssDNA and dsDNA kinetic sub-models derived from the general ssDNA kinetic sub-model and the dsDNA kinetic sub-model. Rows describe the simplifications applied to the general ssDNA and dsDNA kinetic sub-models (2.21 - 2.25, 2.26 - 2.30) to obtain sub-models A-I, and the total number of fit parameters for each sub-model. Total fit parameter counts include the $\mathbf{K_D}$ from the equilibrium sub-model. NA: not applicable. KSM: kinetic sub-model.

KSM	ssDNA	Fit Parameters	dsDNA	Fit Parameters
A	$\mathbf{k}^{\mathbf{d}}_{1...16} = \mathbf{k_s}$	3: $\mathbf{k_s}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1...16} = \mathbf{k_s}$	3: $\mathbf{k_s}$, $\mathbf{k_p}$, $\mathbf{K_D}$
B	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4...16} = \mathbf{k_s}$	4: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2...16} = \mathbf{k_s}$	4: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_p}$, $\mathbf{K_D}$
C	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4,5} = \mathbf{k_s}$ $\mathbf{k}^{\mathbf{d}}_{6...16} = 0$	4: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2} = \mathbf{k_s}$ $\mathbf{k}^{\mathbf{d}}_{3...16} = 0$	4: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_p}$, $\mathbf{K_D}$
D	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5...16} = \mathbf{k_s}$	5: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2...16} = 0$	4: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_p}$, $\mathbf{K_D}$
E	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5} = \mathbf{k_s}$ $\mathbf{k}^{\mathbf{d}}_{6...16} = 0$	5: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{3...16} = \mathbf{k_s}$	5: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_p}$, $\mathbf{K_D}$
F	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5...16} = 0$	5: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{3} = \mathbf{k_s}$ $\mathbf{k}^{\mathbf{d}}_{4...16} = 0$	5: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_p}$, $\mathbf{K_D}$
G	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5...16} = \mathbf{k_{r1}}$	6: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_{r1}}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{3...16} = \mathbf{k_{r1}}$	6: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_{r1}}$, $\mathbf{k_p}$, $\mathbf{K_D}$
H	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5} = \mathbf{k_{r1}}$ $\mathbf{k}^{\mathbf{d}}_{6...16} = 0$	6: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_{r1}}$, $\mathbf{k_p}$, $\mathbf{K_D}$	$\mathbf{k}^{\mathbf{d}}_{1} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{2} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{3} = \mathbf{k_{r1}}$ $\mathbf{k}^{\mathbf{d}}_{4...16} = 0$	6: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_{r1}}$, $\mathbf{k_p}$, $\mathbf{K_D}$
I	$\mathbf{k}^{\mathbf{d}}_{1...3} = \mathbf{k_u}$ $\mathbf{k}^{\mathbf{d}}_{4} = \mathbf{k_q}$ $\mathbf{k}^{\mathbf{d}}_{5} = \mathbf{k_{r1}}$ $\mathbf{k}^{\mathbf{d}}_{6...16} = \mathbf{k_{r2}}$	7: $\mathbf{k_s}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_{r1}}$, $\mathbf{k_{r2}}$, $\mathbf{k_p}$, $\mathbf{K_D}$	NA	NA

ABC-SMC simulations were carried out in MATLAB 2016b to determine the SSRs for each kinetic sub-model (section 2.8). Each ssDNA kinetic sub-model was indi-

vidually fit to the WT RAD51 ssDNA dN-X data, and each dsDNA kinetic sub-model was individually fit to the WT RAD51 dsDNA dN-Xp data. The sum-of-squared residuals (SSRs) was used to then select the simplest best fit ssDNA sub-model (F - Table 2.4) and dsDNA sub-model (D - Table 2.4) (Figs. 2.4, B.4, B.5, B.6, B.7).

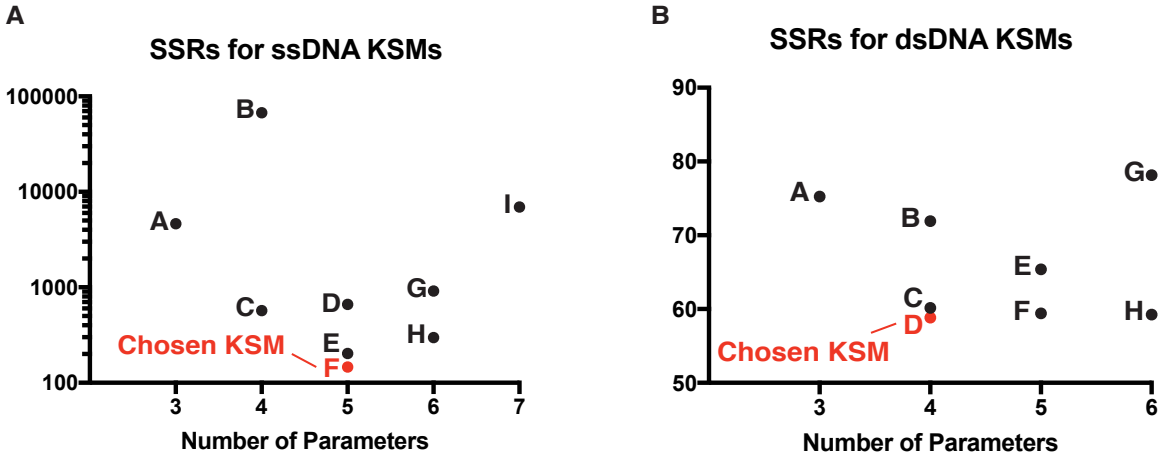


Figure 2.4: **A**. Sum-of-squared residuals of the n-mer ssDNA kinetic sub-models. Each kinetic sub-model together with the equilibrium sub-model was fit to one repeat of the ssDNA dN-X datasets via ABC-SMC. The chosen model (F) is indicated in red. See Figs. B.4, B.5 for fits of KSMs A-E, G-I. **B**. Sum-of-squared residuals of the n-mer dsDNA kinetic sub-models. Each kinetic sub-model together with the equilibrium sub-model was fit to one repeat of the dsDNA dN-Xp datasets via ABC-SMC. The chosen model (D) is indicated in red. See Figs. B.6, B.7 for fits of KSMs A-C, E-H.

The ssDNA kinetic sub-model F and dsDNA kinetic sub-model D were then simultaneously fit to the ssDNA dN-X data and the dsDNA dN-Xp data to obtain best fit parameter values (Table 3.2, Fig. 3.8A-D). Although the $\mathbf{k_u}$ fit for ssDNA dN-X was undetermined (Table 3.2), the probability density of $\mathbf{k_u}$ derived from the ABC-SMC fit nonetheless suggests dN-X $\mathbf{k_u} \gg$ dN-X $\mathbf{k_s}$ (Fig. 3.9A). For this reason, the ssDNA kinetic sub-model F was further simplified to remove the $\mathbf{k_s}S_2$ and $\mathbf{k_s}S_3$ terms from equations 2.22 - 2.24. Using the ssDNA kinetic sub-model F and dsDNA kinetic sub-model D, lsqcurvefit (MATLAB 2016b) was used to then fit the WT RAD51 ssDNA dA-50, dT-50 and dsDNA dN-50p SPR data by individually varying either

$\mathbf{k_p}$, $\mathbf{k_u}$, $\mathbf{k_q}$ or $\mathbf{k_s}$ (Fig. 4.5A-D).

ABC-SMC simulations were also carried out in MATLAB 2016b to simultaneously fit the F86E RAD51 ssDNA dT-50, dN-50, dA-50 and the dsDNA dN-50p SPR data using the ssDNA and dsDNA kinetic sub-model A (Table 2.4) (Figs. 5.8A-D, 5.7A-B).

Chapter 3

Mechanistic Models of RAD51 Polymerisation

3.1 Introduction

DNA double-strand breaks (DSBs) are cytotoxic lesions that can lead to chromosomal breaks, genomic instability and tumourigenesis in mammalian cells [13]. Homologous recombination (HR) offers an error-free DNA repair pathway that accurately repairs DSBs without loss of genetic information, and in this way, contributes to genome stability. During HR-mediated repair, single-stranded DNA overhangs are generated at DSB sites and are rapidly coated with the ssDNA-binding protein RPA [219]. RAD51 is the key ATP-dependent recombinase that is recruited to ssDNA at DSBs following RPA removal to catalyse HR-mediated repair. In this process, RAD51 polymerises on ssDNA to form a ssDNA-nucleoprotein filament and guide homologous strand invasion and DSB repair [56].

In vivo, a plethora of protein co-factors (e.g. BRCA2, PALB2, RAD52) coordinate the recruitment of RAD51 to ssDNA and not on dsDNA, and this is thought to increase the RAD51 local concentration at DSBs [5, 6, 7, 8, 9, 10]. Electrophoretic mobility shift assay (EMSA) studies have suggested human RAD51 binds ssDNA with higher affinity, although this preference is salt-dependent [93, 111, 133]. However, such equilibrium assays have not provided information as to whether kinetics

may contribute to a potential RAD51-dependent ssDNA / dsDNA discrimination mechanism.

Kinetic analyses have suggested that RAD51 polymerisation on both ssDNA and dsDNA consists of a rate-limiting nucleation phase and a growth phase [110, 11, 109], with a minimal polymer nucleus length of either 2-3 RAD51 [110] or 4 - 5 RAD51 protomers on DNA [11, 109]. A more recent study suggested RAD51 displays faster association kinetics on ssDNA [12], while Mine' et al. suggested RAD51 displays slower dissociation kinetics on dsDNA in the presence of ATP hydrolysis [11]. Furthermore, in the absence of ATP hydrolysis, RAD51 was shown to nucleate on average once every 500 bp when incubated with dsDNA at a molar ratio of 1 protomer for every 3 bp [137]. In the presence of ATP hydrolysis, it was shown that RAD51 dissociation from DNA occurs in bursts interspersed by pauses and leads to the contraction of DNA [138]. Despite progress in understanding RAD51 polymerisation on DNA, the mechanism driving the observed differences in RAD51 binding kinetics on ssDNA and dsDNA is not fully understood, and may provide key insight into how RAD51 discriminates between ssDNA and dsDNA.

In this chapter, I analyse the RAD51 binding kinetics to ssDNA and dsDNA oligos of variable length using SPR and derive data-driven, ODE models of RAD51 polymerisation on ssDNA and dsDNA based on standard mass-action kinetics. I also use ABC-SMC to derive model fits to the experimental data that provide novel insight into the mechanisms of RAD51 polymerisation on ssDNA and dsDNA. These results lay the foundations for further analyses to assess how specific steps in the RAD51 polymerisation process may differ between ssDNA and dsDNA.

3.2 RAD51 Preferentially Associates with ssDNA

I first used SPR to confirm that human RAD51 discriminates between ssDNA and dsDNA. Indeed, I assessed the binding kinetics of untagged recombinant human wild-type (WT) RAD51 to a 50-mer mixed sequence ssDNA molecule (dN-50) and a 50-mer mixed sequence paired dsDNA molecule (dN-50p). These analyses performed in the presence of ATP and Ca^{2+}, which blocks RAD51 ATP hydrolysis [106], showed 1) faster RAD51 association to ssDNA compared to dsDNA, and 2) similar RAD51 lifetimes on both ssDNA and dsDNA (Fig. 3.1A-B). These observations validated that RAD51 distinguishes ssDNA from dsDNA through faster polymerisation on ssDNA [12].

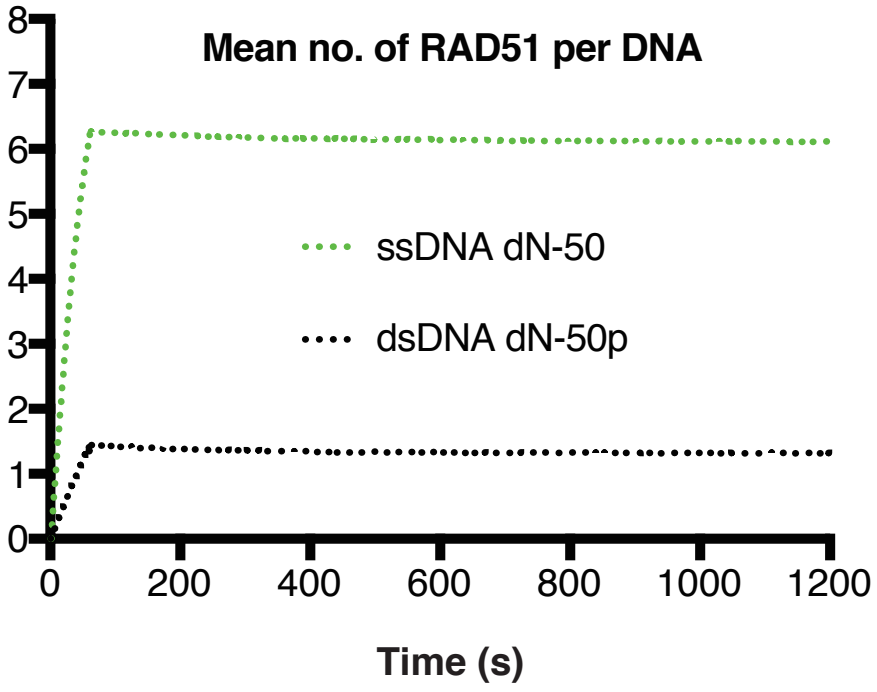


Figure 3.1: RAD51 polymerises faster on ssDNA. WT RAD51 SPR curves for ssDNA dN-50 and dsDNA dN-50p. RAD51 was injected at 150 nM in the presence of 2.5 mM ATP and 10 mM $CaCl_2$. All curves were processed and normalised according to section 2.7.

3.3 RAD51 exhibits stronger binding to dsDNA when DNA is short, but stronger binding to ssDNA when DNA is long

I then aimed to define the initial steps of RAD51 association with ssDNA. To this end, I generated a series of mixed sequence short ssDNA oligos with different lengths, each of which can be bound by a restricted number of RAD51 molecules. As a single RAD51 molecule engages with three nucleotides of DNA [113], ssDNA consisting of 5 nucleotides (dN-5), 8 nucleotides (dN-8), 11 nucleotides (dN-11), 14 nucleotides (dN-14) and 17 nucleotides (dN-17) can accommodate up to one, two, three, four and five RAD51 molecules, respectively (Fig. 3.2A). My systematic SPR measurements of RAD51 binding kinetics revealed no RAD51 binding to the dN-5, dN-8 and dN-11 (30 μM [RAD51]), slow association and moderate dissociation with the dN-14 (3 μM [RAD51]), and slightly faster association and slow dissociation with the dN-17 (3 μM [RAD51]) (Figs. 3.2B, B.1A-E). These observations suggested that three or less RAD51 molecules are unable to generate a stable nucleus on ssDNA, four RAD51 molecules form a quasi-stable nucleus, and five or more RAD51 molecules are able to form a highly stable nucleus.

I next evaluated the initial steps of RAD51 association with dsDNA. I generated an analogous series of mixed sequence dsDNA molecules consisting of 5 base pairs (dN-5p), 8 base pairs (dN-8p) and 11 base pairs (dN-11p), which can accommodate up to one, two or three RAD51 molecules, respectively (Fig. 3.2C). My SPR measurements detected no RAD51 binding to the dsDNA dN-5p (30 μM [RAD51]), as was the case for the ssDNA dN-5. Surprisingly, however, I detected slow association and moderate dissociation of RAD51 with the dN-8p (3 μM [RAD51]), and even faster association and slower dissociation with the dN-11p (3 μM [RAD51]) (Figs. 3.2D, B.2A-C). These observations initially suggested that one RAD51 molecule is unable to associate stably with dsDNA, but two RAD51 molecules form a quasi-

stable nucleus, and three or more RAD51 molecules are able to form a highly stable nucleus. This suggested RAD51 nucleation on dsDNA only requires 2-3 RAD51 molecules versus 4-5 on ssDNA. Altogether, these observations support the idea that RAD51 can bind more strongly to short (≤ 11 bp) dsDNA compared to short (≤ 11 nt) ssDNA molecules, but binds more strongly to long (50 nt) ssDNA molecules compared to long dsDNA molecules.

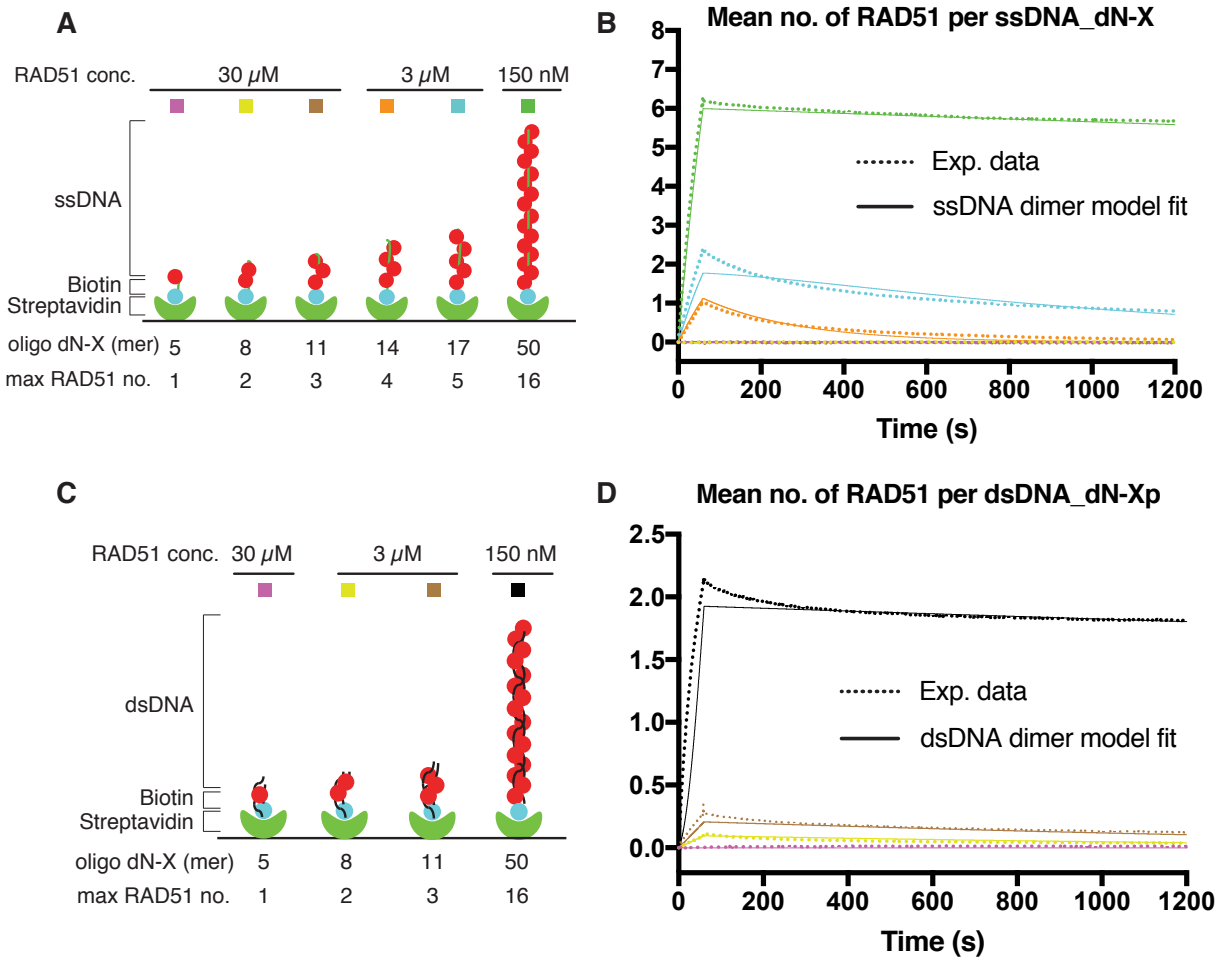


Figure 3.2: WT RAD51 binding kinetics to varying length of ssDNA and dsDNA, with corresponding dimer model fits. **A**. Biotinylated ssDNA oligos of indicated lengths were separately immobilised onto SPR CM5 chips via biotin-streptavidin interaction, and WT RAD51 was injected at the indicated concentrations to measure association and dissociation kinetics. The predicted number of maximum RAD51 protomers per oligo was calculated using the structurally determined RAD51 : nucleotide/base pair stoichiometry of 1:3 [113]. **B**. The dotted and solid curves respectively show the binding of RAD51 to the ssDNA oligos measured by SPR and the ssDNA dimer model fits. **C**. Biotinylated dsDNA oligos of indicated lengths were separately immobilised onto SPR CM5 chips via biotin-streptavidin interaction, and WT RAD51 was injected at the indicated concentrations to measure association and dissociation kinetics. The predicted number of maximum RAD51 protomers per oligo was calculated using the structurally determined RAD51 : nucleotide/base pair stoichiometry of 1:3 [113]. **D**. The dotted and solid curves respectively show the binding of RAD51 to the dsDNA oligos measured by SPR and the dsDNA dimer model fits. **B, D**. The sensorgrams were processed and normalised as described in section 2.7. Normalised sensorgrams represent the mean number of RAD51 molecules per DNA oligo.

3.4 Mechanistic models of RAD51 polymerisation predict faster adsorption and / or elongation to ssDNA

I then moved on to develop ODE models describing RAD51 polymerisation on ssDNA and dsDNA, informed by these SPR datasets. The aim of this was to develop models that could be simultaneously fit to all corresponding ssDNA dN-X and dsDNA dN-Xp datasets to explore the mechanistic differences in RAD51 polymerisation kinetics on ssDNA and dsDNA.

3.4.1 Development of Dimer Models

I first developed a simple set of models where I assumed that 1) RAD51 polymerisation mainly occurs via the addition of RAD51 dimers and to a lesser extent via the addition of RAD51 monomers (i.e. I fixed the concentrations of RAD51 dimers and monomers as described in section 2.9.1), and that 2) RAD51 dissociation only occurs via the removal of monomers. The assumption of RAD51 polymerisation mainly via dimer addition [217] allowed me to simplify model development by only considering two RAD51 species (dimers and monomers) as the key drivers in RAD51 polymerisation. Hence, these models are referred to as the **dimer models** (section 2.9).

Under these model assumptions, the polymerisation of RAD51 on DNA can be broken down into three key steps, namely adsorption, elongation and stabilisation. The adsorption step is characterised by the forward and reverse adsorption rate constants ($\mathbf{k_{ad.f}}$ and $\mathbf{k_{ad.r}}$) and describes the formation of an unstable RAD51 nucleus (1 - 3 protomers). The elongation step is characterised by the forward and reverse elongation rate constants ($\mathbf{k_{el.f}}$ and $\mathbf{k_{el.r}}$), and describes the formation of a quasi-stable nucleus (4 protomers). Finally, the stabilisation step is characterised by the same forward elongation rate constant ($\mathbf{k_{el.f}}$) but a different reverse rate constant,

namely the stabilisation rate constant ($\mathbf{k_s}$). This step describes stable nucleus formation (5 protomers), and also nucleus elongation up to a maximum length of 16, which corresponds to the predicted maximum RAD51 molecules binding to a 50-mer ssDNA molecule [113] (Fig. 3.3A) (model C - Fig. 2.3A). I also generated two simpler ssDNA dimer models that either describe RAD51 polymerisation only via monomer addition (Fig. 2.3A), or without taking account of the reverse elongation rate constant (i.e. $\mathbf{k_{el.r}} = \mathbf{k_s}$)(Fig. 2.3A). I then used ABC-SMC (section 2.8) to independently fit either model A, B or C simultaneously to the ssDNA dN-X data (dN-8, dN-14, dN-17, dN-50). I did not fit the dN-5 and dN-11 curves because they do not provide any additional information on RAD51 binding kinetics compared to the dN-8 curve, as RAD51 binding to these was undetectable. I found that model C (SSR = 463.6) describes the ssDNA dN-X data more accurately compared to A (SSR = 2338.2) and B (SSR = 1049.4) (Figs. 3.3A, 2.3A, B.3A-B).

ODE models A, B and C describing RAD51 polymerisation on dsDNA were then developed using similar assumptions as those applied to the respective ssDNA dimer models A, B and C. Using ABC-SMC to fit either model A, B or C to the dsDNA dN-Xp data (dN-5p, dN-8p, dN-11p, dN-50p), I found that model B (SSR = 101.2) explains the dN-Xp data more accurately compared to A (SSR = 963), and requires only four parameters to generate an accurate fit compared to five parameters for C (SSR = 98.8) (Figs. 3.3B, 2.3B, B.3C-D).

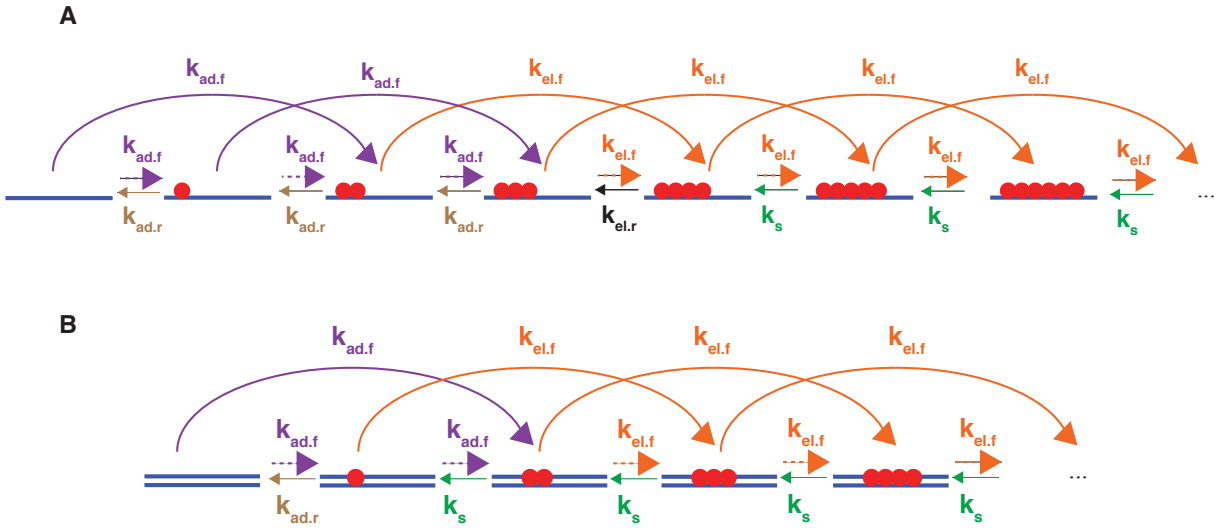


Figure 3.3: Kinetic representations of the dimer models. **A**. Kinetic representation of the ssDNA dimer model describing RAD51 polymerisation on ssDNA, consisting of five parameters: $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_{el.r}}$ and $\mathbf{k_s}$. **B**. Kinetic representation of the dsDNA dimer model describing RAD51 polymerisation on dsDNA, consisting of four parameters: $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$ and $\mathbf{k_s}$. **A-B**. Dotted arrows describe monomer binding, while solid curved arrows indicate dimer binding.

3.4.2 Fitting the Dimer Models

Using ABC-SMC to simultaneously fit the ssDNA dimer model C to the ssDNA dN-X SPR datasets, I obtained values for the five model parameters (Fig. 3.2B, Table 3.1, Fig. 3.4A-B). Fig. 3.4B shows that at 10 iterations the adaptive threshold ϵ of the ABC-SMC algorithm (section 2.8) has reached a plateau, indicating that the model has reached a best fit to the dN-X dataset. I further exploited ABC-SMC to determine a unique set of parameters that can explain the ssDNA SPR data. Specifically, Fig. 3.4A shows the probability densities of each individual parameter along the main diagonal, as well as particle densities for each parameter pair within each heat map.

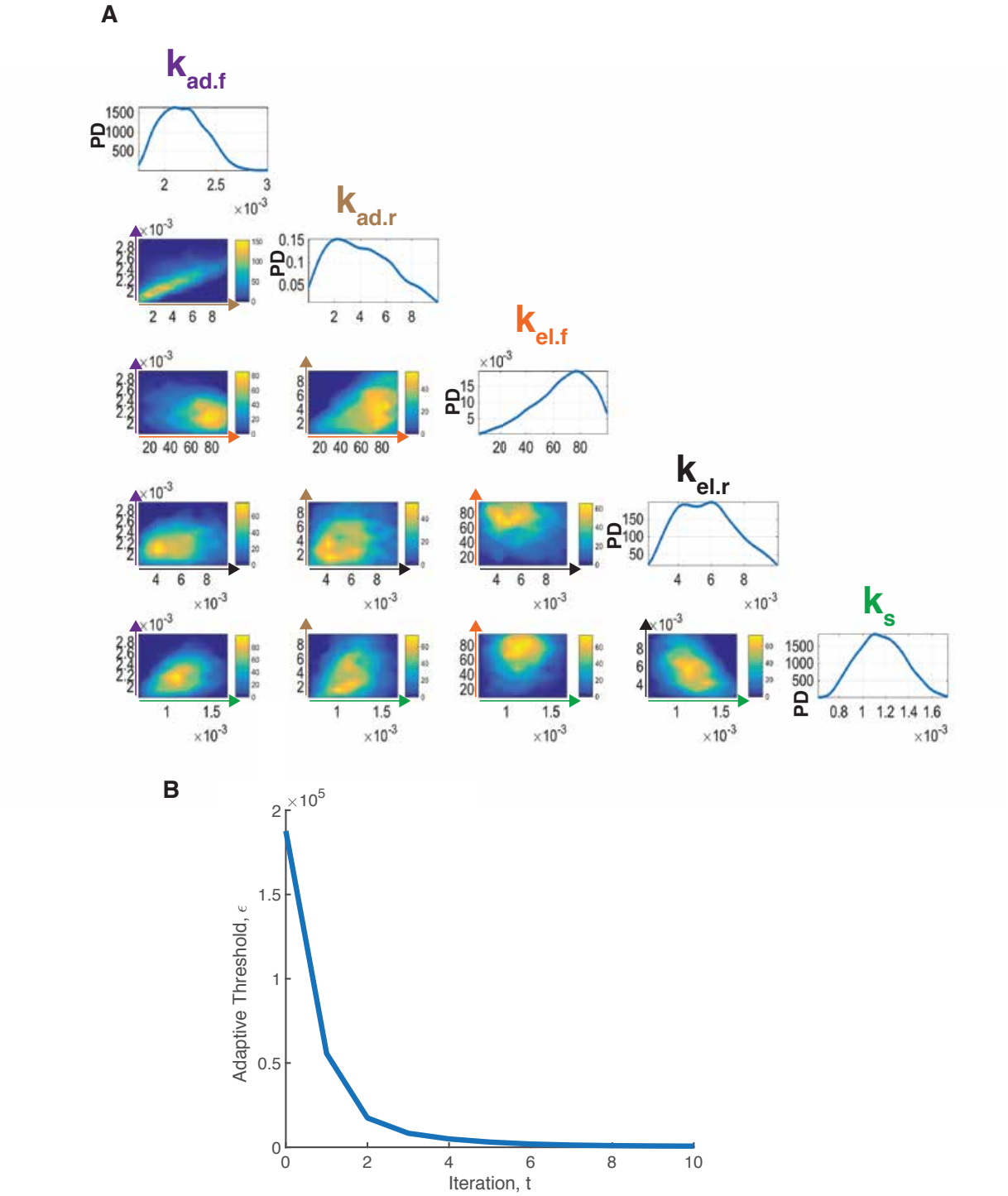


Figure 3.4: ABC-SMC probability densities for the ssDNA dimer model fit parameters for one repeat of dN-X datasets. **A**. Heat maps indicate particle frequency and describe pair-wise correlations between model parameters ($\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_{el.r}}$, and $\mathbf{k_s}$). Each vertical and horizontal axis of each heat map is labelled with the colour of the corresponding parameter ($\mathbf{k_{ad.f}}$ - purple, $\mathbf{k_{ad.r}}$ - brown, $\mathbf{k_{el.f}}$ - orange, $\mathbf{k_{el.r}}$ - black, $\mathbf{k_s}$ - green). A high particle frequency corresponds to a good model fit to the data for the specified parameter pair values. $\mathbf{k_{ad.f}}$ and $\mathbf{k_{ad.r}}$ show some degree of correlation. Probability densities of individual parameters are presented along the top left diagonal and were estimated using a Gaussian kernel. Particles were initialised via uniform priors with the following lower and upper bounds: $10^{-4} < \mathbf{k_{ad.f}} < 10^{-2}$, $10^{-3} < \mathbf{k_{ad.r}} < 20$, $10^{0} < \mathbf{k_{el.f}} < 10^{3}$, $10^{-4} < \mathbf{k_{el.r}} < 10^{-2}$, $10^{-5} < \mathbf{k_s} < 10^{-2}$. **B**. Adaptive threshold for the ABC-SMC simulation. The plateau at 10 iterations means the model has reached a best fit to the data. The simulation was carried out using MATLAB (10 iterations, 2000 particles).

I found that the model can also explain the data when either slightly higher or lower values are concurrently applied to both $\mathbf{k_{ad.r}}$ and $\mathbf{k_{ad.f}}$ (Fig. 3.4A). The reason that these parameters were not uniquely identified was because only a lower bound on $\mathbf{K_{ad}}$ ($\mathbf{K_{ad}} = \mathbf{k_{ad.r}} / \mathbf{k_{ad.f}}$) can be determined because no binding was observed to the dN-8. As a result, large variations of $\mathbf{k_{ad.r}}$ and $\mathbf{k_{ad.f}}$ could fit the data. Additionally, $\mathbf{k_{el.f}}$ displayed a wider range of good fit values ($\approx$ 50 - 80 /μM/s) compared to $\mathbf{k_{ad.f}}$ ($\approx (1.8 \times 10^{-3})$ - (2.6×10^{-3}) /μM/s) (Fig. 3.4A). Nonetheless, the accurate fit to the ssDNA dN-50 (R^2 = 0.95) suggests that the model captures the essential features of RAD51 polymerisation on ssDNA.

Importantly, the model suggested a high forward elongation rate constant ($\mathbf{k_{el.f}}$ = 53.7 ± 10.66 /μM/s) drives rapid RAD51 polymerisation on ssDNA. Indeed, although the experimental data measured an average of six RAD51 molecules per dN-50 at the end of the SPR injection, the model predicts that most RAD51 forms long polymers of 16, 15 and 14 molecules on a fraction of the DNA molecules during each injection (Fig. 3.6A). Similarly, the model predicts that RAD51 forms mainly pentamers and tetramers on the dN-17 (Fig. 3.6B), and only tetramers on the dN-14 (Fig. 3.6C). These results support the notion that stable RAD51 polymer formation on ssDNA requires the nucleation of at least five RAD51 protomers, but may be facilitated by a high elongation rate constant.

The ABC-SMC simulation of the dsDNA dimer model was able to provide more well defined values for the parameters of the dsDNA model compared to those of the ssDNA model (Fig. 3.2D, Table 3.1, Fig. 3.5A-B), suggesting this model explains the SPR data well.

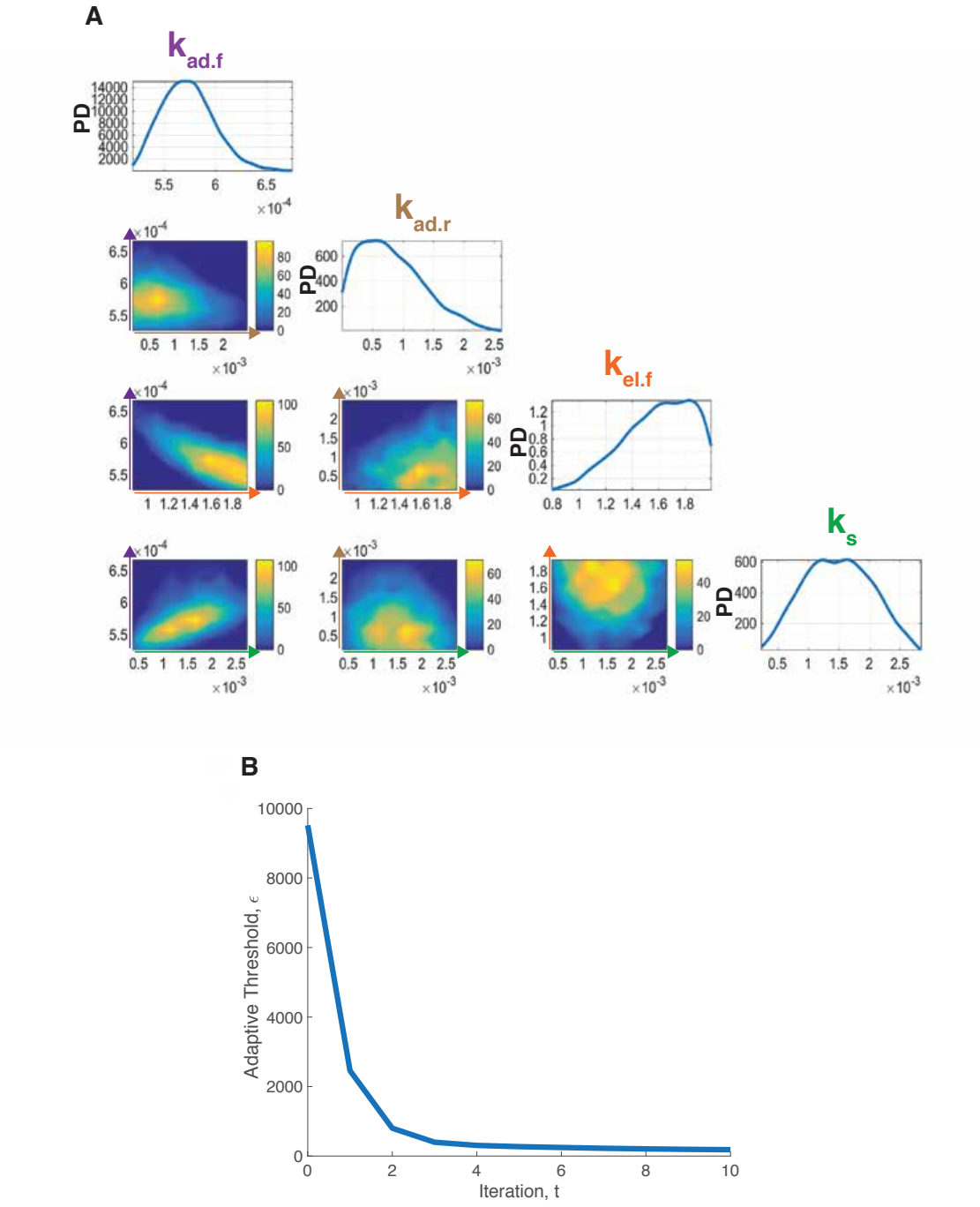


Figure 3.5: ABC-SMC probability densities for the dsDNA dimer model fit parameters for one repeat of dN-Xp datasets. **A**. Heat maps indicate particle frequency and describe pair-wise correlations between model parameters ($\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, and $\mathbf{k_s}$). Each vertical and horizontal axis of each heat map is labelled with the colour of the corresponding parameter ($\mathbf{k_{ad.f}}$ - purple, $\mathbf{k_{ad.r}}$ - brown, $\mathbf{k_{el.f}}$ - orange, $\mathbf{k_s}$ - green). A high particle frequency corresponds to a good model fit to the data for the specified parameter pair values. $\mathbf{k_{ad.f}}$ and $\mathbf{k_{ad.r}}$ show some degree of correlation. Probability densities of individual parameters are presented along the top left diagonal and were estimated using a Gaussian kernel. Particles were initialised via uniform priors with the following lower and upper bounds: $10^{-4} < \mathbf{k_{ad.f}} < 10^{-2}$, $10^{-5} < \mathbf{k_{ad.r}} < 10^{-1}$, $10^{-1} < \mathbf{k_{el.f}} < 10^{2}$, $10^{-5} < \mathbf{k_s} < 10^{-3}$. **B**. Adaptive threshold for the ABC-SMC simulation. The plateau at 10 iterations means the model has reached a best fit to the data. The simulation was carried out using MATLAB (10 iterations, 2000 particles).

Table 3.1: Fit parameter values for the ssDNA and dsDNA dimer models. The ssDNA dimer model was simultaneously fit to the SPR curves of ssDNA dN-8, dN-14, dN-17 and dN-50, while the dsDNA dimer model was simultaneously fit to the dN-5p, dN-8p, dN-11p and dN-50p. Mean values ± 1 SD of 3 mode ABC-SMC particles derived from model fits of 3 dN-X repeats and 3 dN-Xp repeats (n = 3). NA: not applicable.

Fit Parameter	ssDNA dN-X	dsDNA dN-Xp
$\mathbf{k_{ad.f}}$	(2.2 ± 0.18) × 10^{-3} /μM/s	(6.4 ± 0.9) × 10^{-4} /μM/s
$\mathbf{k_{ad.r}}$	(2.7 ± 1.86) /s	(1.1 ± 0.54) × 10^{-3} /s
$\mathbf{k_{el.f}}$	(53.7 ± 10.66) /μM/s	(1.653 ± 0.78) /μM/s
$\mathbf{k_{el.r}}$	(5.3 ± 0.72) × 10^{-3} /s	NA
$\mathbf{k_s}$	(1 ± 0.13) × 10^{-3} /s	(8 ± 1.3) × 10^{-4} /s

Importantly, the model fit suggested a lower adsorption forward rate constant on dsDNA ($\mathbf{k_{ad.f}}$ = (6.4 ± 0.9) × 10^{-4} /μM/s) compared to ssDNA ($\mathbf{k_{ad.f}}$ = (2.2 ± 0.18) × 10^{-3} /μM/s), as well as a considerably lower forward elongation rate constant on dsDNA ($\mathbf{k_{el.f}}$ = 1.653 ± 0.78 /μM/s) compared to ssDNA ($\mathbf{k_{el.f}}$ = 53.7 ± 10.66 /μM/s), describing the unproductive RAD51 polymerisation on dsDNA. This is reflected by the model prediction that most RAD51 does not form long polymers on the dN-50p (Fig. 3.6D). Moreover, the model predicted that RAD51 forms dimers with a small amount of trimers and monomers on the dN-11p (Fig. Fig. 3.6E), and dimers with a small amount of monomers on the dN-8p (Fig. 3.6F). Additionally, the model also predicted only a small difference between the adsorption reverse rate constant ($\mathbf{k_{ad.r}}$ = (1.1 ± 0.54) × 10^{-3} /s) and the stabilisation rate constant ($\mathbf{k_s}$ = (8 ± 1.3) × 10^{-4} /s), suggesting monomeric RAD51 dissociates slowly from dsDNA. However, the model predicts almost no RAD51 binding to the dsDNA dN-5p oligo because it assumes most RAD51 is dimeric and cannot associate to dN-5p (section 2.9).

Collectively, these analyses suggest that, while a RAD51 dimer is sufficient to form a stable nucleus on dsDNA, its elongation on dsDNA is restricted by a lower elongation rate constant compared to ssDNA. It remains undetermined whether a slower adsorption forward rate constant on dsDNA ($\mathbf{k_{ad.f}}$ = (6.4 ± 0.9) × 10^{-4} /μM/s) compared to ssDNA ($\mathbf{k_{ad.f}}$ = (2.2 ± 0.18) × 10^{-3} /μM/s) also contributes to slower RAD51

polymerisation on dsDNA, given that the adsorption reverse rate constant on ds-DNA is also slower compared to that of ssDNA ($\mathbf{k_{ad.r}}$ = 2.7 ± 1.86 /s). Hence, faster RAD51 polymerisation on ssDNA compared to dsDNA may be explained by faster elongation and/or adsorption on ssDNA.

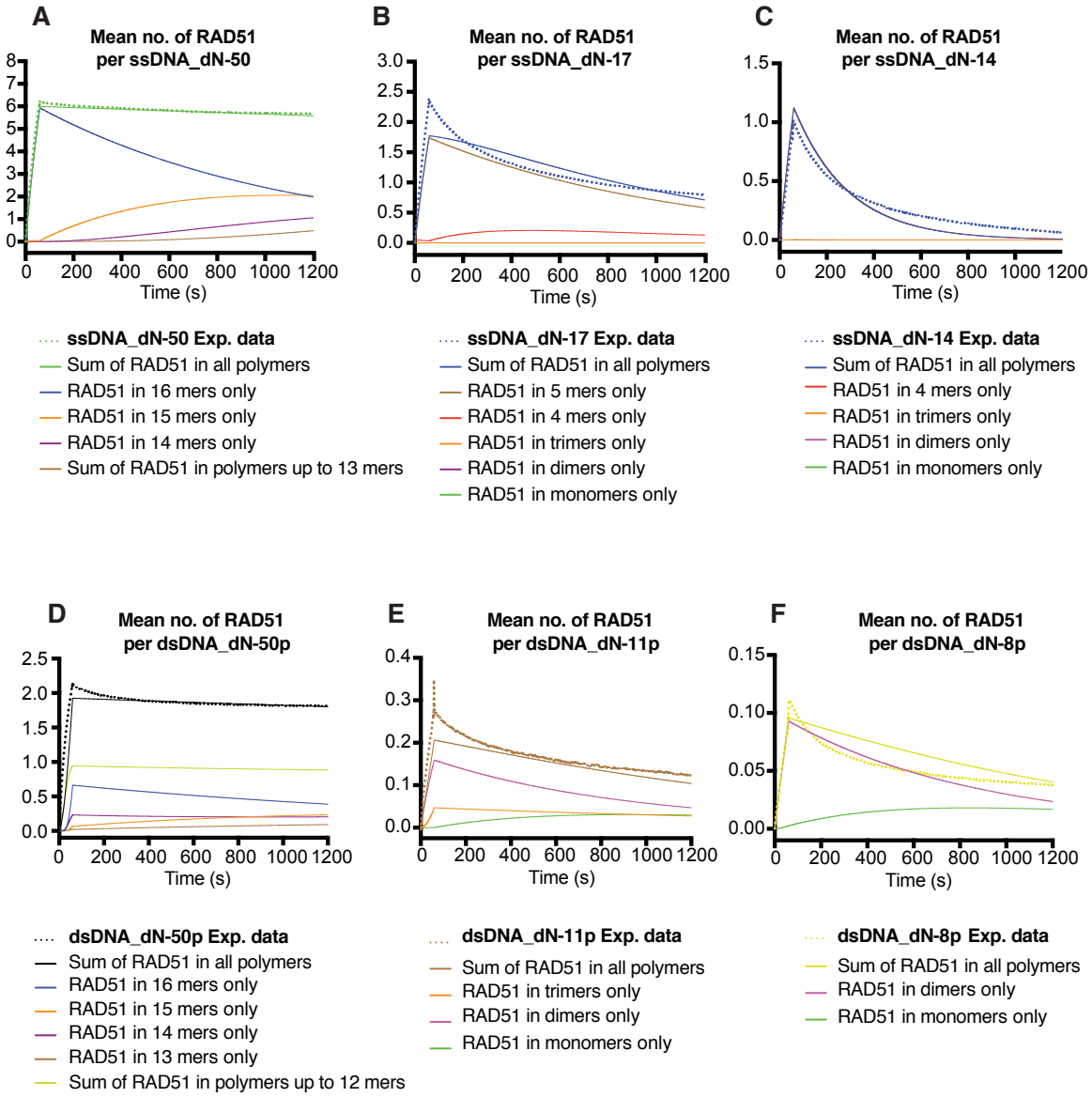


Figure 3.6: Predicted n-mer formation on DNA oligos using dimer models. **A**. Experimental data of ssDNA dN-50 from Fig. 3.2B (dotted lines) and predicted polymers with defined length on each single DNA oligo using the model from Fig. 3.3A (solid lines). Predicted polymer curves were calculated by multiplying the time-dependent concentration of a specific n-mer by its length (i.e. $nS_n(t)$). The model fit (sum of all RAD51 polymers, green solid line) is equal to $\sum_1^n iS_i(t)$. **B-C**. As in **A**, except ssDNA dN-17 (B) and ssDNA dN-14 (C) experimental data and predicted polymers are shown. **D**. Experimental data of dsDNA dN-50p from Fig. 3.2D (dotted lines) and predicted polymers with defined length on each single DNA oligo using the model from Fig. 3.3B (solid lines). Predicted polymer curves were calculated by multiplying the time-dependent concentration of a specific n-mer by its length (i.e. $nD_n(t)$). The model fit (sum of all RAD51 polymers, green solid line) is equal to $\sum_1^n iD_i(t)$. **E-F**. As in **D**, except dsDNA dN-11p (E) and dsDNA dN-8p (F) experimental data and predicted polymers are shown.

3.4.3 Development of n-mers Models

A limitation of the dimer models is that they assume, a priori, that RAD51 in solution exists as either monomers or dimers. This assumption was based on published observations [217] and was used primarily to simplify the mathematical model. However, if RAD51 forms dimers in solution, it is likely that it can form polymers and the degree of solution polymerisation would be expected to be dependent on the concentration of RAD51 relative to the $\mathbf{K_D}$ of the RAD51 protomer-protomer interaction.

Therefore, an extended model was formulated that allowed for RAD51 to form longer polymers in solution (up to a maximum length of 16). These models are collectively referred to as the **n-mer models** (section 2.10). I approached this by first developing an equilibrium sub-model to describe RAD51 polymerisation in solution in the absence of DNA, and then combining this model with kinetic sub-models describing RAD51 polymerisation on ssDNA and dsDNA.

The equilibrium sub-model is a non-linear ODE system describing the formation of RAD51 polymers in solution up to a maximum length of 16 as a function of RAD51 monomer concentration (section 2.10). Importantly, I assume RAD51 polymerisation in solution is in equilibrium because the RAD51 samples for SPR were prepared $\approx$ 20 minutes before the injection. This model allows any RAD51 n-mer ($1 \leq n < 16$) to associate with any other RAD51 m-mer ($1 \leq m < 16$) to form an $(m+n)$-mer ($1 < m+n \leq 16$). In addition, any RAD51 k-mer can fall apart in every possible combination of m-mers and n-mers ($k = n + m$) (e.g. a pentamer can fall apart to form a monomer and a tetramer, or a dimer and a trimer). This model is used to predict the concentrations of RAD51 polymers in solution prior to the injection on the DNA-coated SPR chips, and depends on the fit parameter $\mathbf{K_D}$ (i.e. the RAD51 promoter-promoter dissociation constant). The predicted concentrations are then inserted into the kinetic sub-models to describe RAD51 polymerisation on

ssDNA and dsDNA.

I then developed kinetic sub-models that describe the sequential formation of RAD51 polymers on ssDNA and on dsDNA. I started from the simplest possible ssDNA and dsDNA kinetic sub-models that describe RAD51 polymerisation on DNA via 3 fit parameters: $\mathbf{k_p}$ (the polymerisation rate constant describing the adsorption and elongation of RAD51 n-mers on DNA up to a maximum length of 16), $\mathbf{K_D}$ (the RAD51 protomer-protomer dissociation constant), and $\mathbf{k_s}$ (the single RAD51 protomer dissociation rate constant and the reverse adsorption rate constant for all RAD51 n-mers). Given that it was not possible to determine whether RAD51 displays both faster elongation and faster adsorption on ssDNA compared to dsDNA using the dimer models (section 3.4.2), I decided in the n-mer models to merge the RAD51 adsorption forward rate constant and the elongation forward rate constant into a single rate constant, $\mathbf{k_p}$. I also included stoichiometric factors to account for the different ways in which a RAD51 n-mer can bind to a sequence of n DNA binding sites on a DNA oligo ($1 < n \leq 16$). I then generated further kinetic sub-models by either 1) sequentially increased the number of different adsorption reverse rate constants, or 2) varying the reactions regulated by specific adsorption reverse rate constants (section 2.10.4). Using ABC-SMC, each ssDNA kinetic sub-model was fit with the equilibrium sub-model to the ssDNA dN-X data, and each dsDNA kinetic sub-model was fit with the equilibrium sub-model to the dsDNA dN-Xp data. The SSRs were used to then select the simplest best fit ssDNA kinetic sub-model (Figs. 3.8A, 2.4A, B.4, B.5) and dsDNA kinetic sub-model (Fig. 3.8B, 2.4B, B.6, B.7).

The chosen ssDNA kinetic sub-model (F, Table 2.4, Figs. 2.4A, 3.8A) consists of the polymerisation rate constant ($\mathbf{k_p}$), an unstable reverse rate constant ($\mathbf{k_u}$), a quasi-stable reverse rate constant ($\mathbf{k_q}$) and a stable reverse rate constant ($\mathbf{k_s}$). $\mathbf{k_u}$ describes the dissociation of unstable nuclei (1-3 RAD51 molecules), $\mathbf{k_q}$ describes

the dissociation of quasi-stable nuclei (4 RAD51 molecules), and $\mathbf{k_s}$ here only describes the dissociation of single RAD51 protomers (i.e. RAD51 monomers of a RAD51 polymer) (Figs. 3.7A, 3.8A). Similarly, the chosen kinetic sub-model for dsDNA (D, Table 2.4, Figs. 2.4B, 3.8B) consists of $\mathbf{k_p}$, $\mathbf{k_u}$ and $\mathbf{k_s}$, but without the need for $\mathbf{k_q}$ to describe dissociation of quasi-stable nuclei (Figs. 3.7C, 3.8B). For dsDNA, $\mathbf{k_u}$ describes the rate of dissociation of RAD51 monomers not bound to a RAD51 polymer (Fig. 3.8B).

3.4.4 Fitting the n-mers Models

I also used ABC-SMC to fit these kinetic sub-models together with the equilibrium sub-model simultaneously to the ssDNA dN-X data (Fig. 3.7B) and the dsDNA dN-Xp data (Fig. 3.7D) and obtained values for the model parameters (Table 3.2, Fig. 3.9A-B).

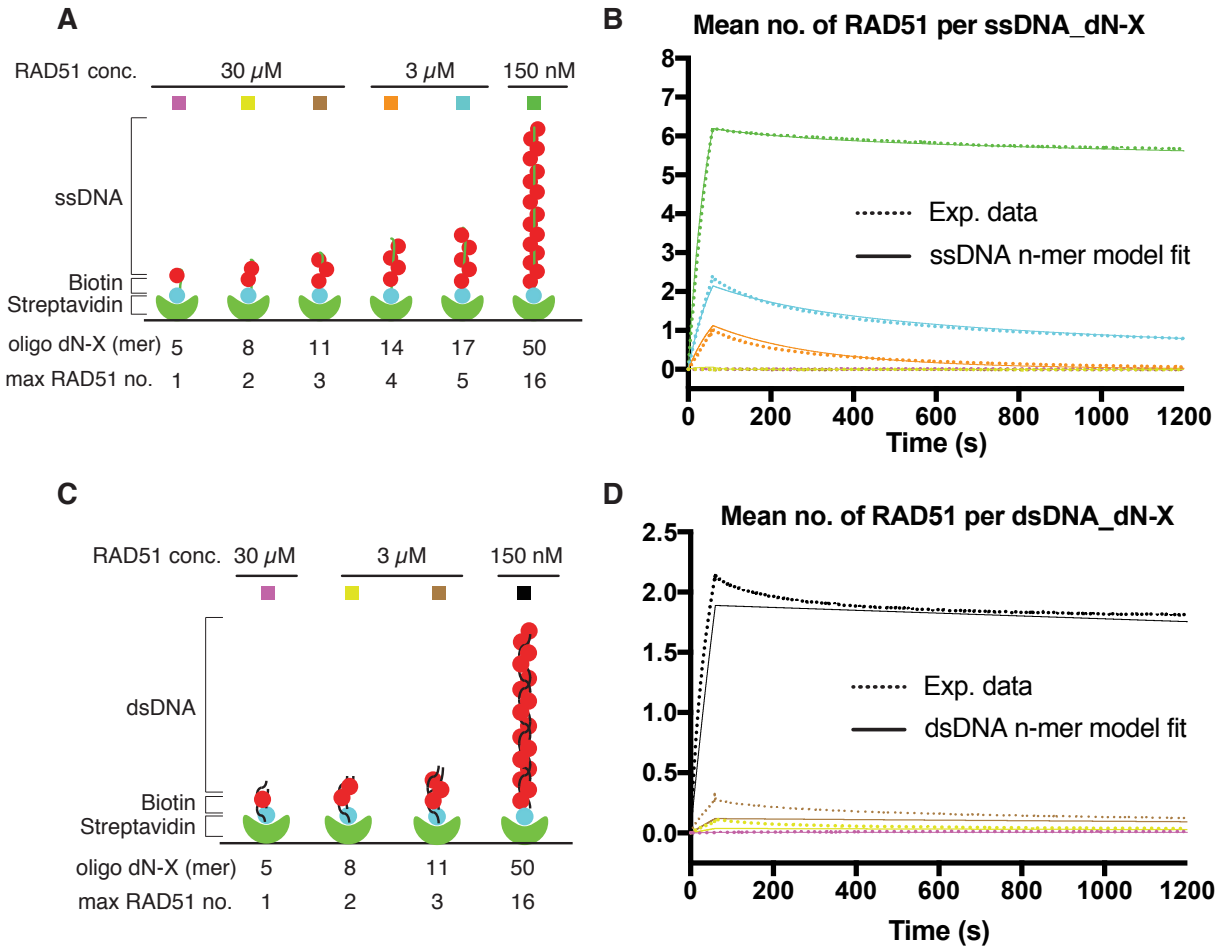


Figure 3.7: WT RAD51 binding kinetics to varying length of ssDNA and dsDNA, with corresponding n-mer model fits. **A** (same as Fig. 3.2A). Biotinylated ssDNA oligos of indicated lengths were separately immobilised onto SPR CM5 chips via biotin-streptavidin interaction, and WT RAD51 was injected at the indicated concentrations to measure association and dissociation kinetics. The predicted number of maximum RAD51 protomers per oligo was calculated using the structurally determined RAD51 : nucleotide/base pair stoichiometry of 1:3 [113]. **B**. The dotted and solid curves respectively show the binding of RAD51 to the ssDNA oligos measured by SPR (same data as in Fig. 3.2B) and the ssDNA n-mer model fits. **C** (same as Fig. 3.2C). Biotinylated dsDNA oligos of indicated lengths were separately immobilised onto SPR CM5 chips via biotin-streptavidin interaction, and WT RAD51 was injected at the indicated concentrations to measure association and dissociation kinetics. The predicted number of maximum RAD51 protomers per oligo was calculated using the structurally determined RAD51 : nucleotide/base pair stoichiometry of 1:3 [113]. **D**. The dotted and solid curves respectively show the binding of RAD51 to the dsDNA oligos measured by SPR (same data as in Fig. 3.2D) and the dsDNA n-mer model fits. **B, D**. The sensorgrams were processed and normalised as described in section 2.7. Normalised sensorgrams represent the mean number of RAD51 molecules per DNA oligo.

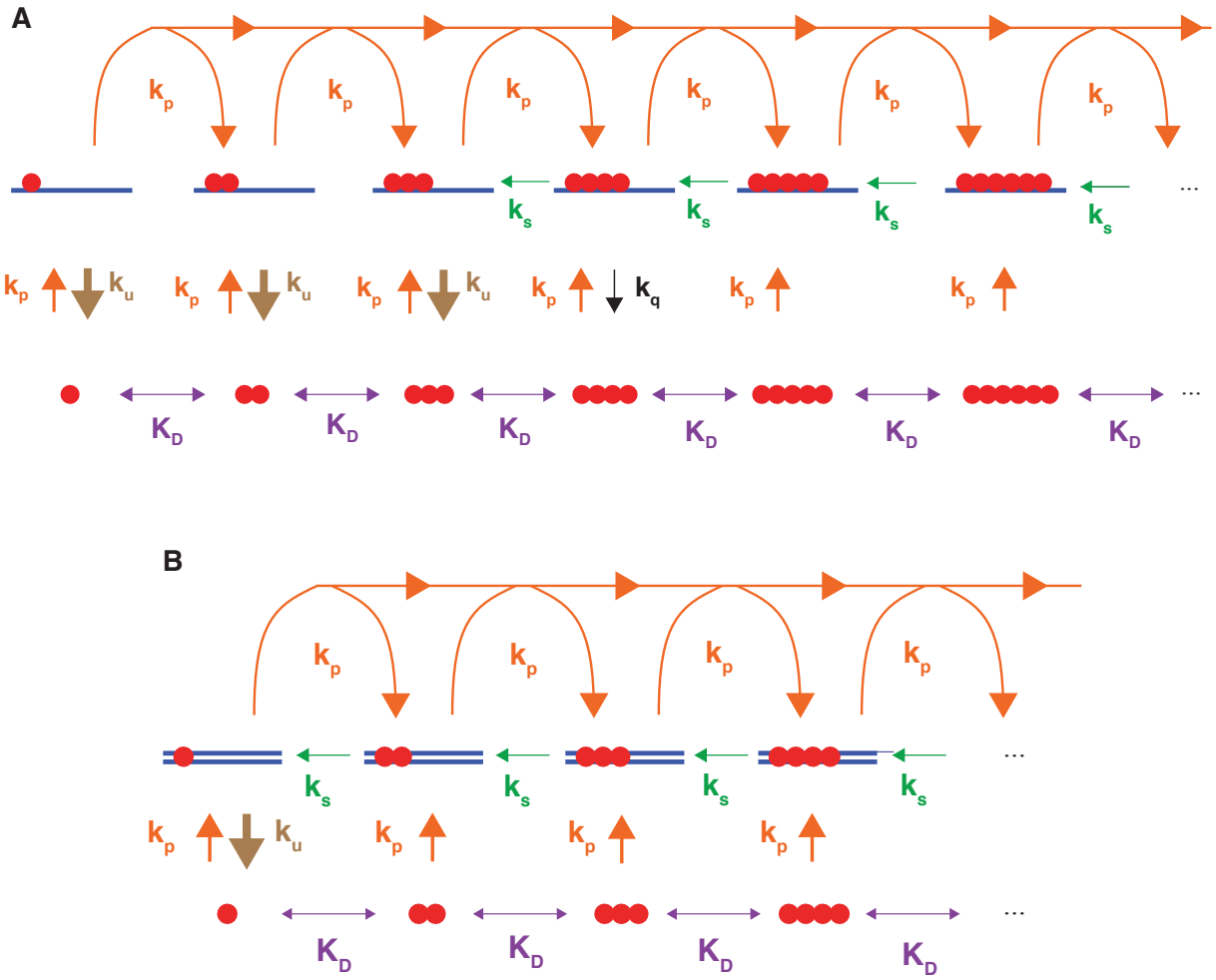


Figure 3.8: Kinetic representations of the ssDNA and dsDNA n-mer models. **A**. Kinetic representation of the ssDNA n-mer model describing RAD51 polymerisation in solution and on ssDNA, consisting of five parameters: $\mathbf{K_D}$, $\mathbf{k_p}$, $\mathbf{k_u}$, $\mathbf{k_q}$ and $\mathbf{k_s}$. **B**. Kinetic representation of the dsDNA n-mer model describing RAD51 polymerisation in solution and on dsDNA, consisting of five parameters: $\mathbf{K_D}$, $\mathbf{k_p}$, $\mathbf{k_u}$ and $\mathbf{k_s}$. **A-B**. $\mathbf{K_D}$ is derived from the equilibrium sub-model and is the same in the ssDNA and dsDNA n-mer models (Table 3.2). Purple arrows depict a simplified cartoon representation of RAD51 polymerisation in solution. The model allows for any RAD51 n-mer ($1 \leq n < 16$) to associate with any other RAD51 m-mer ($1 \leq m < 16$) to form an $(m+n)$-mer ($1 < m+n \leq 16$), and any RAD51 k-mer to dissociate into any combination of n-mers and m-mers ($k = n+m$.

dN-X & dN-Xp $\mathbf{K_D}$, dN-X $\mathbf{k_p}$, $\mathbf{k_q}$, $\mathbf{k_s}$, and dN-Xp $\mathbf{k_p}$ are well determined (Fig. 3.9A). dN-X $\mathbf{k_u}$ and dN-Xp $\mathbf{k_u}$ were undetermined (Fig. 3.9A), due to the fact that no RAD51 binding was observed to the dN-8 and dN-5p oligos (Figs. 3.7A-D, B.1A, B.2A). Finally, I could only determine a well-defined upper bound for the dsDNA stable reverse rate constant (dN-Xp $\mathbf{k_s} < 10^{-3}$) (Fig. 3.9A). Given that both dN-X and dN-Xp $\mathbf{k_u}$ were undetermined and only a lower bound for dN-Xp $\mathbf{k_s}$ was determined, analyses similar to Fig. 3.6 were not carried out for the n-mer models. I identified a nano-molar range RAD51 protomer-protomer dissociation constant (dN-X & dN-Xp $\mathbf{K_D}$ = 1.14 ± 0.5 nM) as a key factor driving rapid RAD51 polymerisation on both ssDNA and dsDNA. This is due to the fact that this $\mathbf{K_D}$ value enables RAD51 to form abundant long polymers in solution at both 150 nM and 3 μM [RAD51] (Fig. 3.10A-B). These RAD51 polymers can then directly adsorb onto DNA and elongate, therefore enabling RAD51 to polymerise faster on the dN-50 (150 nM [RAD51]) compared to the dN-14, dN-17 (3 μM [RAD51]), and faster on the dN-50p (150 nM [RAD51]) compared to the dN-11p and dN-8p (3 μM [RAD51]), despite a 20-fold lower RAD51 injection concentration for dN-50 and dN-50p.

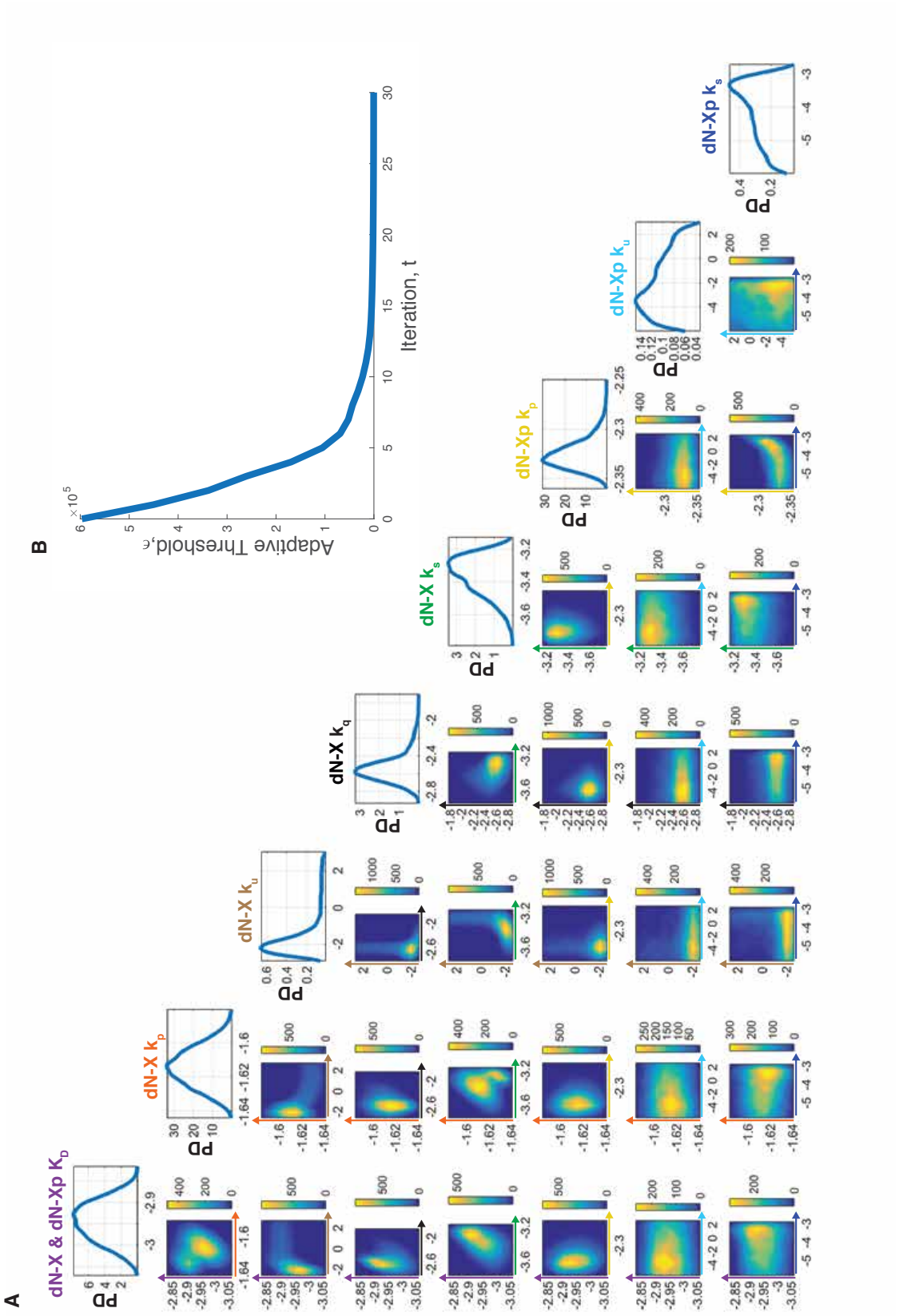

A
dN-X & dN-Xp K_D
dN-X k_p
dN-X k_u
dN-X k_q
dN-X k_s
dN-Xp k_p
dN-Xp k_u
dN-Xp k_s
PD
B
Adaptive Threshold, ε
Iteration, t
$\times 10^5$

Figure 3.9

Figure 3.9: ABC-SMC probability densities for the ssDNA and dsDNA n-mer model fit parameters for one repeat of dN-X and dN-Xp datasets using WT RAD51. **A**. ABC-SMC of the simultaneous WT RAD51 ssDNA (Fig. 3.7) and dsDNA model (Fig. 3.8) fits for the ssDNA dN-X (Fig. 3.7A) and dsDNA dN-Xp (Fig.3.7C) SPR data. Heat maps indicate particle frequencies and describe any pair-wise correlations between model parameters (dN-X & dN-Xp $\mathbf{K_D}$, dN-X $\mathbf{k_p}$, $\mathbf{k_u}$, $\mathbf{k_q}$, $\mathbf{k_s}$, dN-Xp $\mathbf{k_p}$, $\mathbf{k_u}$, $\mathbf{k_s}$). All heat map axes are in log-log scale and are labelled according to the corresponding parameter colour: dN-X & dN-Xp $\mathbf{K_D}$ (purple), dN-X $\mathbf{k_p}$ (orange), dN-X $\mathbf{k_u}$ (brown), dN-X $\mathbf{k_q}$ (black), dN-X $\mathbf{k_s}$ (green), dN-Xp $\mathbf{k_p}$ (yellow), dN-Xp $\mathbf{k_u}$ (cyan), dN-Xp $\mathbf{k_s}$ (blue). A high particle frequency corresponds to a good model fit to the data for the specified parameter pair. The probability densities for individual parameters are presented along the top left diagonal and were estimated using a Gaussian kernel. Particles were initialised via log-uniform priors with the following lower and upper bounds: 10^{-4} < dN-X & dN-Xp $\mathbf{K_D}$ < 10^{-1}, 10^{-4} < dN-X $\mathbf{k_p}$ < 10^4, 10^{-6} < dN-X $\mathbf{k_u}$ < 10^3, 10^{-6} < dN-X $\mathbf{k_q}$ < 10^3, 10^{-6} < dN-X $\mathbf{k_s}$ < 10^3, 10^{-4} < dN-Xp $\mathbf{k_p}$ < 10^4, 10^{-6} < dN-Xp $\mathbf{k_u}$ < 10^3, 10^{-6} < dN-Xp $\mathbf{k_s}$ < 10^3. PD: probability density. **B**. Adaptive threshold for the ABC-SMC simulation. The plateau at 30 iterations means the model has reached a best fit to the data. The simulation was carried out using MATLAB (30 iterations, 10000 particles).

Table 3.2: Fit parameter values for the ssDNA and dsDNA n-mer models. The ssDNA and dsDNA n-mer models were simultaneously fit to the ssDNA dN-X data (ssDNA dN-8, dN-14, dN-17 and dN-50) and the dsDNA dsDNA dN-Xp data (dN-5p, dN-8p, dN-11p and dN-50p), respectively. Mean values ± 1 SD of 3 mode ABC-SMC particles derived from model fits of 3 dN-X and dN-Xp repeats (n = 3). NA: not applicable. [Undetermined]: both dN-X and dN-Xp $\mathbf{k_u}$ were undetermined ($\mathbf{k_u}$ values ranging from 10^{-5} to 10^{2} generated similar model fits to the SPR data (Fig. 3.9A)).

Fit Parameter	ssDNA dN-X	dsDNA dN-Xp
$\mathbf{K_D}$	1.14 ± 0.5 nM	1.14 ± 0.5 nM
$\mathbf{k_p}$	(2.61 ± 0.18) × 10^{-2} /μM/s	(4 ± 0.6) × 10^{-3} /μM/s
$\mathbf{k_u}$	[Undetermined]	[Undetermined]
$\mathbf{k_q}$	(3.6 ± 1) × 10^{-3} /s	NA
$\mathbf{k_s}$	(4.3 ± 0.6) × 10^{-4} /s	<10^{-3} /s

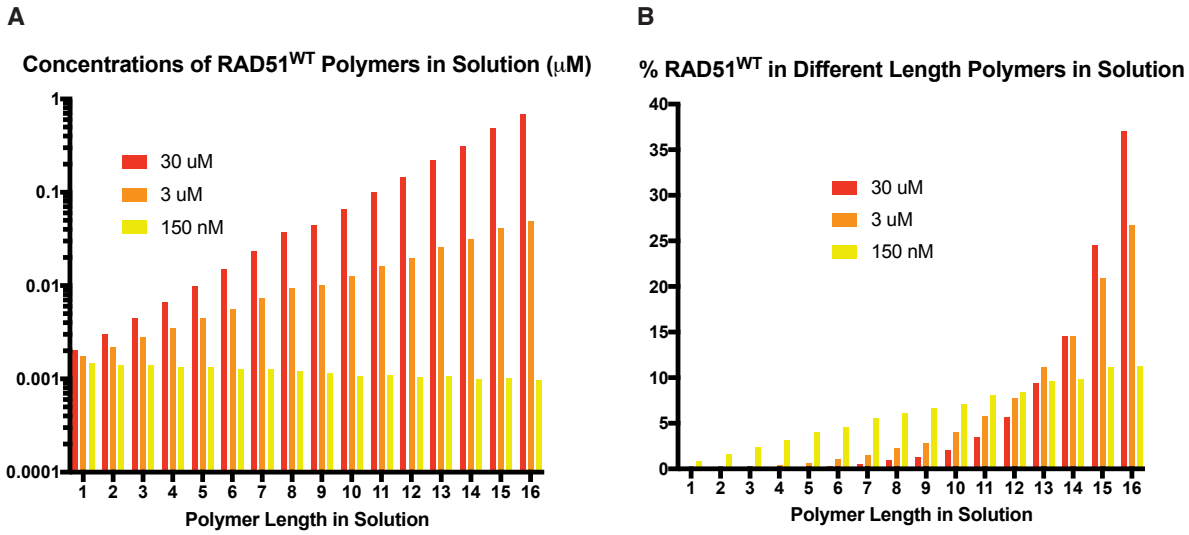


Figure 3.10: WT RAD51 polymerisation in solution. **A**. Predicted concentrations of WT RAD51 polymers in solution at equilibrium as a function of WT RAD51 monomer concentration (150 nM, 3 μM, 30 μM), prior to RAD51 injection onto DNA-coated SPR CM5 chips. **B**. Predicted % RAD51^{WT} within each polymer state in solution at equilibrium as a function of WT RAD51 monomer concentration (150 nM, 3 μM, 30 μM), prior to RAD51 injection onto DNA-coated SPR CM5 chips. A-B. Polymer concentrations were predicted using the WT RAD51 equilibrium sub-model. The $\mathbf{K_D}$ was fixed to the mean value (Table 3.2) (i.e. $\mathbf{K_D}$ = 1.14 nM), and the concentration of RAD51 monomer concentration was varied accordingly. In panel **B**, the % RAD51^{WT} values were calculated by multiplying the polymer concentrations by their respective polymer length, and dividing each value by the total RAD51 monomer concentration (i.e. % $\text{RAD51}^{\text{WT}}{}_n$ = nR_n/R_1). R_1 (i.e. the RAD51 SPR injection concentration, [RAD51]) was calculated via coomassie staining image quantification using a BSA standard curve. **A-B**. It is important to mention that the assumed maximum polymer length in solution (16) leads to an overestimation of the concentration of each n-mer in solution, given that RAD51 is likely to form polymers of length greater than 16 at 150 nM, 3 μM and 30 μM RAD51 monomer concentration. However, this overestimation is likely to not alter the overall conclusions of section 3.4.4. See also section 3.5.

Importantly, I also identified a 6.5-fold higher ssDNA polymerisation forward rate constant (dN-X $\mathbf{k_p}$ = (2.61 ± 0.18) × 10^{-2} /μM/s) compared to dsDNA (dN-Xp $\mathbf{k_p}$ = (4 ± 0.6) × 10^{-3} /μM/s), explaining the overall faster RAD51 polymerisation on ssDNA compared to dsDNA (Fig. 3.8A-B, Table 3.2). Overall, these analyses using the n-mer models suggest RAD51 elongation and/or adsorption is faster on ssDNA compared to dsDNA, as previously suggested using the dimer models. In addition, the analyses suggest a high protomer-protomer affinity enables RAD51 to nucleate and elongate effectively on ssDNA and dsDNA even at low concentrations of monomeric RAD51.

3.5 Discussion

In this chapter, the RAD51 binding kinetics to ssDNA and dsDNA of variable length were analysed using SPR to derive data-driven ODE models of RAD51 polymerisation on ssDNA and dsDNA. I observed faster RAD51 association to ssDNA compared to dsDNA, and developed two sets of ODE models to further characterise the mechanisms of RAD51 polymerisation on ssDNA and dsDNA: the dimer models and the n-mer models. For the dimer models, I assume polymerisation mainly via the addition of pre-formed RAD51 dimers. Although this assumption has experimental support [217], it is possible that RAD51 polymerisation on DNA occurs also via the addition of pre-formed oligomers of variable length (e.g. trimers, tetramers, pentamers, etc.) [12]). The n-mer models were developed to test whether allowing RAD51 polymerisation both in solution and on DNA would alter the conclusions derived from the dimer models.

Both the ssDNA dimer model and the ssDNA n-mer model suggests 1) RAD51 polymerisation on ssDNA can be described via the sequential formation of an unstable nucleus (1 - 3 protomers), a quasi-stable nucleus (4 protomers) and a stable nucleus (5 protomers). In contrast, both the dsDNA dimer model and the dsDNA n-mer model suggest RAD51 has a minimal nucleation length of two monomers on dsDNA. Importantly, both the dimer models and the n-mer models highlighted that RAD51 polymers elongate and/or adsorb faster on ssDNA compared to dsDNA, despite requiring more RAD51 molecules for stable nucleation on ssDNA compared to dsDNA (five versus two).

An important assumption of the n-mer equilibrium sub-model is the truncation of RAD51 polymer lengths in solution to a maximum length of 16. This was chosen to simplify the development of the equilibrium sub-model. However, in my experimental system RAD51 is likely to form polymers of lengths greater than 16, given that

tetramers and pentamers have been detected at RAD51 monomer concentrations as low as 12.5 nM [12]. Hence, the assumed maximum polymer length in solution leads to an overestimation of the concentration of each n-mer in solution, as mentioned in Fig. 3.10. In the case of an even more elaborate equilibrium sub-model model that allows for RAD51 polymerisation up to a maximum length of e.g. 32, the conclusions of the n-mer models would remain unaltered as higher dN-X $\mathbf{k_p}$ and dN-Xp $\mathbf{k_p}$ values would be able to compensate for lower concentrations of RAD51 n-mers in solution to adequately fit the dN-X and dN-Xp SPR data, respectively.

Furthermore, my estimated RAD51 protomer-protomer dissociation constant ($\mathbf{K_D}$ = 1.14 ± 0.5 nM) is lower compared to the $\mathbf{K_D}$ reported in [12] ($\mathbf{K_D} \approx$ 17 nM). While both values point to a high affinity RAD51 protomer-protomer dissociation constant in the presence of ATP and Ca^{2+}, the apparent discrepancy may be due to assuming a maximum polymer length in solution of 16, while the stochastic model in [12] did not truncate the maximum polymer length in solution. This may have lead to an overestimation of the ($\mathbf{K_D}$) using the n-mer models. However, the reported value of $\mathbf{K_D} \approx$ 17 nM may also be biased because all stochastic model fits in [12] were carried out on single molecule fluorescence microscopy experiments using a single RAD51 concentration (12.5 nM) and were not further validated using different RAD51 concentrations. On the other hand, my reported $\mathbf{K_D}$ = 1.14 ± 0.5 nM may be more accurate in this regard as it explains the SPR data in the presence of a 200-fold variation of RAD51 concentration (150 nM, 3 μM and 30 μM, Fig. 3.7). Interestingly, a study reporting pressure perturbation fluorescence spectroscopy data analysed with a model similar to that of [12] described a RAD51 $\mathbf{K_D}$ of $\approx$ 1.6 nM in the presence of ATP and Ca^{2+}, similar to my reported value [220]. However, the study also reported a $\approx$ 100-fold decrease in the $\mathbf{K_D}$ to $\approx$ 0.018 nM in the presence of ATP, Ca^{2+} and ssDNA, suggesting the association to ssDNA may further enhance the protomer-protomer interaction affinity. Nonetheless, the $\mathbf{K_D}$ values re-

ported in [12, 220] and in this chapter all point to a high affinity protomer-protomer interaction and support the idea that pre-formed RAD51 polymers may directly bind from solution to DNA.

Altogether, these results suggest that RAD51 can distinguish ssDNA from dsDNA *in vitro* in the absence of additional protein co-factors (e.g. BRCA2, PALB2, RAD52), potentially via faster elongation and/or adsorption of RAD51 polymers on ssDNA, and are in broad agreement with single molecule DNA studies using magnetic tweezers [11] or optical tweezers [12] showing faster association to ssDNA compared to dsDNA. The empirically derived minimal stable nucleation length for ssDNA (i.e. 5 RAD51 molecules) is in agreement with fitted RAD51 nucleation lengths from other studies [11, 109], while the nucleation length for dsDNA (i.e. 2 RAD51 molecules) is in agreement with the predicted nucleation length in a previous study also using optical tweezers [110].

The observation that ssDNA is more flexible compared to dsDNA [149, 148, 150] raises an interesting question: does DNA flexibility impact the ability of RAD51 to discriminate between ssDNA and dsDNA? To tackle this question, in the following chapter I measure the RAD51 binding kinetics to ssDNA molecules of variable flexibility using SPR and fit the dimer and n-mer models from Figs. 3.3 and 3.8 to these data.

Chapter 4

RAD51 Polymerises Faster on Flexible DNA

4.1 Introduction

DNA is a highly dynamic structure that can both bend and twist, although the impact of DNA flexibility on DNA repair remains poorly characterised.

As mentioned in section 1.4.1, Human HMGB DNA-binding proteins can bend and compact DNA, and play critical but poorly understood roles in transcription, chromatin remodelling and DNA repair [221]. In particular, HMGB1 has been suggested to promote repair of psoralen and UV-induced DNA interstrand-crosslinks *in vivo* by bending dsDNA at sites of photoproducts or pyrimidine dimers and enhancing the recruitment of nucleotide excision repair factors XPA and XPC-RAD23 [162, 167]. Furthermore, the human HMGB2 and the *S. cerevisiae* HMO1 proteins were shown via optical tweezers to increase dsDNA flexibility by $\approx$ 10 - fold [163, 164], suggesting that increased DNA flexibility may enhance the recruitment of DNA repair factors. Nonetheless, how DNA flexibility may directly impact the DNA-binding kinetics of DNA repair factors, such as RAD51, remains unanswered.

The flexibility of a DNA molecule can be characterised by its persistence length (L_p), a mechanical parameter quantifying polymer stiffness: the higher the persistence

length, the stiffer the polymer. Interestingly, a key structural difference between ssDNA and dsDNA is flexibility: ssDNA is known to be more flexible compared to dsDNA. In the presence of monovalent or divalent salt, dsDNA is shown to display an L_p of 30 - 55 nm [149, 150], while ssDNA is much more flexible, with an L_p of 1.5 - 3 nm [146, 147, 148].

In this chapter, I investigate the impact of DNA bending flexibility on RAD51 polymerisation kinetics and present the Bend-To-Capture (BTC) mechanism to explain how RAD51 may preferentially associate to ssDNA compared to dsDNA. The mechanism proposes that the higher flexibility of ssDNA compared to dsDNA leads to faster RAD51 association to ssDNA. To test this hypothesis, I analysed the RAD51 binding kinetics to ssDNA molecules of variable flexibility using SPR and fit the ssDNA dimer model (Fig. 3.3A) and ssDNA n-mer model (Fig. 3.8A) to these data. I conclude that RAD51 adsorbs and/or elongates faster on flexible ssDNA compared to stiff ssDNA or dsDNA.

4.2 The Bend-To-Capture Mechanism

Overall, the analyses from chapter 3 suggested RAD51 has a shorter minimum nucleation length on dsDNA compared to ssDNA (2 versus 5), but adsorbs and/or elongates significantly faster on ssDNA compared to dsDNA. I speculate that the difference in RAD51 polymerisation speed is due to the higher flexibility of ssDNA compared to dsDNA. Explicitly, I propose the BTC mechanism to explain how DNA flexibility can impact RAD51 polymerisation kinetics (Fig. 4.1). A free RAD51 monomer or polymer in solution would need to generate two sequential, non-covalent interactions to be incorporated into the growing polymer: the interaction with the exposed interface of an existing RAD51 DNA-bound polymer and with the exposed scaffold DNA. To avoid steric clashes, I reason that each of these interactions requires naked DNA immediately next to the existing RAD51 DNA-bound

polymer in a configuration that can bend away from the preferred direction of the polymer. In this way, flexible DNA is expected to explore more conformations compatible with the further addition of RAD51 per unit time.

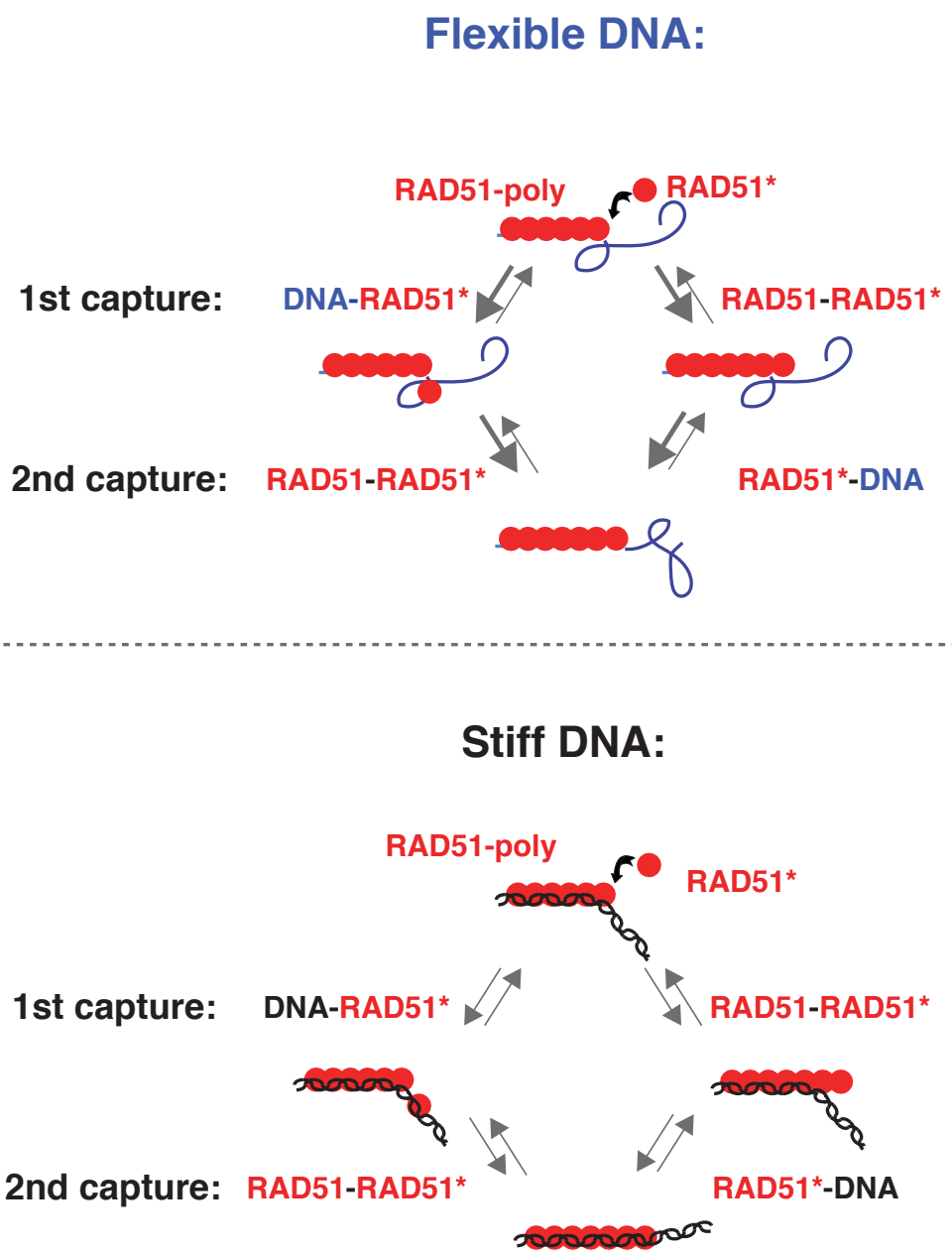


Figure 4.1: The Bend-To-Capture (BTC) mechanism. RAD51 generates two sequential, non-covalent interactions (1st capture, 2nd capture: a RAD51 protomer-protomer interaction and a RAD51-DNA interaction, or vice versa) faster on flexible DNA (depicted as ssDNA) compared to stiff DNA (depicted as dsDNA). The free RAD51 monomer or polymer in solution (depicted here as a monomer) to be incorporated into the existing DNA-bound polymer is labelled with an asterisk. RAD51-poly: RAD51 DNA-bound polymer.

4.3 Persistence Length Measurements of DNA molecules using SAXS

To test this notion, I designed an experiment to measure the kinetics of RAD51 binding to DNA oligos of varying flexibility. It has been shown that poly-dT ssDNA, which has been widely used for RAD51 binding assays, is highly flexible, while poly-dA ssDNA is highly stiff due to base stacking interactions [222, 223]. I used SAXS to measure the persistence lengths of the ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p oligos, and found that ssDNA dT-50 is the most flexible oligo, followed by dN-50, dA-50 and dsDNA dN-50p (Fig. 4.2 and Table 4.1).

Table 4.1: Persistence lengths (L_p) fits and 95% confidence intervals (L_p error) derived from fitted Kuhn lengths (L_k) for the ssDNA dN-50, dA-50, dT-50 and the dsDNA dN-50p SAXS plots. $L_p = \frac{L_k}{2}$. It is likely that the relatively low dsDNA dN-50p L_p (7.34 nm) compared to the average dsDNA L_p (30 - 55 nm [149, 150]) is due to the low GC content of the dsDNA dN-50p (36%). All data manipulation and model fitting was done using SasView.

Molecule Name	L_k (nm)	L_p (nm)	L_p error (nm)
ssDNA_dT-50	2.6702	1.3351	0.200185
ssDNA_dN-50	4.5177	2.25885	0.048095
ssDNA_dA-50	9.3488	4.6744	0.13126
dsDNA_dN-50p	14.674	7.337	0.07379

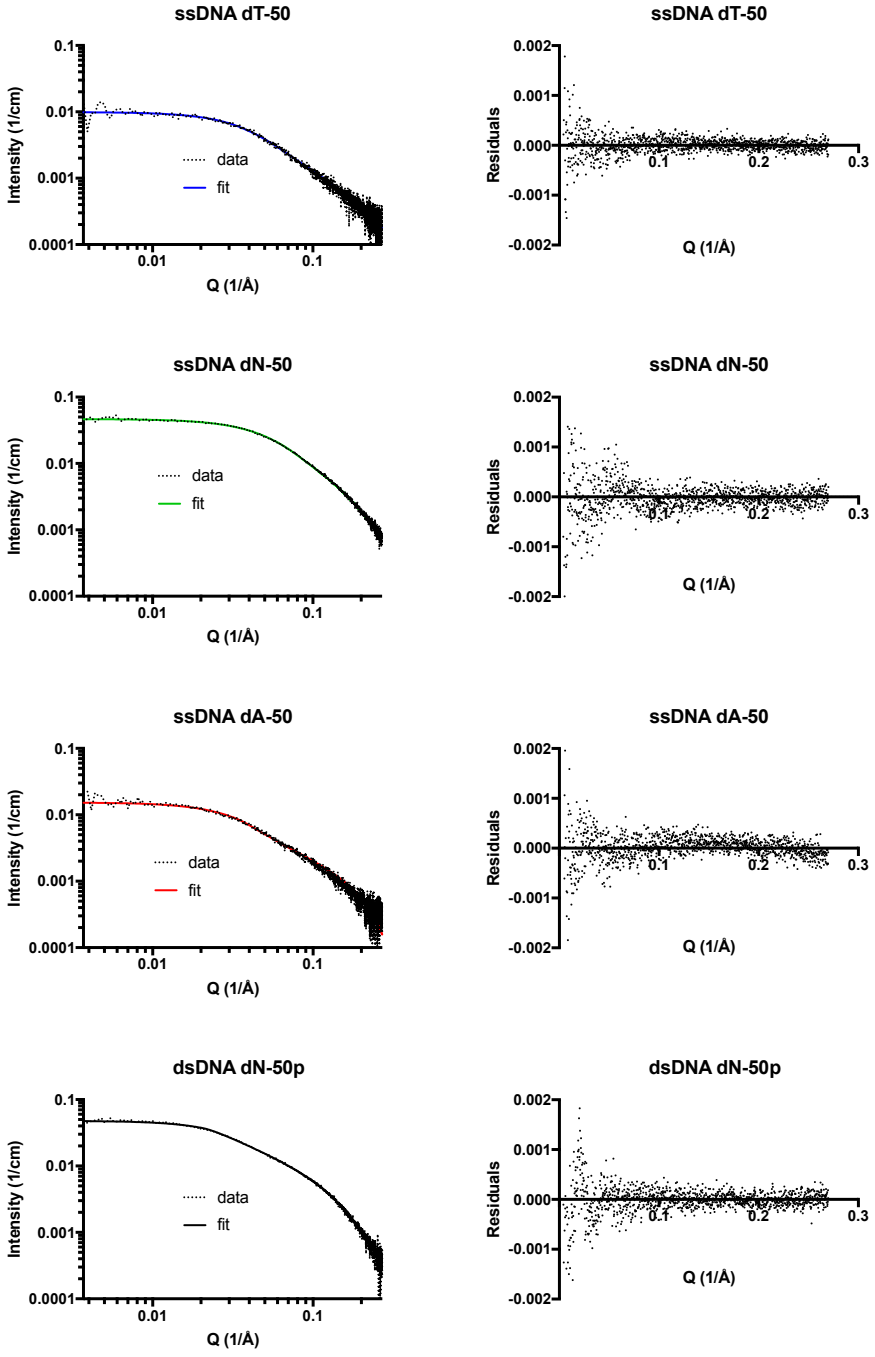


Figure 4.2: The flexibilities of ssDNA dT-50 (**A**), dN-50 (**B**), dA-50 (**C**) and dsDNA dN-50p (**D**) were assessed by SAXS. The left panels show the absolute scattering intensity (1/cm) measured as a function of the scattering vector (1/Å). The flexible cylinder model was fit to the data (fitting range: 0.0037 - 0.27 1/Å) to predict the kuhn lengths. The right panels show the residuals of the flexible cylinder model fits.

4.4 RAD51 Adsorbs and/or Elongates Faster on Flexible DNA

I then measured RAD51 binding kinetics to these DNA oligos and used the ssDNA dimer model (Fig. 3.3A) and the ssDNA n-mer model (Fig. 3.8A) to fit these data. Specifically, I used the lsqcurvefit function in MATLAB to fit the dT-50 and dA-50 curves by individually varying a single parameter within the ssDNA dimer and n-mer models. This allowed me to test for potential correlations between DNA flexibility and any of the individual model parameters.

First, I found that RAD51 indeed displayed faster association to the dT-50 compared to the dN-50 (Fig. 4.3A-B). The dimer model fit suggests this can be explained either via a higher elongation forward rate constant ($\mathbf{k_{el.f}}$) (Fig. 4.3B-C), a higher adsorption forward rate constant ($\mathbf{k_{ad.f}}$) (Fig. 4.4A) or a lower adsorption reverse rate constant ($\mathbf{k_{ad.r}}$) (Fig. 4.4B). Similarly, the n-mer model fit suggests faster association to the dT-50 compared to the dN-50 can be explained by a higher polymerisation rate constant ($\mathbf{k_p}$) (Fig. 4.5B-C).

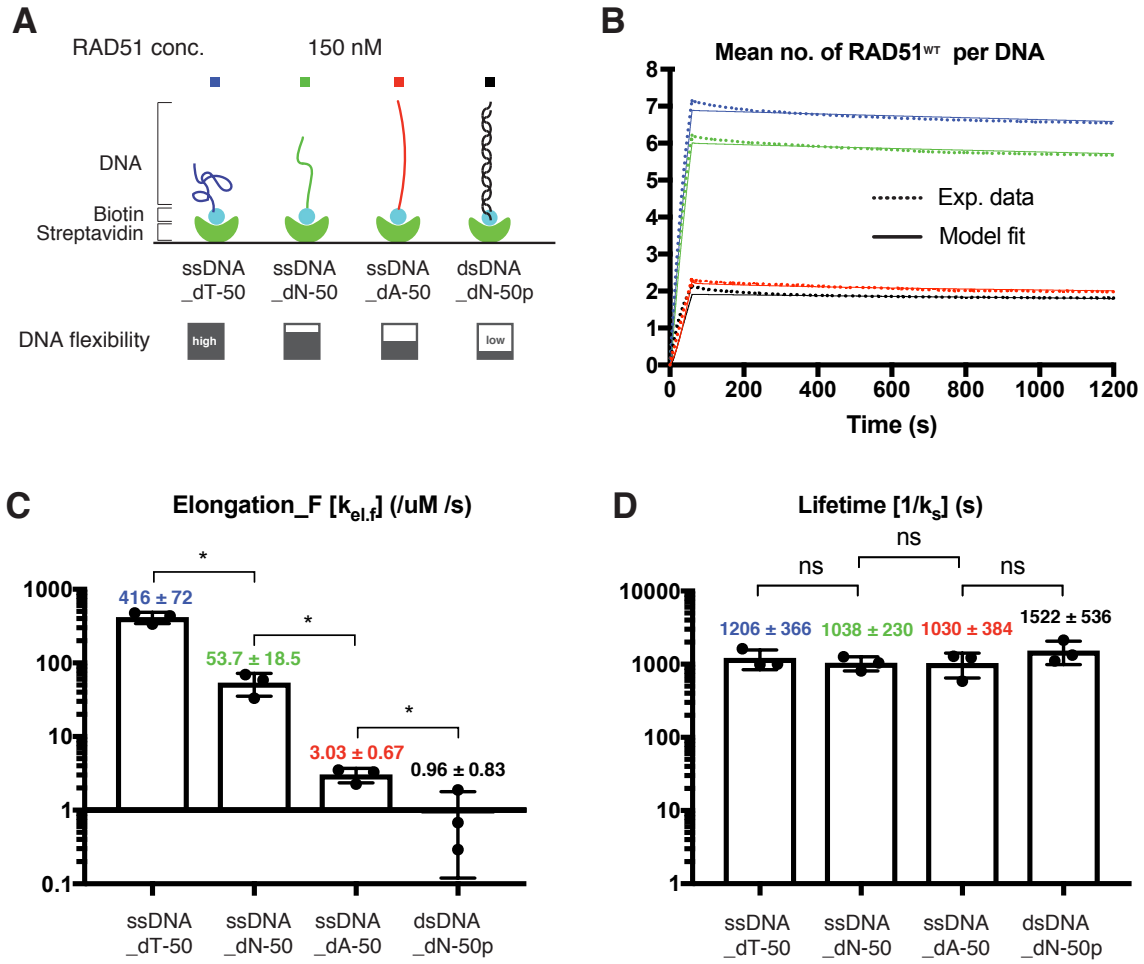


Figure 4.3: WT RAD51 binding kinetics to DNA with varying flexibility and corresponding dimer model fits. **A**. The Bend-To-Capture (BTC) mechanism predicts the efficiency of RAD51 polymerisation on the basis of the relative flexibility of the DNA oligos. **B**. WT RAD51 SPR curves for each DNA oligos (dotted lines) and corresponding dimer model fits (solid lines). **C-D**. Bar plots of fitted $\mathbf{k_{el.f}}$ values (**C**) and $\mathbf{k_s}$ values (**D**) (reported as $1/\mathbf{k_s}$) for each SPR curve. **C**. When fitting the ssDNA dT-50 and dA-50 SPR data, parameters $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.r}}$ and $\mathbf{k_s}$ in the ssDNA dimer model (Fig. 3.3A) were fixed to the mean values as identified in Table 3.1, and $\mathbf{k_{el.f}}$ was fitted using lsqcurvefit (MATLAB). **D**. When fitting the ssDNA dT-50 and dA-50 SPR data, parameters $\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$ and $\mathbf{k_{el.r}}$ in the ssDNA dimer model (Fig. 3.3A) were fixed to the mean values as identified in Table 3.1, and $\mathbf{k_s}$ and $\mathbf{k_{el.f}}$ were fitted using lsqcurvefit (MATLAB). Only the $\mathbf{k_s}$ values are reported here. **C-D**. Mean ± 1 SD of 3 dimer model fits (n = 3). The ssDNA dN-50 and dsDNA dN-50p $\mathbf{k_{el.f}}$ values (**C**) and $\mathbf{k_s}$ (**D**) values are the modes of the parameter distributions derived from the ABC-SMC simulations (Figs. 3.4, 3.5, Table 3.1). **C**: un-paired, one-tailed Mann-Whitney tests. **D**: un-paired, two-tailed Mann-Whitney tests. * $p \leq 0.05$; ns = non-significant.

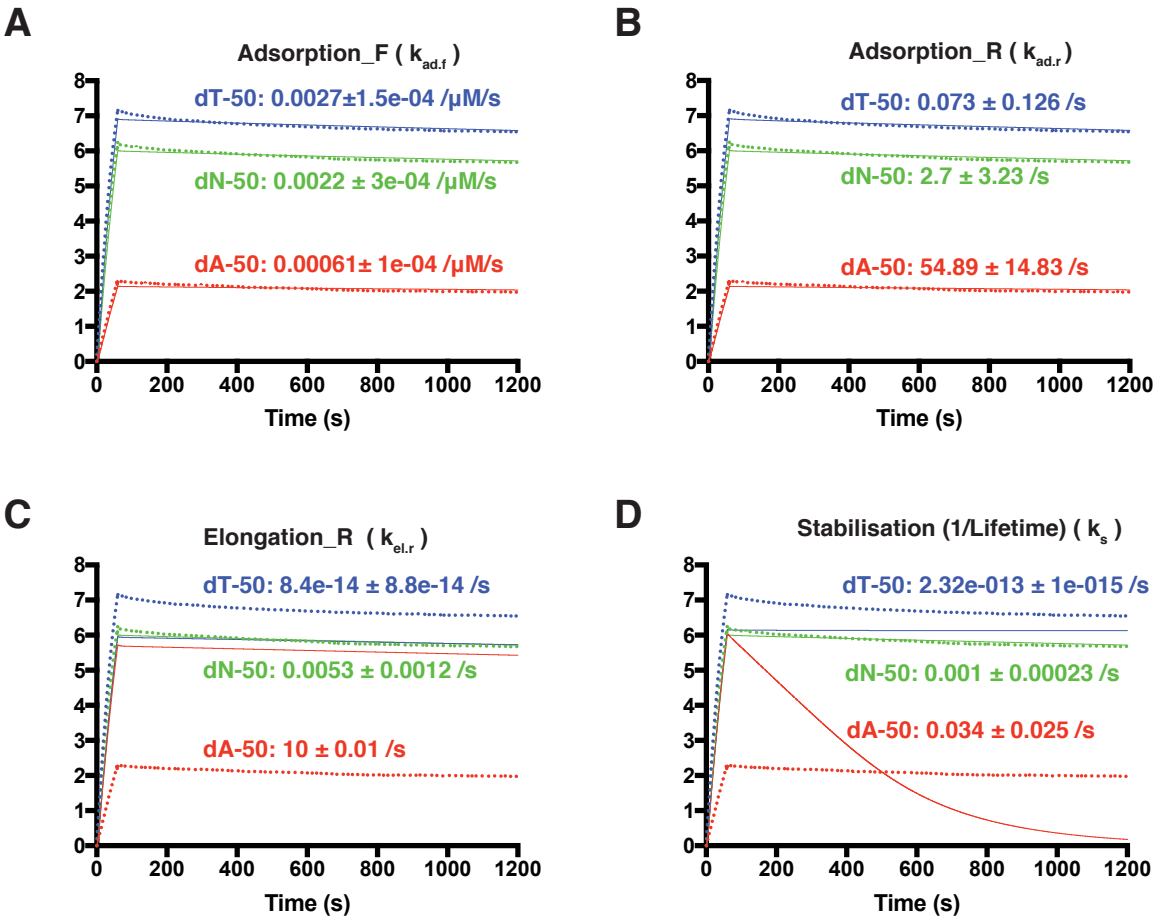


Figure 4.4: **A-D**. RAD51 SPR curves (dotted lines) of the ssDNA dT-50, dN-50 and dA-50 are fit using lsqcurvefit (MATLAB) by varying a single parameter of the ssDNA dimer model (Fig. 3.3A). Each panel shows the ODE model fits (solid lines) based on varying the adsorption forward rate constant ($\mathbf{k_{ad.f}}$) (**A**), the adsorption reverse rate constant ($\mathbf{k_{ad.r}}$) (**B**), the elongation reverse rate constant ($\mathbf{k_{el.r}}$) (**C**), or the stabilisation rate constant ($\mathbf{k_s}$) (**D**). A higher or lower $\mathbf{k_{ad.f}}$ or $\mathbf{k_{ad.r}}$, but not $\mathbf{k_{el.r}}$ or $\mathbf{k_s}$, can lead to accurate model fits.

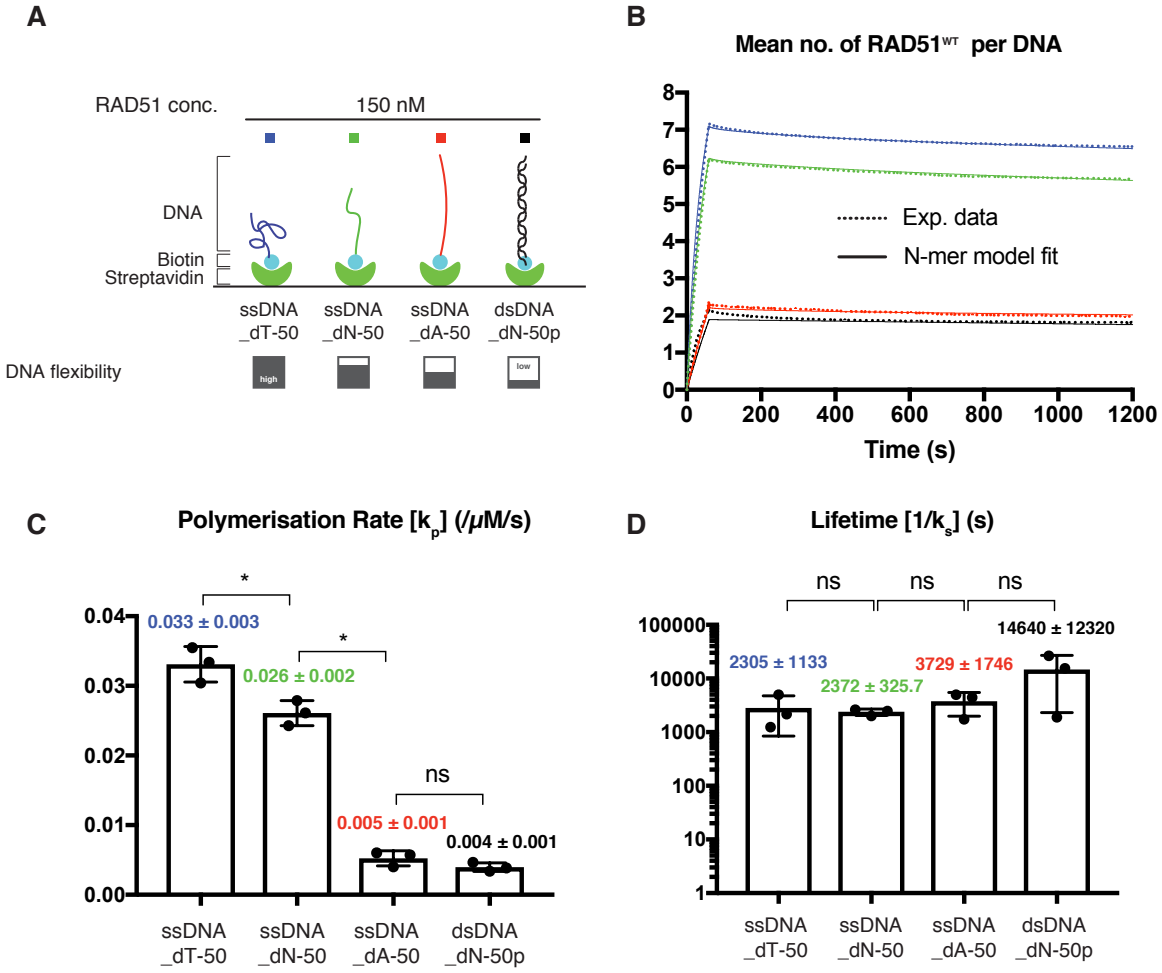


Figure 4.5: WT RAD51 binding kinetics to DNA with varying flexibility and corresponding n-mer model fits. **A**. The Bend-To-Capture (BTC) mechanism predicts the efficiency of RAD51 polymerisation on the basis of the relative flexibility of the DNA oligos. **B**. WT RAD51 SPR curves for each DNA oligos (dotted lines) and corresponding n-mer model fits (solid lines). **C-D**. Bar plots of fitted $\mathbf{k_p}$ values (C) and $\mathbf{k_s}$ (reported as $1/\mathbf{k_s}$) values (D) for each SPR curve. **C**. When fitting the ssDNA dT-50 and dA-50 SPR data, parameters $\mathbf{K_D}$, $\mathbf{k_q}$, $\mathbf{k_s}$ in the ssDNA n-mer model (Fig. 3.8A) were fixed to the mean values as identified in Table 3.2, and $\mathbf{k_p}$ was fitted using lsqcurvefit (MATLAB). **D**. When fitting the ssDNA dT-50 and dA-50 SPR data, parameters $\mathbf{K_D}$ and $\mathbf{k_q}$ in the ssDNA n-mer model (Fig. 3.8A) were fixed to the mean values as identified in Table 3.2, and $\mathbf{k_s}$ and $\mathbf{k_p}$ were fitted using lsqcurvefit (MATLAB). Only the $\mathbf{k_s}$ values are reported here. **C-D**. Mean ± 1 SD of 3 n-mer model fits (n = 3). The ssDNA dN-50 and dsDNA dN-50p $\mathbf{k_p}$ values (**C**) and $\mathbf{k_s}$ (**D**) values are the modes of the parameter distributions derived from the ABC-SMC simulations (Fig. 3.9, Table 3.2). The SD for dN-50p $\mathbf{k_s}$ is high because only a lower bound was determined for this parameter (Table 3.2). **C**: unpaired, one-tailed Mann-Whitney tests. **D**: unpaired, two-tailed Mann-Whitney tests. * p $\leq$ 0.05; ns = non-significant.

Furthermore, RAD51 displayed very slow association to the dA-50, comparable to that of dsDNA dN-50p (Fig. 4.3A-B). The dimer model fit suggests this is explained by a lower elongation forward rate constant (Fig. 4.3B-C), a lower adsorption forward rate constant (Fig. 4.4A) or a higher adsorption reverse rate constant (Fig. 4.4B). Similarly, the n-mer model fit suggests the slower association to dA-50 is explained by a lower polymerisation rate constant ($\mathbf{k_p}$) (Fig. 4.5B-C). Importantly, individually fitting the dimer model elongation reverse rate constant ($\mathbf{k_{el.r}}$) or the dimer model stabilisation rate constant ($\mathbf{k_s}$) could not explain the dT-50 and dA-50 data (Fig. 4.4C-D). Similarly, individually fitting the n-mer model $\mathbf{k_u}$, $\mathbf{k_q}$ or $\mathbf{k_s}$ also failed to explain the dT-50 and dA-50 data (Fig. 4.6A-C).

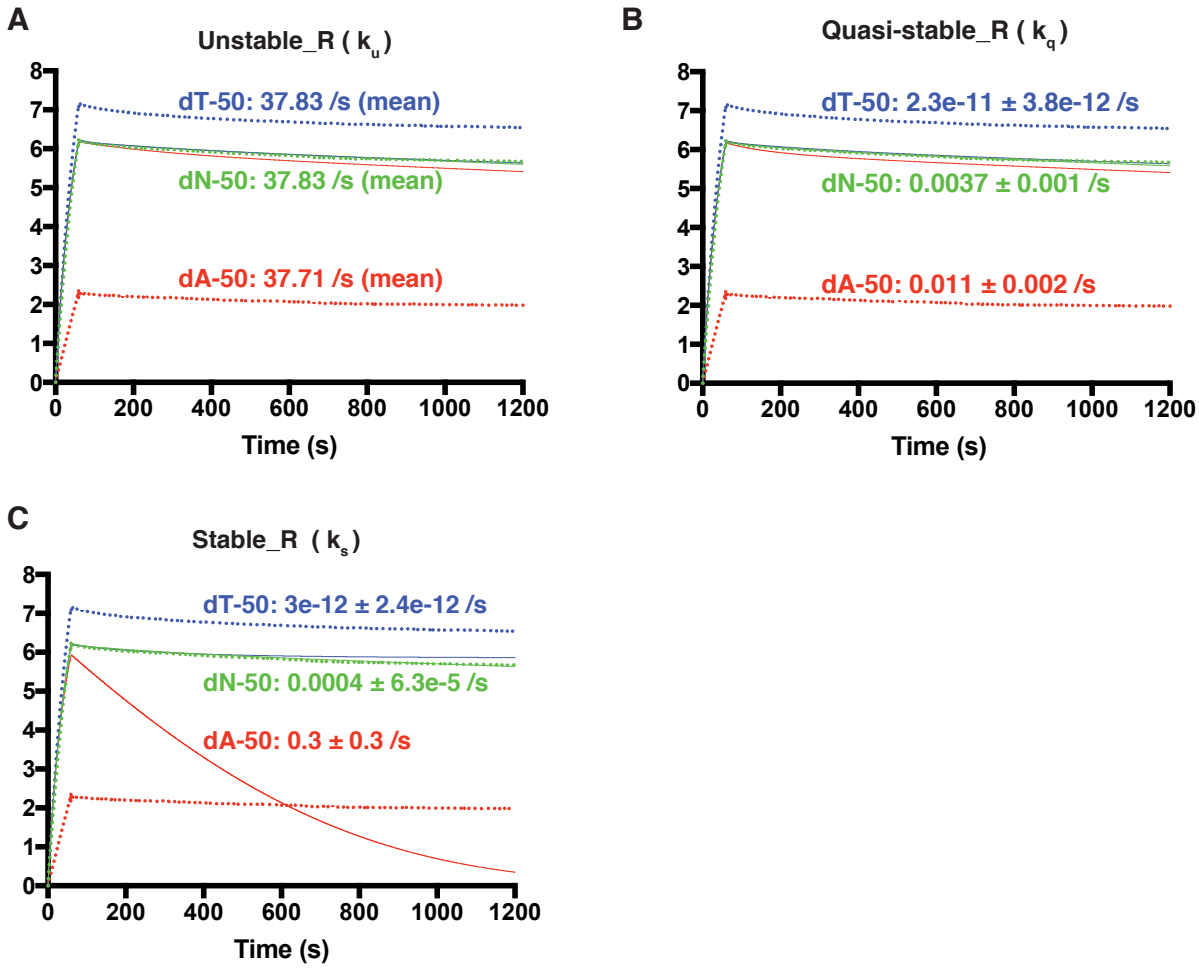


Figure 4.6: **A-C**. RAD51 SPR curves (dotted lines) of the ssDNA dT-50, dN-50 and dA-50 are fit using the ssDNA n-mer model via lsqcurvefit (MATLAB) (solid lines) by individually varying either $\mathbf{k_u}$ (**A**), $\mathbf{k_q}$ (**B**) or $\mathbf{k_s}$ (**C**). Individually varying either $\mathbf{k_u}$, $\mathbf{k_q}$ or $\mathbf{k_s}$ does not lead to accurate model fits. $\mathbf{k_u}$ was undetermined (Table 3.2), and for this reason fit values in **A** are reported only as means.

I also quantified the lifetime of RAD51 by simultaneously fitting both $\mathbf{k_s}$ and $\mathbf{k_{el.f}}$ to the dT-50 and dA-50 datasets with the dimer model (Fig. 3.3A) using lsqcurvefit, while keeping all other parameter values fixed according to the mean values in Table 3.1. I observed that the lifetime of single RAD51 protomers (here defined as $\frac{1}{\mathbf{k_s}}$) was independent of the flexibility of the naked DNA (Figs. 4.3D).

Using the n-mer model and keeping $\mathbf{K_D}$ and $\mathbf{k_q}$ fixed according to the mean values in Table 3.2, I also observed similar RAD51 lifetimes on all four DNA molecules (Fig. 4.5D).

Hence, both the ssDNA dimer model and the ssDNA n-mer model suggest RAD51 displays faster polymerisation on flexible DNA such as ssDNA dT-50, and slower polymerisation on stiff DNA such as ssDNA dA-50 or dsDNA dN-50p. Taken together, these observations suggest that although RAD51 is able to form a stable nucleus with fewer molecules on dsDNA compared to ssDNA (two versus five), it elongates and/or adsorbs more efficiently on ssDNA because of the higher flexibility of ssDNA compared to dsDNA, consistent with the proposed BTC mechanism.

4.5 Discussion

In this chapter, I have presented the BTC mechanism to explain how RAD51 may preferentially associate to ssDNA compared to dsDNA. The SPR data and corresponding dimer and n-mer model fits, together with the SAXS-derived persistence length measurements of the ssDNA (dT-50, dN-50, dA-50) and dsDNA (dN-50p) oligos, support the idea that the higher flexibility of ssDNA promotes faster RAD51 adsorption and/or polymer elongation on ssDNA compared to dsDNA. Overall, these observations point to a previously unappreciated role of DNA flexibility in the recruitment of RAD51 to ssDNA at DSBs.

In contrast to the BTC mechanism, is it possible that RAD51 simply polymerise faster on more extensible DNA? Base stacking interactions are known to pose an energetic barrier to DNA extension, and dT-50 is likely to display less base-stacking interactions compared to dN-50, dA-50 and dN-50p [224, 225]. Therefore, RAD51 protomers may encounter a lower energetic barrier to form the required protein-DNA interactions and extend DNA on dT-50 due to less base-stacking interactions compared to dN-50, dA-50 and dN-50p, potentially driving faster polymerisation on dT-50, followed by dN-50, dA-50 and dN-50p. This may be a viable alternative explanation to the BTC mechanism, suggesting that the extent of DNA base-stacking, rather than bending, determines whether DNA is in a configuration compatible with RAD51 binding (Stack-To-Reject, STR). However, while it has been previously suggested that RAD51 extends dsDNA less than mixed sequence ssDNA ($\approx$ 135 % versus $\approx$ 154 % extension compared to the length of B-form DNA) [97], it has more recently been shown that RAD51 extends both mixed sequence ssDNA and dsDNA to very similar extents ($\approx$ 145 %) [109]. This argues that faster polymerisation on ssDNA compared to dsDNA is independent of the extent of RAD51-induced DNA extension. However, it is unknown whether the extension of RAD51-bound poly-dT or poly-dA ssDNA differs from the previously reported mixed sequence ssDNA and

dsDNA extensions [109].

How may the BTC mechanism act *in vivo*? During the initial steps of homologous recombination, resected ssDNA is bound by the ssDNA-binding replication protein A (RPA) prior to RAD51 polymerisation. RPA is a heterotrimer complex with six oligonucleotide/oligosaccaride-binding (OB) folds, four of which can associate tightly with ssDNA in a stepwise manner [226, 227]. When fully associated, one RPA has a 30 nucleotide footprint, while undertaking dynamic exchange with free RPA in solution, an event known as the facilitated exchange [10]. RPA binding to ssDNA is believed to resolve ssDNA secondary structures and facilitate efficient RAD51 polymerisation. Furthermore, DSS1 in complex with BRCA2 has been shown to enhance RPA removal from ssDNA by attenuating its affinity for ssDNA, therefore promoting the RPA-RAD51 exchange and providing free ssDNA for RAD51 polymerisation [7]. As RAD51 binds to 3 nucleotides of ssDNA [113], the unbinding of a single RPA heterotrimer could provide enough space on ssDNA to accommodate up to ten RAD51 protomers. Hence, it is conceivable that RPA dissociation provides highly flexible ssDNA which can promote RAD51 nucleation and polymer elongation.

While higher DNA flexibility may enhance the speed of RAD51 polymerisation, there is also an energetic cost to binding flexible DNA because RAD51 polymerisation is expected to stiffen DNA and reduce the number of spatial configurations flexible DNA can explore. This additional energy cost can be described as an entropic penalty. Therefore, in the next chapter I ask the following question: how does RAD51 overcome the entropic penalty for binding flexible DNA? To address this, I measure the RAD51 binding kinetics to ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p using a RAD51 mutant with a reduced protomer-protomer affinity and fit the dimer models and n-mer models to these data.

Chapter 5

RAD51 Overcomes the Entropic Penalty to Binding Flexible DNA

5.1 Introduction

Protein-DNA interactions are regulated by changes in enthalpy and entropy. As mentioned in section 1.4, the change in Gibbs free energy (ΔG) relates changes in enthalpy (ΔS) and entropy (ΔH) to predict whether a protein-DNA interaction is energetically favourable. Although evidence describing the interplay of entropy and enthalpy in RAD51-ssDNA and RAD51-dsDNA interactions is lacking, the interactions of site-specific proteins with dsDNA are known to be regulated by changes in entropy and enthalpy. DNA-binding proteins such as the bacteriophage λ cI repressor [176], *E. coli* catabolite gene activator protein (CAP) [172], *S. cerevisiae* TATA binding protein (TBP) [228, 173] and general control protein 4 (GCN4) [177] in complex with dsDNA were shown to alter the flexibility of dsDNA to obtain favourable binding energies ($\Delta H < T\Delta S$).

It is unclear whether site-non-specific DNA-binding proteins such as RAD51 may also adopt similar strategies to stably bind DNA. RAD51 polymerisation on DNA incurs an entropic penalty because polymerisation can reduce the number of spatial configurations a DNA molecule can explore. ssDNA is more flexible compared to dsDNA, and is expected to contribute a larger entropic penalty. This is further

supported by the observation that RAD51 polymerisation stiffens ssDNA and dsDNA to a similar extent ($L_p \approx 200$ nm for both RAD51-ssDNA and RAD51-dsDNA complexes [11]). Therefore, how RAD51 polymerisation balances enthalpic and entropic contributions to achieve stable binding on flexible DNA is still unknown.

In this chapter, I investigate how RAD51 can stably bind to flexible DNA. I propose that the enthalpic contribution of RAD51 polymerisation is sufficient to offset the large entropic penalty associated with polymerisation on flexible DNA, therefore enabling the formation of stable RAD51 polymers. To test this hypothesis, I measured RAD51 binding kinetics to the same DNA molecules as in chapter 4 via SPR using a RAD51 point mutant (F86E) associated with a reduced enthalpic contribution [129, 92, 3], and fit the dimer and n-mer models to these data. I conclude that the enthalpic contribution of WT RAD51 polymerisation, driven by a high affinity RAD51 protomer-protomer interaction, enables WT RAD51 polymers to have similar, long lifetimes on both flexible and stiff DNA.

5.2 The Entropic Compensation Mechanism

So far, I have observed that RAD51 associates faster to flexible ssDNA compared to dsDNA, and the proposed BTC mechanism decribes how higher DNA flexibility can induce faster RAD51 polymerisation. However, the formation of a stiff RAD51 polymer on highly flexible DNA is expected to be associated with a large entropic penalty, because RAD51 polymerisation can dramatically reduce the number of spatial configurations that a flexible DNA polymer can explore. RAD51 polymerisation on flexible ssDNA is expected to incur a higher entropic penalty compared to its polymerisation on stiff ssDNA or dsDNA, and it would be expected that the enthalpic contribution of RAD51 polymerisation on DNA is sufficiently large to offset (or compensate) any entropic penalty to associate with ssDNA and dsDNA. However, if this enthalpic contribution were to be reduced, it is predicted that at equilibrium

there would be less RAD51 bound to flexible DNA compared to stiff DNA. I summarise this as the Entropic Compensation Mechanism (EPC), which proposes that the favourable enthalpic energy of RAD51 polymerisation on DNA is sufficient to compensate for any DNA-dependent entropic penalty. Specifically I propose that, upon reduction of the enthalpic contribution, RAD51 will exhibit faster unbinding from flexible DNA compared to stiff DNA, as this is consistent with reduced RAD51 binding at equilibrium (Fig. 5.1).

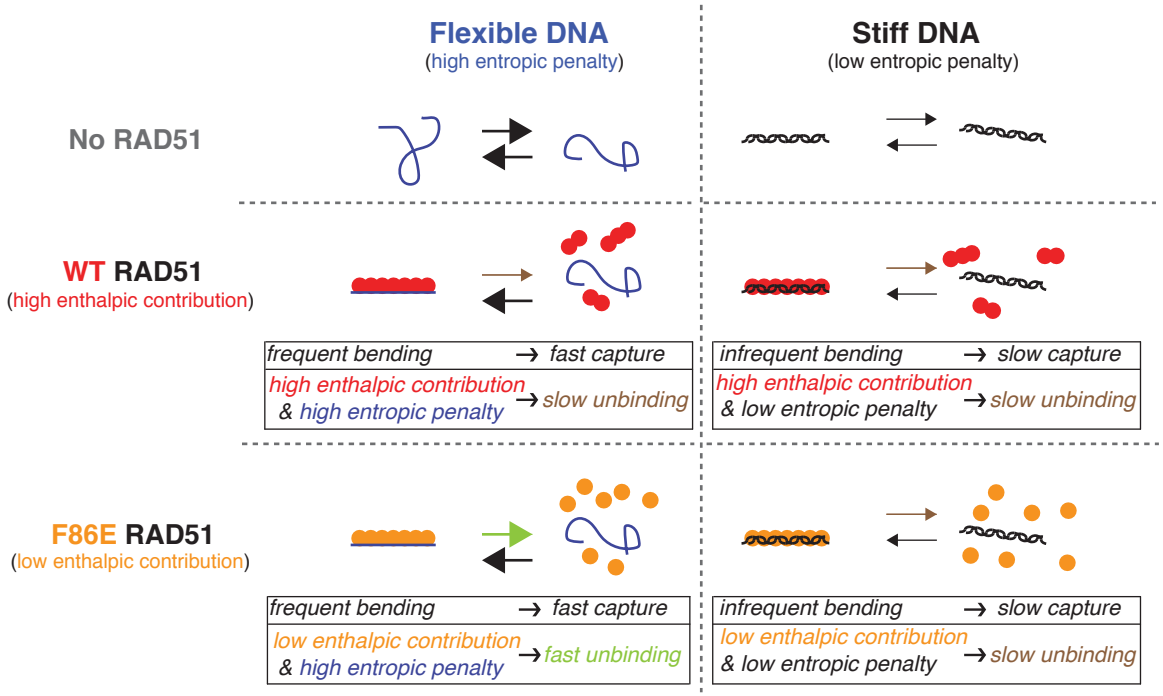
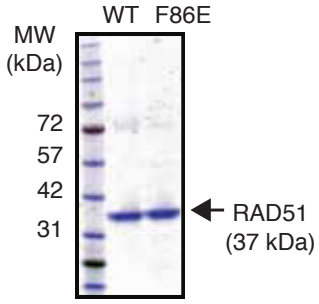


Figure 5.1: The Entropic Penalty Compensation (EPC) mechanism. The enthalpic contribution of WT RAD51 polymerisation enables RAD51 to overcome the entropic penalty to binding flexible DNA (here depicted as ssDNA). F86E RAD51 has a lower protomer-protomer affinity and consequently a lower enthalpic contribution to polymerisation, and cannot overcome the entropic penalty to binding flexible DNA. Both WT and F86E RAD51 can overcome the modest entropic penalty to binding stiff DNA (here depicted as dsDNA). Thick arrows indicate fast capture and fast unbinding, while thin arrows indicate slow capture and slow unbinding.

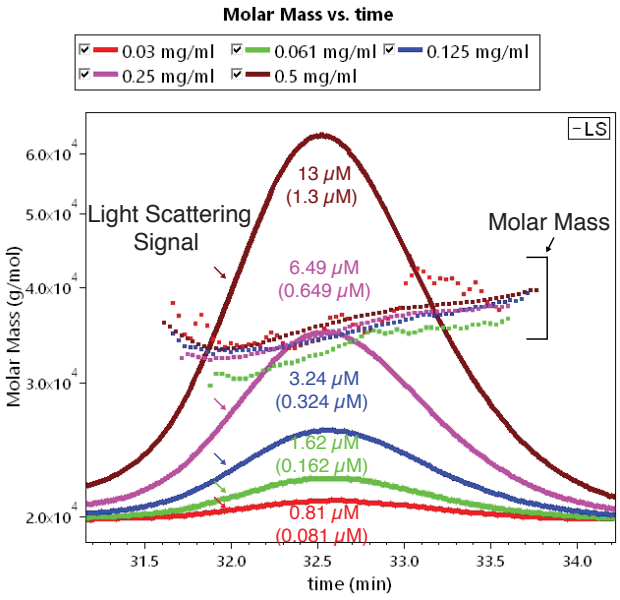
5.3 The RAD51 Protomer-Protomer Interaction Compensates for the Entropic Penalty of Flexible DNA

To directly test this hypothesis, I assessed the kinetics of a RAD51 mutant, which confers a reduced enthalpic contribution compared to WT RAD51. I predicted that, with such a mutation, the stability of the mutant RAD51 polymers on DNA would become more dependent on the flexibility of the naked DNA (Fig. 5.1). I took advantage of a phenylalanine to glutamate substitution at RAD51 residue 86 (F86E), which reduces the RAD51 protomer-protomer interaction affinity [129, 92, 3] (section 1.2.4). Indeed, using SEC-MALS, I confirmed that F86E RAD51 is primarily monomeric in solution (Fig. 5.2A-C).

A



B



C

[F86E] (mg/ml)	MW (kDa)	RSE
0.03	36.2	4.6%
0.061	33.3	1.1%
0.125	34.9	0.8%
0.25	34.9	0.2%
0.5	35.9	0.4%
Average	35	

Figure 5.2: **A**. Purified WT and F86E RAD51 on an SDS-PAGE gel visualised by coomassie blue staining. **B**. Size-exclusion multi-angle light scattering (SEC-MALS) curves for F86E RAD51 at five injection concentrations as indicated. The Superdex200 10/300 GL size exclusion column leads to a $\approx$ 10-fold dilution of the injected sample. Hence, approximate elution concentrations are given in brackets below the injection concentrations. Dotted lines represent the molar mass derived values for each elution peak, un-dotted lines represent light scattering. F86E RAD51 concentrations are shown in both mg/ml and μM. **C**. Calculated molecular weights of the five F86E RAD51 elution peaks. Data was analysed using the ASTRA 6.1 chromatography software for SEC-MALS. MW: molecular weight. RSE: relative standard error.

I then used SPR to measure the binding kinetics of F86E RAD51 to ssDNA dT-50, dN-50, dA-50, and dsDNA dN-50p. First, it was immediately evident that F86E RAD51 shows significantly reduced affinity for ssDNA and dsDNA compared to WT RAD51, with no detectable binding to DNA at the concentration of 150 nM or 3 μM [F86E RAD51] (Fig. 5.3A-B). Nonetheless, at 30 μM [F86E RAD51], I observed association to both ssDNA and dsDNA (Fig. 5.3A-B). Importantly, I observed F86E RAD51 associates faster to flexible DNA (Fig. 5.6A-B), similarly to WT RAD51 (Fig. 4.3A-B). Strikingly, I also observed faster dissociation of F86E RAD51 on more flexible DNA (Fig. 5.6A-B), in contrast to WT RAD51 (Fig. 4.3A-B).

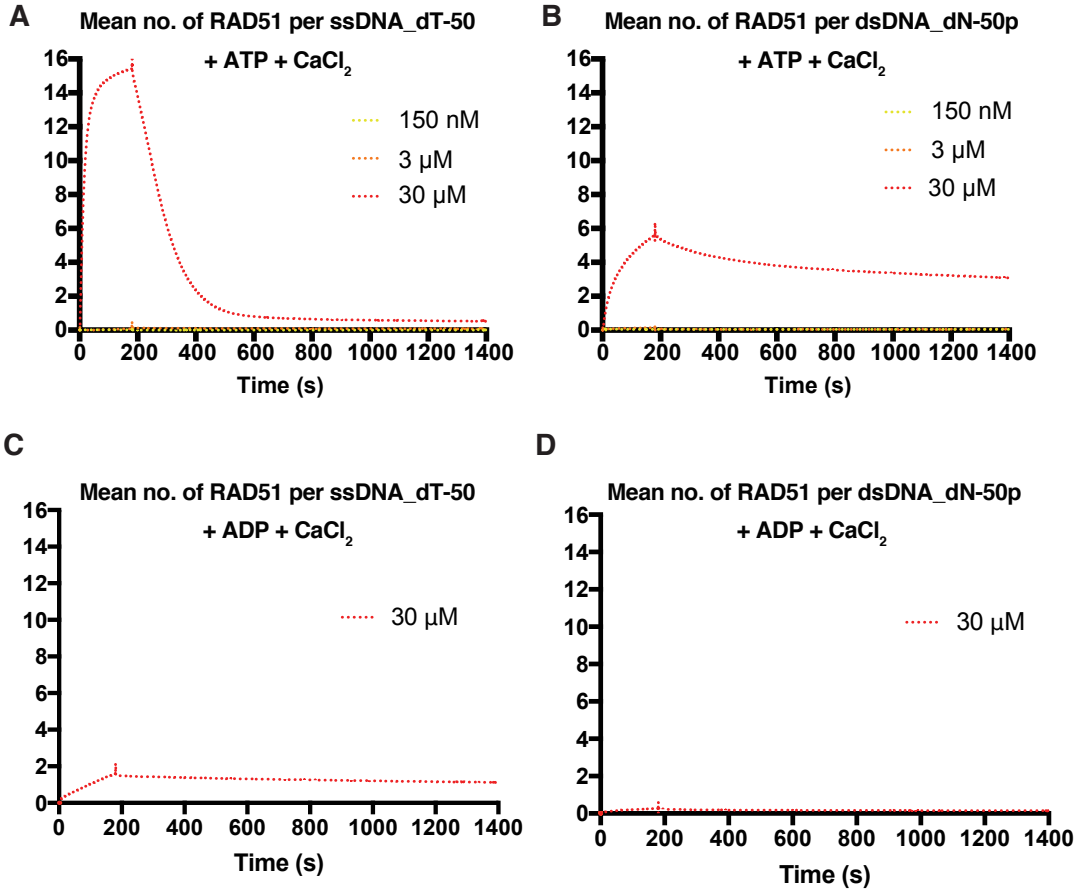


Figure 5.3: F86E RAD51 displays reduced affinity to both ssDNA and dsDNA. **A**. F86E RAD51 SPR binding curves for the ssDNA dT-50 at the indicated concentrations in the presence of 2.5 mM ATP and 10 mM $CaCl_2$. **B**. F86E RAD51 SPR binding curves for the dsDNA dN-50p at the indicated concentrations in the presence of 2.5 mM ATP and 10 mM $CaCl_2$. **C**. F86E RAD51 SPR binding curve for the ssDNA dT-50 at 30 μM [F86E RAD51] in the presence of 2.5 mM ADP and 10 mM $CaCl_2$. The apparently slower dissociation in the presence of ADP compared to ATP is due to a small degree of non-specific binding to the negative control (DNA-free) flow cell. **D**. F86E RAD51 SPR binding curve for the dsDNA dN-50p at 30 μM [F86E RAD51] in the presence of 2.5 mM ADP and 10 mM $CaCl_2$. **C-D**. The SPR buffer used is identical to the buffer used in Fig. 5.3A-B (section 2.6), except that ATP was substituted for ADP. **A-D**. F86E RAD51 was injected for 180 seconds.

I then fit simplified versions of the dimer and n-mer models to these data to compare and contrast the kinetics of F86E RAD51 and WT RAD51 polymerisation. Importantly, I fit the dimer model to these data under the assumption that F86E RAD51 polymerisation takes place only via monomer addition. For the dimer model, I use lsqcurvefit (MATLAB) to fit a single forward rate ($\mathbf{k_a} = \mathbf{k_{ad.f}} = \mathbf{k_{el.f}}$) and a single reverse rate ($\mathbf{k_d} = \mathbf{k_{ad.r}} = \mathbf{k_{el.r}} = \mathbf{k_s}$) to the F86E RAD51 SPR data (Fig. 5.4A). Similarly, for the n-mer model I use ABC-SMC to fit a single RAD51 protomer-protomer dissociation constant ($\mathbf{K_D}$) to all SPR datasets, and a single forward rate constant ($\mathbf{k_p}$, as in Fig. 3.8A-B) and a single reverse rate constant ($\mathbf{k_{pr}} = \mathbf{k_u} = \mathbf{k_q} = \mathbf{k_s}$) to each dataset (i.e. the ssDNA and dsDNA kinetic sub-models A, Table 2.4) (Fig. 5.4B). For both the dimer and n-mer models, I decided to reduce the number of dissociation rate constants to fit because the F86E RAD51 binding kinetics to the short ssDNA oligos (dN-5, dN-8, dN-11, dN-14, dN-17) and dsDNA oligos (dN-5p, dN-8p, dN-11p) (Figs. 3.2A, 3.2C) were not systematically measured due to low F86E purification yield (section 2.2).

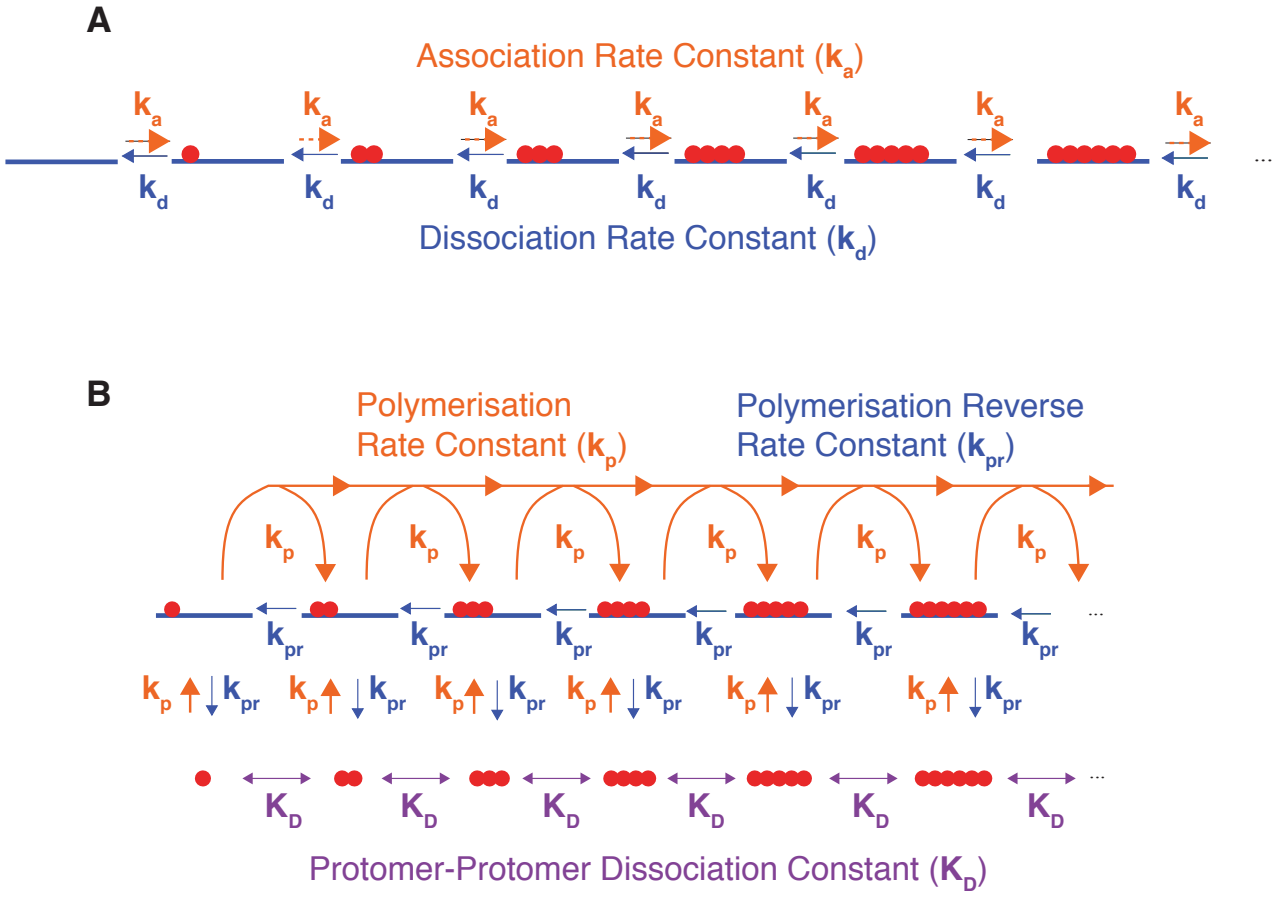


Figure 5.4: Simplified dimer model (**A**) and n-mer model (**B**) used to fit the F86E RAD51 SPR data generated using the dT-50, dN-50, dA-50 and dN-50p DNA molecules. **A**. The simplified dimer model is composed of two parameters: an associate rate constant $\mathbf{k_a}$ ($\mathbf{k_a} = \mathbf{k_{ad.f}} = \mathbf{k_{el.f}}$) and a dissociation rate constant $\mathbf{k_d}$ ($\mathbf{k_d} = \mathbf{k_{ad.r}} = \mathbf{k_{el.r}} = \mathbf{k_s}$). In this model, polymerisation only occurs via monomer addition. This model was used to individually fit the dT-50, dN-50, dA-50 and dN-50p datasets. **B**. The simplified n-mer model consists of the n-mer equilibrium sub-model and kinetic sub-model A (Table 2.4, the ssDNA and dsDNA kinetic sub-models A are the same) and is composed of three parameters: a polymerisation rate constant $\mathbf{k_p}$ (as in Fig. 3.8A-B), a polymerisation reverse rate constant $\mathbf{k_{pr}}$ ($\mathbf{k_{pr}} = \mathbf{k_u} = \mathbf{k_q} = \mathbf{k_s}$) and the RAD51 protomer-protomer dissociation constant ($\mathbf{K_D}$, as in Fig. 3.8A-B). This model was used to simultaneously fit the dT-50, dN-50, dA-50 and dN-50p datasets via ABC-SMC. Purple arrows depict a simplified cartoon representation of RAD51 polymerisation in solution. The model allows for any RAD51 n-mer ($1 \leq n < 16$) to associate with any other RAD51 m-mer ($1 \leq m < 16$) to form an $(m+n)$-mer ($1 < m+n \leq 16$), and any RAD51 k-mer to dissociate into any combination of n-mers and m-mers ($k = n + m$.

The ABC-SMC n-mer model fits suggested F86E RAD51 displays a reduced protomer-protomer dissociation constant compared to WT RAD51 ($\mathbf{K_D}$ = 10.2 ± 7 μM versus 1.14 ± 0.5 nM) (Fig. 5.8B, Table 3.2). This supports the idea that F86E RAD51 displays a weaker protomer-protomer affinity compared to WT RAD51 and is consistent with previous experimental evidence suggesting reduced F86E RAD51 polymerisation in solution compared to WT RAD51 [3]. Importantly, the equilibrium sub-model predicts that F86E RAD51 is monomeric at 150 nM [F86E RAD51], and primarily monomeric at 3 μM [F86E RAD51] (Fig. 5.5B). This is in agreement with my SEC-MALS analysis that also suggested F86E RAD51 is monomeric within the 81 nM - 1.3 μM range (Fig. 5.2B-C). However, the model also predicts F86E RAD51 forms abundant polymers in solution at 30 μM [F86E RAD51] (Fig. 5.5A-B), although to a lesser extent compared to WT RAD51 (Fig. 3.10A-B). This suggests F86E RAD51 association to DNA only occurs when the concentration of F86E RAD51 is sufficiently high such that it forms polymers in solution that can directly adsorb and elongate on DNA, and provides a possible explanation for why F86E RAD51 association to DNA was only detected at 30 μM (Figs. 5.5A-B, 5.3A-B).

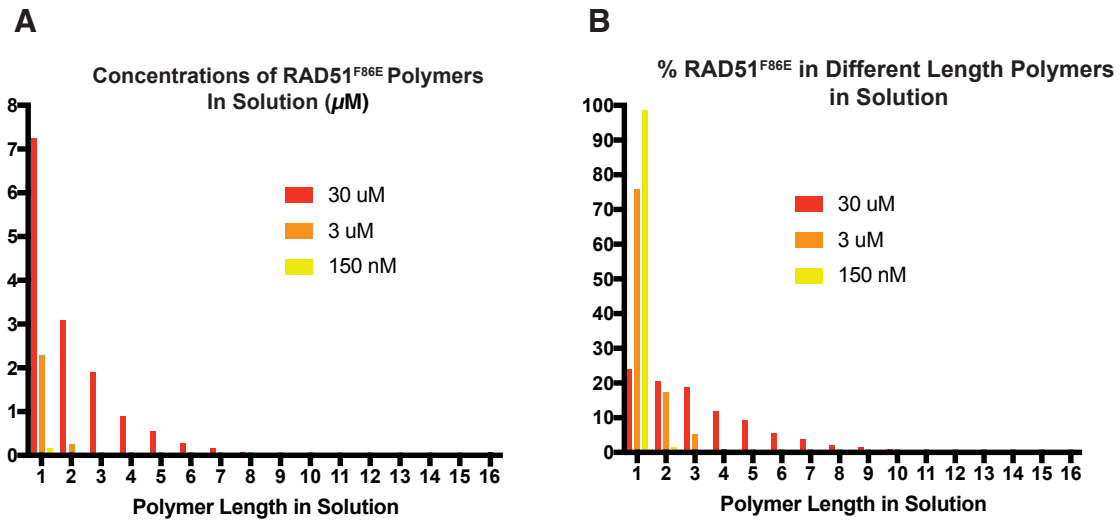


Figure 5.5: F86E RAD51 polymerisation in solution. **A**. Predicted concentrations of F86E RAD51 polymers in solution at equilibrium as a function of F86E RAD51 monomer concentration (150 nM, 3 µM, 30 µM), prior to RAD51 injection onto DNA-coated SPR CM5 chips. **B**. Predicted % RAD51^{F86E} within each polymer state in solution at equilibrium as a function of F86E RAD51 monomer concentration (150 nM, 3 µM, 30 µM), prior to RAD51 injection onto DNA-coated SPR CM5 chips. **A-B**. Polymer concentrations were predicted using the F86E RAD51 equilibrium sub-model. The $\mathbf{K_D}$ was fixed to the mean value (Fig. 5.8B) (i.e. $\mathbf{K_D}$ = 10.2 μM), and the concentration of F86E RAD51 monomer concentration was varied accordingly. In panel **B**, the % RAD51^{F86E} values were calculated by multiplying the polymer concentrations by their respective polymer length, and dividing each value by the total RAD51 monomer concentration (i.e. % RAD51$^{F86E}{}_n$ = nR_n/R_1). R_1 (i.e. the F86E RAD51 SPR injection concentration, [F86E RAD51]) was calculated via coomassie staining image quantification using a BSA standard curve. Due to the relatively high standard deviation of the estimated $\mathbf{K_D}$ ($\mathbf{K_D}$ = 10.2 ± 7 μM) the predicted values of the polymer concentrations are not to be considered as exact predictions.

Furthermore, the dimer model fit suggests F86E RAD51 has a higher association rate constant to more flexible DNA (Fig. 5.6A-C), similarly to WT RAD51 (Fig. 4.3A-C). The n-mer model fit also suggests a higher polymerisation rate constant on more flexible DNA, as for WT RAD51 (Figs. 5.8A-C, Figs. 4.5A-C). The values of the n-mer model $\mathbf{K_D}$, as well as the dT-50, dN-50, dA-50 and dN-50p $\mathbf{k_p}$ values are well determined (Fig. 5.7A-B).

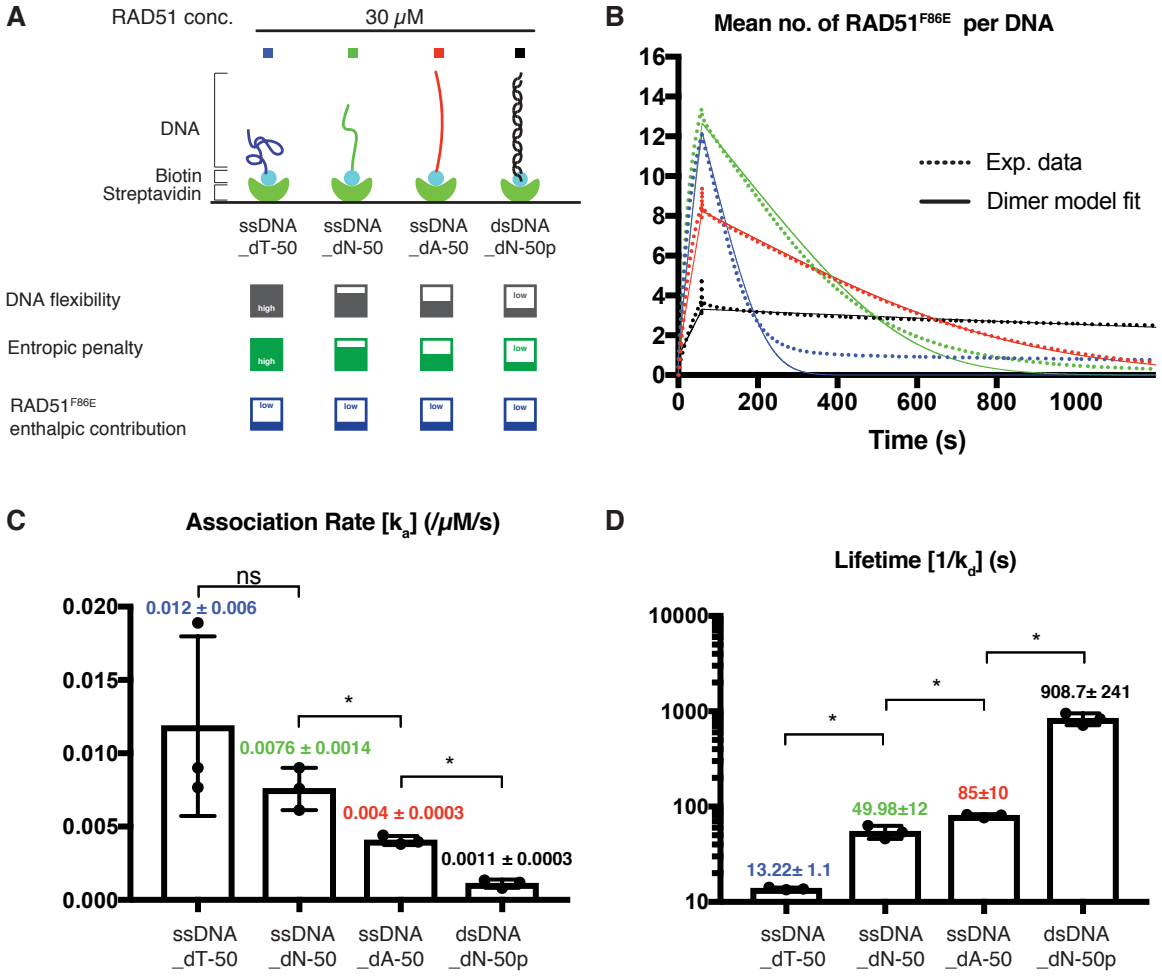


Figure 5.6: F86E RAD51 binding kinetics to DNA with varying flexibility with associated dimer model fits. **A**. The enthalpic contribution of F86E RAD51 polymerisation is insufficient to offset the entropic penalty to binding flexible DNA. **B**. F86E RAD51 SPR curves (dotted lines) and corresponding model fits (solid lines) using the simplified dimer model (Fig. 5.4A). A single forward rate ($\mathbf{k_a} = \mathbf{k_{ad.f}} = \mathbf{k_{el.f}}$) and a single reverse rate ($\mathbf{k_d} = \mathbf{k_{ad.r}} = \mathbf{k_{el.r}} = \mathbf{k_s}$) were fitted to the experimental data using lsqcurvefit (MATLAB). **C-D**. Bar plots of fitted $\mathbf{k_a}$ values (**C**) and $\mathbf{k_d}$ values (reported as $1/\mathbf{k_d}$) (**D**) for each SPR curve. Mean ± 1 SD of 3 dimer model fits (n = 3). Asterisks indicate unpaired, one-tailed Mann-Whitney tests. * $p \leq 0.05$; ns = non-significant.

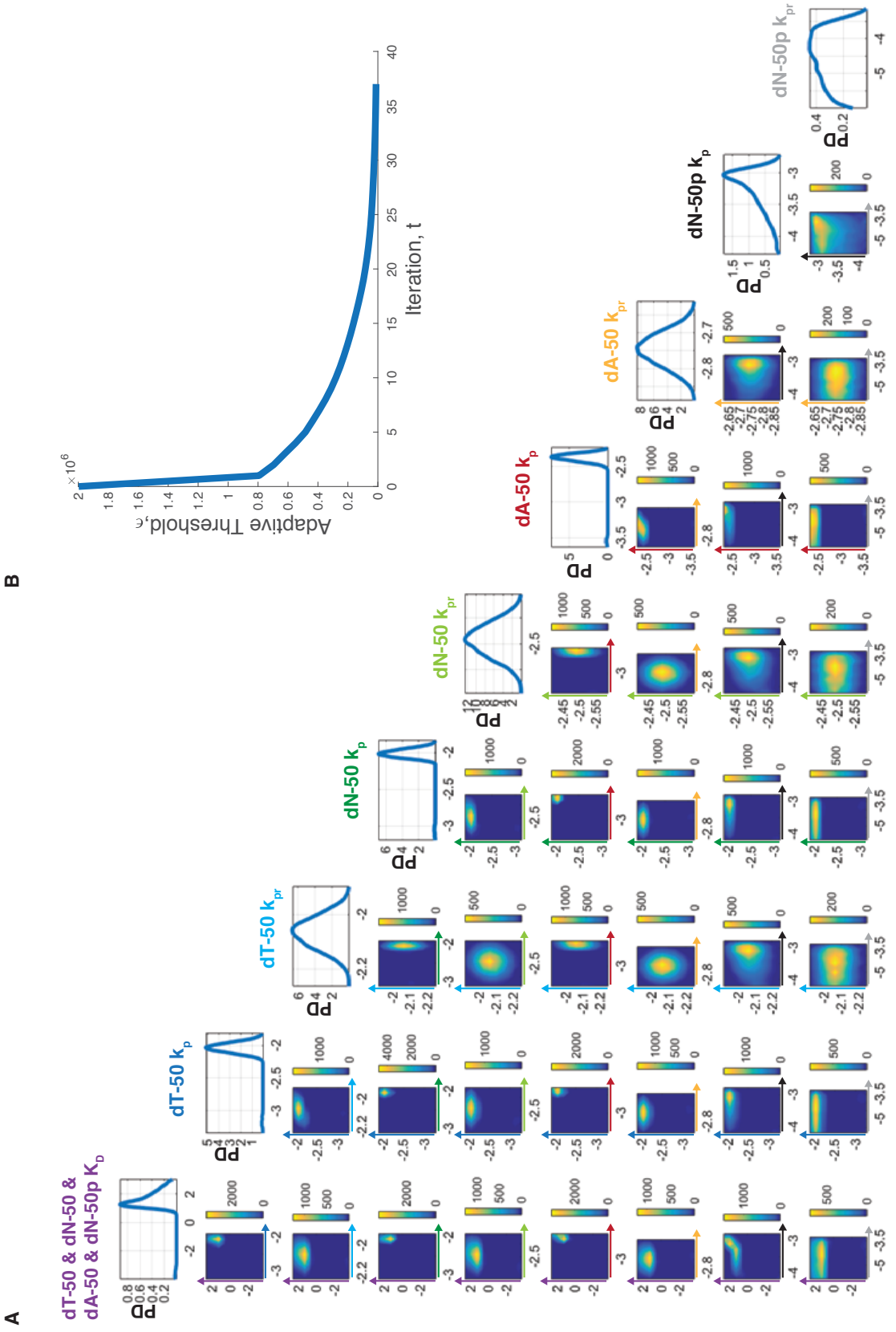

A
B
dT-50 & dN-50 &
dA-50 & dN-50p K_D
dT-50 k_p
dT-50 k_pr
dN-50 k_p
dN-50 k_pr
dA-50 k_p
dA-50 k_pr
dN-50p k_p
dN-50p k_pr
PD
Adaptive Threshold, ε
Iteration, t

Figure 5.7

Figure 5.7: ABC-SMC probability densities for the simplified n-mer model fit parameters for one repeat of the dT-50, dN-50, dA-50 and dN-50p datasets using F86E RAD51. **A**. ABC-SMC of the simultaneous F86E RAD51 n-mer model (Fig. 5.4B) fits for the ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p SPR data (Fig. 5.8B-D). Heat maps indicate particle frequencies and describe any pair-wise correlations between model parameters ($\mathbf{K_D}$, dT-50 $\mathbf{k_p}$, dT-50 $\mathbf{k_{pr}}$, dN-50 $\mathbf{k_p}$, dN-50 $\mathbf{k_{pr}}$, dA-50 $\mathbf{k_p}$, dA-50 $\mathbf{k_{pr}}$, dN-50p $\mathbf{k_p}$, dN-50p $\mathbf{k_{pr}}$. All heat map axes are in log-log scale and are labelled according to the corresponding parameter colour: $\mathbf{K_D}$ (purple), dT-50 $\mathbf{k_p}$ (blue), dT-50 $\mathbf{k_{pr}}$ (cyan), dN-50 $\mathbf{k_p}$ (green), dN-50 $\mathbf{k_{pr}}$ (light green), dA-50 $\mathbf{k_p}$ (red), dA-50 $\mathbf{k_{pr}}$ (salmon), dN-50p $\mathbf{k_p}$ (black), dN-50p $\mathbf{k_{pr}}$ (grey). A high particle frequency corresponds to a good model fit to the data for the specified parameter pair. The probability densities for individual parameters are presented along the top left diagonal and were estimated using a Gaussian kernel. Particles were initialised via log-uniform priors with the following lower and upper bounds: $10^{-4} < \mathbf{K_D} < 10^3$, 10^{-6} < dT-50 $\mathbf{k_p} < 10^4$, 10^{-6} < dT-50 $\mathbf{k_{pr}} < 10^3$, 10^{-6} < dN-50 $\mathbf{k_p} < 10^4$, 10^{-6} < dN-50 $\mathbf{k_{pr}} < 10^3$, 10^{-6} < dA-50 $\mathbf{k_p} < 10^4$, 10^{-6} < dA-50 $\mathbf{k_{pr}} < 10^3$, 10^{-6} < dN-50p $\mathbf{k_p} < 10^4$, 10^{-6} < dN-50p $\mathbf{k_{pr}} < 10^3$. All parameters except dN-50p $\mathbf{k_{pr}}$ are well determined. Only a lower bound could be determined for dN-50p $\mathbf{k_{pr}}$ ($\mathbf{k_{pr}} < 10^{-3.25}$). PD: probability density. **B**. Adaptive threshold for the ABC-SMC simulation. The plateau at 37 iterations means the model has reached a best fit to the data. The simulation was carried out using MATLAB (37 iterations, 10000 particles).

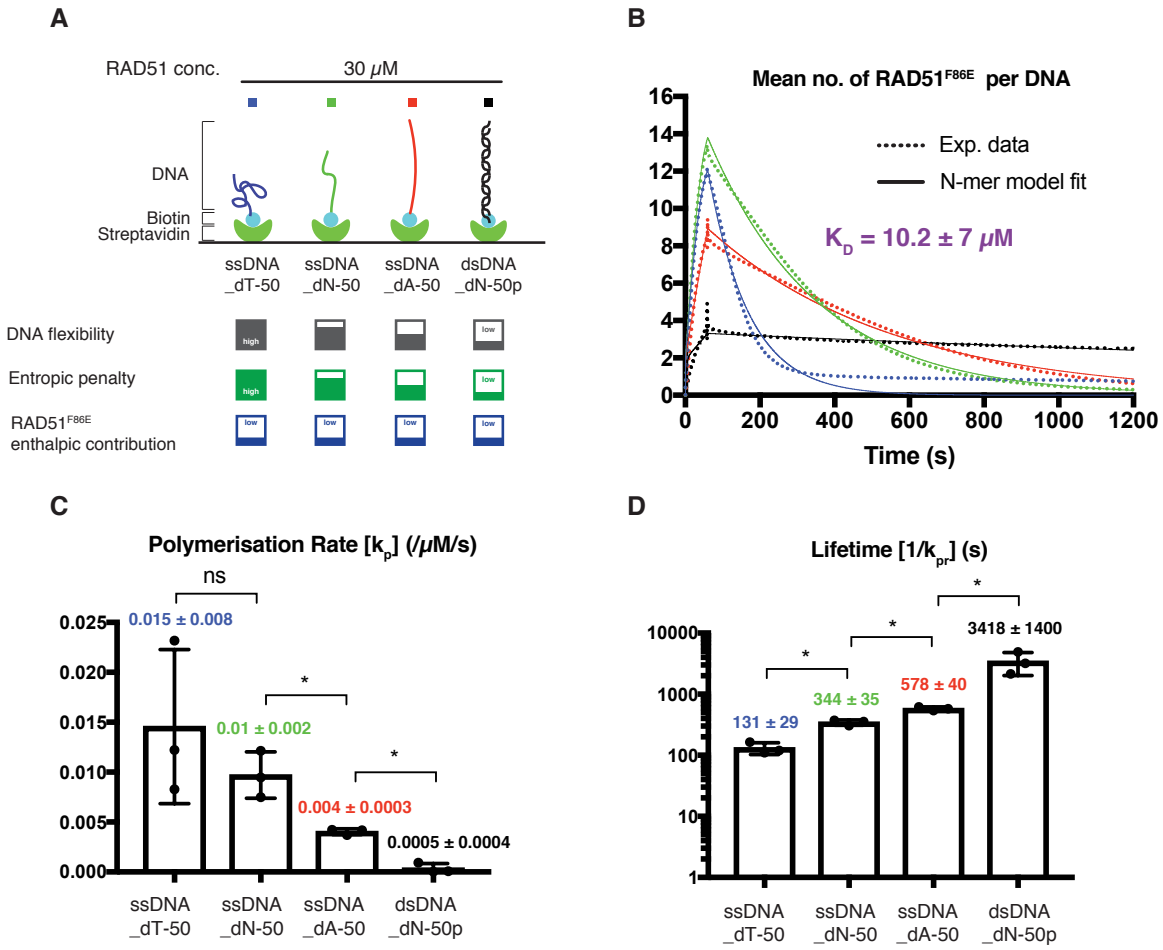


Figure 5.8: F86E RAD51 binding kinetics to DNA with varying flexibility with associated n-mer model fits. **A**. The enthalpic contribution of F86E RAD51 polymerisation is insufficient to offset the entropic penalty to binding flexible DNA. **B**. F86E RAD51 SPR curves (dotted lines) and corresponding model fits (solid lines) using the ssDNA n-mer model (ssDNA dT-50, dN-50, dA-50) and the dsDNA n-mer model (dsDNA dN-50p). A single RAD51 protomer-protomer dissociation constant ($\mathbf{K_D}$) was fit to all SPR datasets (dT-50, dN-50, dA-50, dN-50p), and a single forward rate ($\mathbf{k_p} = \mathbf{k_{ad.f}} = \mathbf{k_{el.f}}$) and a single reverse rate ($\mathbf{k_{pr}} = \mathbf{k_{ad.r}} = \mathbf{k_{el.r}} = \mathbf{k_s}$) were fit to each SPR dataset using ABC-SMC. **C-D**. Bar plots of fitted $\mathbf{k_p}$ values (**C**) and $\mathbf{k_{pr}}$ values (reported as $1/\mathbf{k_{pr}}$) (**D**) for each SPR curve. Mean ± 1 SD of 3 n-mer model fits (n = 3). The $\mathbf{K_D}$ (**B**), $\mathbf{k_p}$ (**C**) and $\mathbf{k_s}$ values (**D**) are the modes of the parameter distributions derived from the ABC-SMC simulations (Fig. 5.7A). Unpaired, one-tailed Mann-Whitney tests. * $p \leq 0.05$; ns = non-significant.

To compared the dissociation speeds of F86E RAD51 on dT-50, dN-50, dA-50 and dN-50p, I then defined the lifetime of F86E RAD51 with the dimer model as $\frac{1}{\mathbf{k_d}}$ (i.e. the average lifetime of single F86E RAD51 protomers on DNA), and with the n-mer model as $\frac{1}{\mathbf{k_{pr}}}$ (i.e. the average lifetime of a F86E RAD51 n-mer on DNA, $1 \leq n \leq 16$). Importantly, the n-mer model $\mathbf{k_{pr}}$ values for dT-50, dN-50 and dA-50 obtained via ABC-SMC are well determined, although only a lower bound could be determined for dN-50p $\mathbf{k_{pr}}$ ($\mathbf{k_{pr}} < 10^{-3.25}$) (Fig. 5.7A). Importantly, both the dimer model and the n-mer model fits suggest the dissociation speed of F86E RAD51 inversely correlates with the flexibility of DNA (Figs. 5.6A-B, D, 5.8A-B, D). This observation is in sharp contrast with WT RAD51, which formed stable, long-lived polymers independently of DNA flexibility (Figs. 4.3A-B, D, 4.5A-B, D).

Dissociation from dT-50 plateaus at $\approx$ 1 F86E RAD51 per DNA molecule within the timescale of the experiment because there was a small portion of F86E RAD51 that remained tightly bound to dT-50 (Figs. 5.6B, 5.8B). For this reason, the dimer and n-mer model fits for the dT-50 accurately describe the first $\approx$ 200 s of the dissociation phase, and this was deemed sufficient given that most F86E RAD51 dissociation from dT-50 occurs within the first 200 s (Figs. 5.6B, 5.8B).

Association and dissociation occur simultaneously when RAD51 is injected over the DNA-coated SPR chip. The slightly higher maximum binding observed with dN-50 ($\approx$ 13 F86E RAD51s per DNA molecule) compared to dT-50 ($\approx$ 12 F86E RAD51s per DNA molecule) can be explained by the fact that F86E RAD51 displays similar forward rate constants on dN-50 and dT-50 ($\mathbf{k_a}$ in the dimer model, $\mathbf{k_p}$ in the n-mer model) (Figs. 5.6B-C, 5.8B-C), but displays a lower lifetime on dT-50 compared to dN-50 ($\frac{1}{\mathbf{k_d}}$ in the dimer model, $\frac{1}{\mathbf{k_{pr}}}$ in the n-mer model) (Figs. 5.6B, D, 5.8B, D).

Taken together, these observations suggest that the enthalpic contribution of WT

RAD51 polymer formation, associated with its high-affinity protomer-protomer interaction, is sufficient to offset the large entropic penalty associated with polymerisation on flexible DNA, therefore supporting the EPC mechanism. As a result, RAD51 is able to polymerise more rapidly on flexible DNA despite incurring a larger entropic penalty.

5.4 Discussion

In this chapter, I have addressed how RAD51 may overcome the entropic penalty to stably bind flexible DNA. I have measured RAD51 binding kinetics to DNA oligos of different flexibilities using the F86E RAD51 mutant with a reduced protomer-protomer affinity. I fit these data using simplified versions of the RAD51 dimer and n-mer models, which respectively assume F86E RAD51 polymerises on DNA only via monomer addition (dimer model) or via the addition of pre-formed polymers (n-mer model), and determined with both models decreasing F86E RAD51 lifetimes on more flexible DNA. I conclude that the enthalpic contribution of WT RAD51 polymerisation is sufficient to offset any entropic penalty to binding flexible DNA, enabling RAD51 polymers to have similar lifetimes on both flexible and stiff DNA.

SPR does not enable direct observation of RAD51 polymerisation on DNA, but rather measures the addition of RAD51 mass to a DNA-coated chip matrix as a proxy for RAD51-DNA interactions. For this reason, a potential caveat of this analysis may be that F86E RAD51 does not form polymers either in solution or on DNA, given that it has been biochemically characterised only as a monomer (Fig. 5.2). Hence, there is a possibility that F86E RAD51 simply adsorbs and dissociates from DNA as a monomer and does not form any protomer-protomer interactions at 30 μM concentration. However, I would suggest this is unlikely. Structural studies have suggested the F86E mutation inhibits the protomer-protomer interaction surface mediated by the short β-strand within the RAD51 linker region (residues 85 - GFTTA - 89) [129, 92] (Fig. 1.4D). However, it is likely F86E RAD51 molecules can still form ATP-dependent protomer-protomer interactions via the Walker A and B motifs (Fig. 1.4E) [60], as well as hydrophobic protomer-protomer interactions via the Y54 and F195 residues (section 1.2.4) [60], to enable polymerisation. Indeed, I show that substituting ATP with ADP in the presence of $CaCl_2$ drammati-

cally reduces F86E RAD51 association to both dT-50 and dN-50p (Fig. 5.3A-D), supporting the idea that ATP-dependent protomer-protomer interactions promote F86E RAD51 polymerisation on DNA. Whether F86E RAD51 forms compressed filaments in the presence of ADP, as suggested for WT RAD51 [112, 121], remains unknown. Therefore, I would argue that the F86E mutation strongly reduces the RAD51 protomer-protomer affinity by inhibiting one of the three interaction surfaces such that F86E RAD51 is primarily monomeric at concentrations below 3 μM (Figs. 5.2A-C, 5.5A-B), but does not prevent polymerisation at higher F86E RAD51 concentrations (Figs. 5.3A-B, 5.5A-B).

On the basis of the findings in chapters 3, 4 and 5, I propose the **DUET** (**D**na molec**U**lar fl**E**xibili**T**y) model to describe how RAD51 polymerises on ssDNA in preference to dsDNA: 1) RAD51 polymerises faster on flexible DNA because flexible DNA explores more conformations compatible with the binding of both RAD51 monomers and polymers per unit time (BTC) (Fig. 4.1), and 2) the enthalpic contribution of RAD51 polymerisation enables RAD51 to overcome the entropic penalty to binding flexible DNA (EPC) (Fig. 5.1).

The diploid human genome consists of $\approx$ 6.4 billion base pairs, and $\approx$ 50 endogenous DSBs are estimated to occur in every cell cycle in each cell [229]. In HR-mediated DSB repair, ssDNA is generated to the length of up to 3500 nucleotides in human cells [15], which then serves as a substrate for RAD51 polymerisation. However, ssDNA of this sort would constitute only up to 0.005 % of total genomic DNA (3500 nts x 2 overhangs x 50 DSBs / 6.4 billion bps), and it is easily imaginable that the polymerisation of RAD51 exclusively on such ssDNA is a rare event. The nuclear concentration of RAD51 is estimated at $\approx$ 100 nM [230], and a plethora of protein co-factors (e.g. BRCA2, PALB2, RAD52) coordinate the recruitment of RAD51 to ssDNA and not on dsDNA, therefore increasing the RAD51 local con-

centration at DSBs [8, 6, 5, 38, 9, 7]. However, it remains unclear how RAD51 polymerises on the insignificant ssDNA generated at DSBs in preference to the bulk of undamaged dsDNA.

Another major unanswered issue about RAD51 polymerisation is how RAD51 can discriminate ssDNA generated by DSB resection from ssDNA generated during normal cellular processes, such as DNA replication and transcription. Spontaneous RAD51 polymerisation on all ssDNA would be problematic, as it could trigger undesired recombination or disruption of DNA replication and transcription, resulting in genomic instability. In this context, the DUET model is particularly appealing, as ssDNA generated during transcription and replication does not have free ends and could therefore be less flexible compared to resected ssDNA with a free end. This in turn may reduce RAD51 polymerisation on transcription- and replication-derived ssDNA, while promoting it on ssDNA at DSB sites.

These results also suggest a more general mechanism for the recruitment of DSB repair factors. Indeed, numerous DNA repair machineries are comprised of multiple subunits with several binding interfaces to DNA and other co-factors. The high flexibility of ssDNA at DSBs may enhance recruitment of such repair complexes, and could be involved in promoting the clustering of multiple DSBs into repair centers [231, 232]. However, poly-dA-rich (stiff) ssDNA may instead inhibit the recruitment of repair complexes. In support of this, it was recently shown that RPA fails to associate to poly-dA ssDNA tracts at replication origins in mammalian cells and exposes the origins to DSBs [233]. Further studies in combination with RAD51 mediators, such as RAD51 paralogs and BRCA2, which have been shown to facilitate RAD51 filament formation [44, 54], will be important to further advance the understanding of the RAD51 polymerisation mechanism. These results provides evidence for a previously unappreciated role of DNA flexibility in the process of DNA repair.

In the next chapter, I will summarise the contents of the thesis, discuss experimental caveats and compare and contrast the RAD51 polymerisation models I have developed to other similar models in the literature. I will also provide ideas on how to further improve the mechanistic understanding of RAD51 polymerisation by expanding the experimental and mathematical modelling approaches used in this study.

Chapter 6

Conclusion

6.1 Summary of Thesis

The main focus of this thesis has been to investigate how human RAD51, the key catalyst of HR, discriminates between ssDNA and dsDNA.

To begin, I measured the RAD51 binding kinetics to ssDNA and dsDNA oligos of different lengths using SPR (chapter 3). I then developed the dimer and n-mer RAD51 polymerisation models, and used ABC-SMC to fit the models to these data and derived three conclusions. First, both models suggested RAD51 polymerisation on ssDNA proceeds via the formation of an unstable nucleus (three RAD51 molecules), followed by a quasi-stable nucleus (four RAD51 molecules) and a stable nucleus (five RAD51 molecules). In contrast, both models suggested RAD51 polymerisation on dsDNA proceeds via the formation of an unstable nucleus (RAD51 monomer) followed by a stable nucleus (RAD51 dimer). I also concluded with both models that RAD51 displays faster adsorption and/or elongation on ssDNA compared to dsDNA, despite RAD51 requiring less molecules to form a stable nucleus on dsDNA versus ssDNA (two versus five).

I then assessed the impact of DNA flexiblity on RAD51 polymerisation kinetics by measuring RAD51 binding kinetics to highly flexible ssDNA (ssDNA dT-50) and stiff

ssDNA (ssDNA dA-50) via SPR (chapter 4). I fit the ssDNA dimer and n-mer models to these data and concluded that RAD51 polymerises faster on flexible DNA compared to stiff DNA (BTC). In particular, DNA flexibility may affect both RAD51 adsorption and polymer elongation kinetics. In addition, I observed that RAD51 had similar lifetimes on both flexible and stiff DNA.

Finally, I investigated how RAD51 forms stable polymers on flexible DNA (chapter 5). I purified the F86E RAD51 point mutant with a reduced RAD51 protomer-protomer affinity, and measured its binding kinetics to flexible and stiff DNA. By fitting simplified versions of the dimer and n-mer models from chapters 3 to these data, I showed that the lifetime of F86E RAD51 on DNA inversely correlated with the flexibility of the naked DNA. From this observation I concluded that the enthalpic contribution of WT RAD51 polymerisation, driven by a high affinity RAD51 protomer-protomer interaction, is sufficient to offset any entropic penalty to binding flexible DNA (EPC).

Based on the conclusions from this study, I proposed the **DUET** model describing how RAD51 preferentially polymerises on ssDNA compared to dsDNA. Although RAD51 requires fewer molecules to form a stable nucleus on dsDNA compared to ssDNA, its polymerisation on flexible ssDNA is specifically enhanced via two mechanisms: 1) RAD51 polymerises faster on flexible ssDNA because flexible ssDNA explores more spatial conformations compatible with the further addition of RAD51 monomers or polymers per unit time (BTC), and 2) the enthalpic contribution of RAD51 polymerisation, driven by a strong RAD51 protomer-protomer interaction, is sufficient to offset any entropic penalty to stably polymerise on flexible ssDNA (EPC) (Fig. 6.1).

In the remainder of the Conclusion chapter, I will describe the significance of the

results, compare and contrast the RAD51 polymerisation models I have developed to other similar models in the literature, and propose further work to refine the DUET model and better understand how RAD51 preferentially polymerises on ssDNA compared to dsDNA *in vivo*.

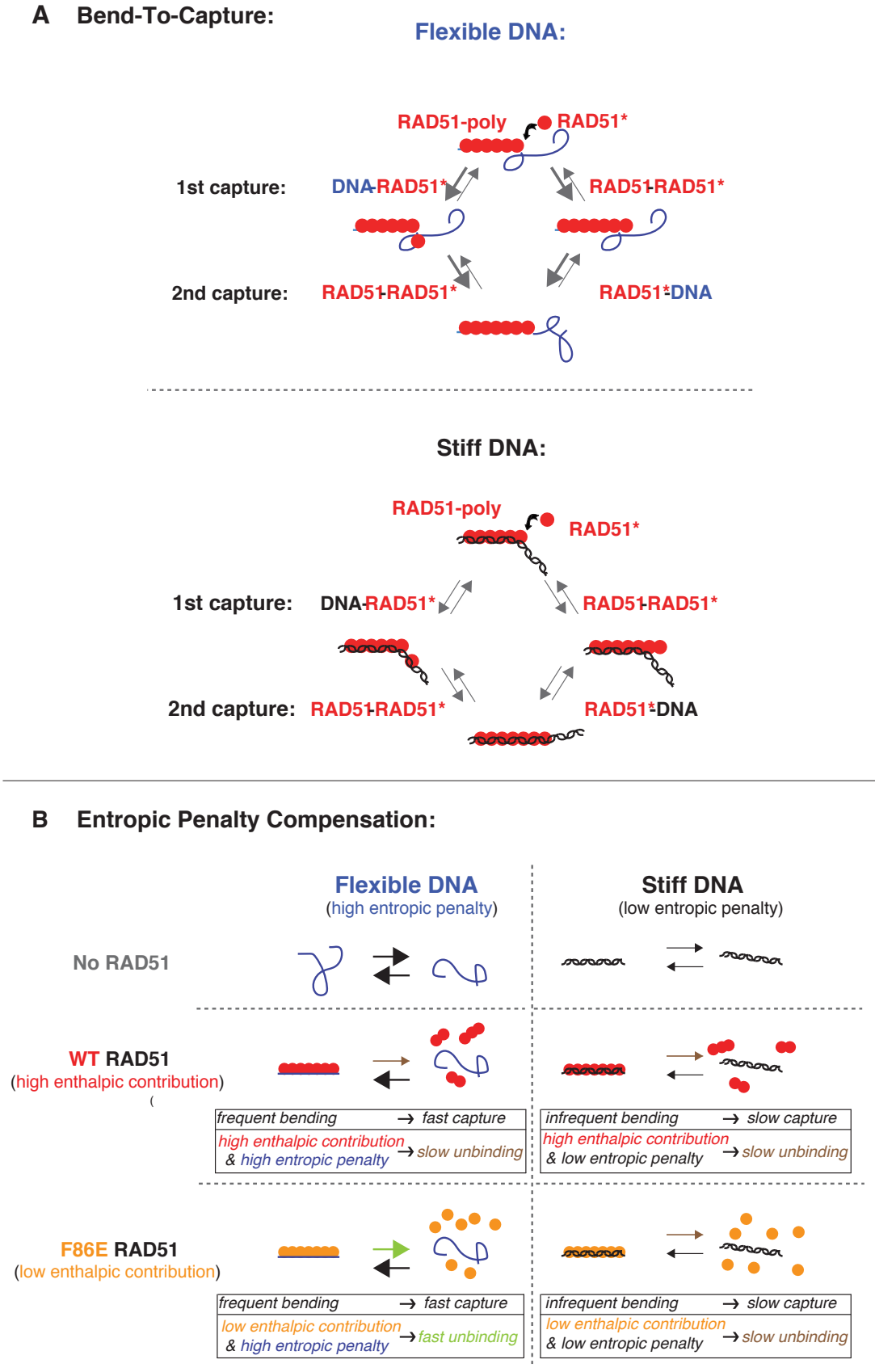


Figure 6.1: The DUET model consists of the Bend-To-Capture (BTC) and the Entropic Penalty Compensation (EPC) mechanisms. **A**. The BTC mechanism, as depicted in Fig. 4.1. **B**. The EPC mechanism, as depicted in Fig. 5.1.

6.2 Significance of Results

The results presented in this thesis support a previously unappreciated role for DNA flexibility as a determinant of RAD51 binding to ssDNA over dsDNA, and suggest DNA flexibility may regulate the recruitment of other DNA repair proteins that multimerise at DNA damage sites (e.g. RecA, RadA). Furthermore, they also complement previous observations describing the correlation between increased dsDNA flexibility due to HMGB-induced bending and enhanced DNA repair [162, 167, 166, 159, 164, 221] (sections 1.4.1 and 4.1). Interestingly, these results are also supported by the conclusions of a computational study which analysed the impact of linear polymer flexibility on polymer-polymer binding affinities [234]. Although this study did not analyse polymerisation kinetics, it suggested that higher polymer flexibilities are associated with higher polymer-polymer binding affinities, and that high enthalpic contributions can compensate for entropic penalties to binding flexible polymers [234]. Hence, it is possible that the proposed role of DNA flexibility in modulating RAD51 binding kinetics may be an example of a more general physico-chemical mechanism regulating the self-assembly of linear polymers on linear polymers. However, it is important to highlight that the dimer and n-mer models developed in chapter 3 represent approximations of RAD51 polymerisation kinetics on DNA used to interpret SPR data. An interesting next step would be to extend the biophysical models developed to fit SPR to describe RAD51 polyermisation on DNA *in vivo*, which is likely to include effects of longer DNA substrates, longer RAD51 filaments, and the effects of protein co-factors such as BRCA2, RPA, PALB2 and RAD51 paralogs.

6.2.1 Alternative Mechanism to BTC

As mentioned in section 4.5, a possible alternative to the BTC mechanism is the STR mechanism, whereby DNA base-stacking, rather than bending, determines

whether RAD51 can effectively bind to and extend DNA: lower base-stacking correlates with faster RAD51 binding. However, this mechanism would suggest that RAD51 should extend ssDNA more compared to dsDNA, given that ssDNA has less base stacking compared to dsDNA. While one study supports this claim [97], another more recent study suggests RAD51 extends mixed sequence ssDNA and dsDNA to similar extents ($\approx$ 145 % the length of B-form DNA) [109]. It would be useful to combine single molecule fluorescence microscopy and optical tweezers (similarly to [138]) to monitor and compare in real-time the extension of single poly-dT ssDNA, mixed sequence ssDNA, poly-dA ssDNA and mixed sequence dsDNA molecules upon the addition of RAD51. This would allow to test the STR mechanism and better understand how base-stacking affects the ability of RAD51 to bind to and extend DNA.

6.2.2 Experimental Caveats

As mentioned in section 5.4, one of the possible experimental caveats of this study was that F86E RAD51 may not form polymers either in solution or on DNA, given that in this study it has only been characterised as a monomer (Fig. 5.2). While this possibility may not be formally excluded, I argue that F86E RAD51 still retains the ability to form polymers, although to a much lesser extent compared to WT RAD51 due to a much lower protomer-protomer interaction affinity ($\mathbf{K_D}$). While the F86E mutation abrogates one of the protomer-protomer interaction surfaces, F86E RAD51 is still expected to have functional Walker A and B motifs to form polymers via ATP-mediated protomer-protomer interactions (Fig. 1.4A, E). I then supported this claim by showing that F86E RAD51 displays reduced binding to both ssDNA dT-50 and dsDNA dN-50p in the presence of ADP instead of ATP (Fig. 5.3C-D), suggesting F86E RAD51 binding to DNA is nucleotide-dependent and that ATP-mediated protomer-protomer interactions enhance polymerisation on DNA.

Nonetheless, the F86E RAD51 SPR data (Figs. 5.6B, 5.8B) could also have been analysed using the Langmuir binding model [195] accounting for the RAD51:DNA binding stoichiometry of 1:16 for the dT-50, dN-50, dA-50 and dN-50p DNA molecules, under the assumption that F86E RAD51 does not form polymers on DNA. The exponential decay fits of these data would have most likely shown decreasing RAD51 lifetime with increasing DNA flexibility, as shown using the simplified dimer and n-mer models in Figs. 5.6D, 5.8D.

6.3 Comparisons with Other Polymerisation Models

The most well known mathematical model describing the binding of analytes (e.g. protein molecules such as RAD51) to linear lattices (e.g. ssDNA, dsDNA) at equilibrium is the Mcghee & von Hippel model [141]. In this section, I will introduce the Mcghee & von Hippel model and describe its advantages and drawbacks compared to the dimer and n-mer models developed in chapter 3. I will then describe the Schwarz & Watanabe model [235], an extension of the Mcghee & von Hippel model that allows simple analysis of binding kinetics of analytes to linear lattices under specific experimental conditions. I will conclude by highlighting the advantages and drawbacks of this model compared to the dimer and n-mer models.

6.3.1 The Mcghee & von Hippel Model

The Mcghee & von Hippel model [141] is a simple mathematical model that describes the cooperative binding of protein molecules to DNA as a function of three parameters: the binding site size, ρ (unitless) (i.e. the number of nts or bps occupied by a single protein molecule), an intrinsic binding constant, K (in units of inverse moles per liter, M^{-1}) (i.e the affinity of a single protein molecule to an isolated binding site), and a protein-protein cooperativity factor, ω (unitless) (i.e. the equilibrium constant for the process of moving a bound ligand from an isolated site to a singly contiguous site, or from a single contiguous to a doubly contiguous site) (Fig.

6.2A) [141]. Importantly, it assumes all protein molecules to be homogenous (i.e. ρ is the same for all protein molecules), and considers DNA as an infinite size lattice such that the length of the DNA, N (in nts or bps) $\gg \rho$, and neglecting any effects associated with binding DNA ends. The model takes the form of an implicit, closed form equation that relates the "binding density", $\nu = \theta/\rho$ (i.e. the fractional amount of DNA nts or bps bound by protein molecules, θ, divided by ρ: $0 \leq \nu < 1/\rho$), and the concentration of unbound protein, L (in units of moles per liter), to the above mentioned parameters:

$$\frac{\nu}{L} = K(1-\rho\nu)\left(\frac{(2\omega+1)(1-\rho\nu)+\nu-R}{2(\omega-1)(1-\rho\nu)}\right)^{\rho-1}\left(\frac{1-(\rho+1)\nu+R}{2(1-\rho\nu)}\right)^{2} \tag{6.1}$$

where $R = \sqrt{[1-(\rho+1)\nu]^2+4\omega\nu(1-\rho\nu)}$. When $\omega = 1$, there is no cooperativity and the model can be further simplified [141]. Negative cooperativity occurs when $0 < \omega < 1$ (i.e. bound protein molecules "repel" unbound protein molecules), while positive cooperativity occurs when $\omega > 1$ (i.e. bound protein molecules "attract" unbound protein molecules). Specifically, the model describes equilibrium protein-DNA binding, and has previously been used to fit data obtained from steady-state titration experiments [142, 236]. Given that experimental data is often measured as a function of total protein concentration L_T, and that θ can be measured directly as a function of L_T, the following variable changes facilitate data fitting [142]:

$$L = L_T - N\nu \tag{6.2}$$

$$\nu = \theta/\rho \tag{6.3}$$

The main finding from the model analysis was that, assuming no protein-protein interactions (i.e $\omega = 1$), protein molecules that bind more than one DNA residue (i.e.

$\rho \geq 2$) display an ”anti-cooperative” behaviour. This is due to the fact that there are overlapping binding sites on DNA when $\rho \geq 2$. Intuitively, if $\rho = 1$, a single protein molecule can only exclude a single binding site, while if $\rho = 2$, a single protein molecule excludes three binding sites, and if $\rho = 3$, a single protein molecule excludes five binding sites (Fig. 6.2B) [141]. Furthermore, overlapping binding sites also lead to the formation of gaps that are too short to accommodate further protein molecules. This overall negative effect can be antagonised by a sufficiently high $\omega > 1$ that promotes efficient polymerisation on DNA by enabling protein molecules to bind to DNA in long clusters with little or no gaps (e.g. $\omega \geq 10\rho$) [141].

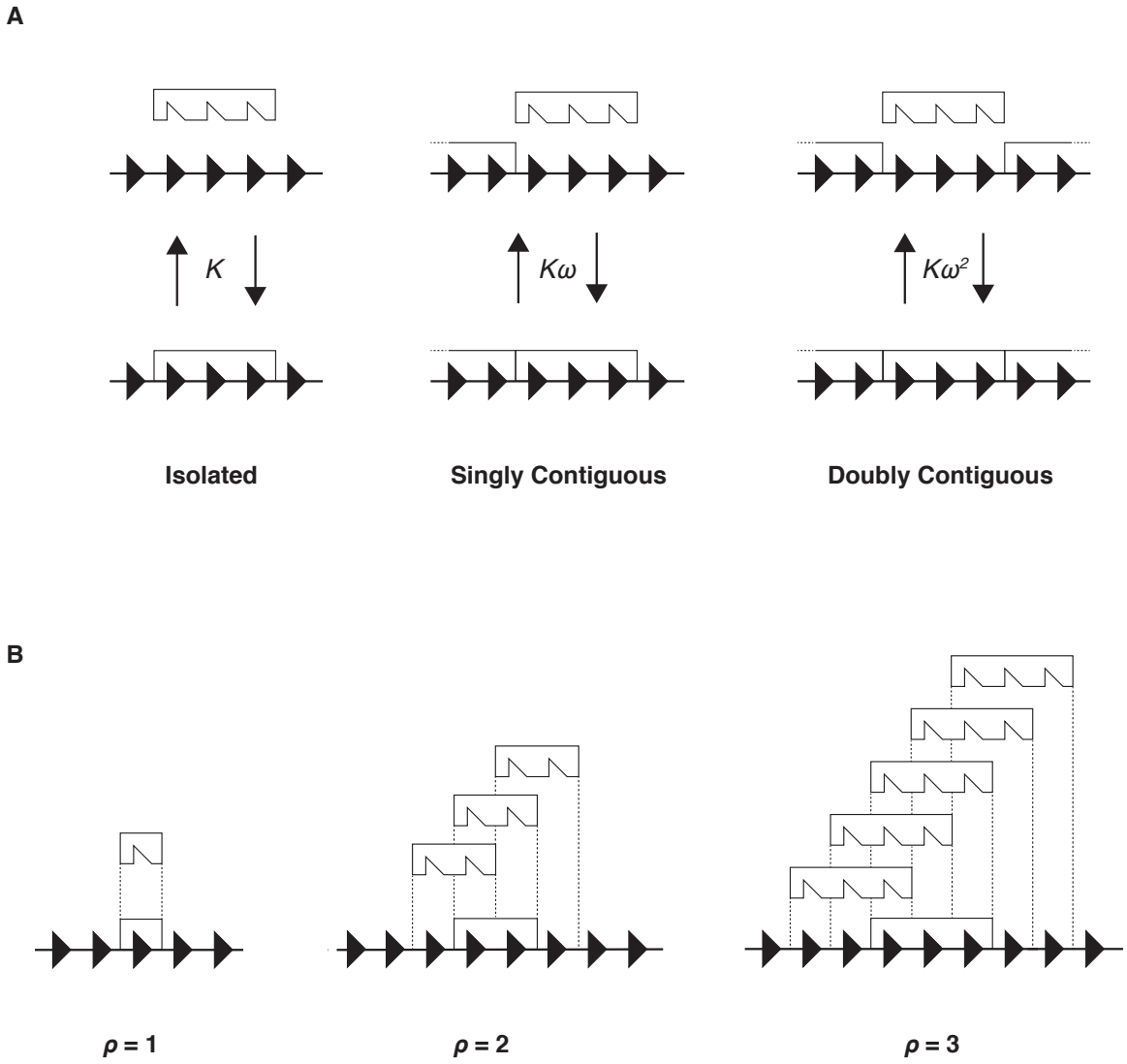


Figure 6.2: The Mcghee & von Hippel model. **A**. The binding of a protein molecule (here depicted with $\rho = 3$) to isolated DNA sites is characterised by the affinity constant K. Binding to singly contiguous sites is characterised by $K\omega$, and binding to doubly contiguous sites characterised by $K\omega^2$. Based on Fig. 1 from [141]. **B**. When $\rho = 1$, only a single binding site is excluded. When $\rho = 2$, three binding sites are excluded, and when $\rho = 3$, five binding sites are excluded. The number of excluded binding sites is given by $2\rho - 1$. Based on Fig. 2a from [141].

Interestingly, this model was previously used to analyse cooperativity in RAD51 binding to ssDNA [140]. The study analysed RAD51 binding to ssDNA via SPR (although the SPR experiments never reached equilibrium) using different RAD51 injection concentrations and fit the cooperativity model to these data under the assumption that RAD51 polymerisation occurs only via monomer addition. The model fit suggested RAD51 cooperativity increases $\approx$ 3-fold in the presence of ATP ($\omega \approx 27$ in the absence of nucleotide, $\omega \approx 83$ in the presence of ATP), while the intrinsic binding constant K decreases $\approx$ 7-fold in the presence of ATP ($K \approx$ 1.26 × 10^6 in the absence of nucleotide,$K \approx$ 1.7 × 10^5 in the presence of ATP) [140]. This analysis suggested that ATP limits gap formation by enhancing RAD51 polymerisation on DNA instead of the adsorption of individual, interspersed RAD51 monomers [140].

Overall, the model provides a mechanistic description of protein binding to DNA that takes into account the interplay between cooperativity (ω) due to protein-protein interactions and inhibition due to overlapping binding sites that cause gap formation ($\rho \geq 2$), and is computationally inexpensive to fit using a least-squares methods [141, 236] given its closed form. However, the model has two main limitations. First, in this formulation it can only be used for analysis of equilibrium binding, while kinetic analysis involving determination of forward and reverse rate constants is not accounted for. Second, it cannot be applied in its current formulation to analyse binding data derived from mixtures of different protein species with different ρ values (e.g. pre-formed RAD51 polymers of different lengths in solution) binding to DNA. More recently, the model has been generalised to account for binding of two protein species [237] and of more than two protein species [238, 239], although these models formulations only support analysis of equilibrium binding data.

In comparison, the dimer and n-mer models developed in this thesis enable de-

termination of kinetic rate constants. This is particularly important given that the inverse correlation between DNA flexibility and F86E RAD51 lifetime on DNA (chapter 5) was detected by fitting the dimer model dissociation rate constant ($\mathbf{k_d}$) and the n-mer model polymerisation reverse rate constant ($\mathbf{k_{pr}}$) directly to the SPR data. Using the Mcghee & von Hippel model [141] to fit these data, it would not have been possible to determine whether F86E RAD51 has a lower association rate constant or higher dissociation rate constant on flexible DNA. Furthermore, it has been proposed that RAD51 polymerisation on DNA occurs via the addition of pre-formed polymers [12], a mechanism which cannot be accounted for using the Mcghee & von Hippel model [141]. Indeed, the n-mer model was formulated as an extension to the dimer model to characterise RAD51 binding kinetics to DNA by considering the abundance of pre-formed RAD51 polymers in solution determined by its protomer-protomer affinity ($\mathbf{K_D}$) [141].

However, the dimer and n-mer models assume non-overlapping binding sites, and therefore neglect the inhibitory effect of overlapping binding sites. While this effect may play a role in determining the kinetics in my SPR experiments, I would argue this effect is most likely negligible for three reasons. First, it was proposed that RAD51 in the presence of ATP has an 83-fold higher preference for binding to a site adjacent to a DNA-bound RAD51 protomer rather than to an isolated DNA binding site ($\omega \approx 83$), suggesting RAD51 polymerisation on DNA involves little to no gap formation [140]. Second, an isolated RAD51 molecule is likely to be unstable on DNA (especially ssDNA, as suggested by the dimer model $\mathbf{k_{ad.f}}$ and $\mathbf{k_{ad.r}}$ fits, Table 3.1), and therefore unlikely to maintain gaps on DNA. In other words, given that nucleation appears to be slower than elongation ($\mathbf{k_{ad.f}}$ and $\mathbf{k_{el.f}}$, Table 3.1), I expect that once a nucleus forms it will elongate very rapidly. Thus, although it may be possible for two nuclei to form on a single filament, this is a low probability event. Third, it was more recently shown using scanning force microscopy that RAD51

nucleates on average once every 500 bps on dsDNA (when RAD51 is incubated with dsDNA at a molar ratio of 1 protomer per 3 bps) [137], suggesting RAD51 polymers on average elongate starting from a single nucleus on individual dT-50, dN-50, dA-50 and dN-50p molecules (chapters 4, 5). In this scenario, a RAD51 polymer of length 15 on e.g. dT-50 could potentially leave a single 2 nt gap on one DNA end and a single 1 nt gap on the opposite end, therefore inhibiting the binding of another RAD51 protomer required to fully saturate the dT-50 molecule. I assume this effect is negligible given that 15 out of 16 binding sites per DNA molecule would nonetheless be available for RAD51 adsorption and polymer elongation.

Another potential drawback of using the dimer and n-mer models compared to the Mcghee & von Hippel model [141] is that the dimer and n-mer models do not explicitly account for cooperativity (ω). The dimer and n-mer models implicitly account for cooperativity by describing the binding and dissociation of RAD51 polymers from DNA using different dissociation rate constants (Figs. 3.3A-B, 3.8A-B). If SPR data in chapters 4 and 5 had been obtained at equilibrium, the Mcghee & von Hippel model could have been used to test whether the increase in RAD51 polymerisation speed due to increasing DNA flexibility (Figs. 4.5C, 5.8C) were either explained via different intrinsic association constants (i.e. different K values) and/or different degrees of cooperativity (i.e. different ω values) for dT-50, dN-50, dA-50 and dN-50p.

6.3.2 The Schwarz & Watanabe Model

The model developed by Schwarz & Watanabe [235] represents an extension of the Mcghee & von Hippel model [141] to account for kinetics of protein-DNA binding. Here, binding to DNA binding sites is described via an association rate constant k_{on}, a dissociation rate constant k_{off}, the cooperativity factor ω, and two extra kinetic parameters γ' and γ'' that allow to define differences in association between

isolated, singly contiguous and doubly contiguous binding sites (Fig. 6.3). The authors did not develop a closed-form solution to the kinetic ODE system, but derived strategies to approximate the rate constants under simplifying conditions. One example is the condition of irreversible dissociation [235], which can be considered analogous to the dissociation phase of an SPR experiment where only analyte dissociation occurs (Fig. 1.9A).

The authors assumed the rate of dissociation from within the DNA-bound protein polymers can be neglected compared to that from the polymer ends, given a sufficiently high ω [235]. In this situation, dissociation of protein molecules from isolated sites is also negligible so that only dissociation from polymer ends is significant, and a single relaxation time τ (i.e. the time required to reach equilibrium) describes the kinetics of the system:

$$\tau = \frac{1}{2k'_{\text{off}}}\sqrt{\omega/\rho}\sqrt{\frac{\theta_0}{1-\theta_0}} \tag{6.4}$$

Where $k'_{\text{off}} = k_{\text{off}}/\omega\gamma'$ is the dissociation rate constant for singly contiguous sites, ω is the cooperativity factor, ρ is the number of DNA bp or nt occupied by a single protein molecule, and θ_0 is the fractional amount of DNA nts or bps bound by protein molecules at the start of the irreversible dissociation. τ can often be extrapolated graphically by plotting θ from the beginning of the start of irreversible dissociation as a function of time and estimating the time for the experimental system to reach equilibrium (i.e. no change in θ). k'_{off} can then be determined by 1) directly fitting k'_{off} using this equation, or by 2) plotting $\frac{1}{\tau^2}$ as a function of the ratio of DNA nts or bps to total protein molecules in the system (p), and determining k'_{off} from the slope of the resulting straight line [235]. Indeed, the authors used a similar approach to determine the k'_{off} from kinetic data describing the interaction between fluorescein-labelled clupein Z and dsDNA [240].

Hence, the Schwarz & Watanabe model provides approximation of kinetic rate constants, such as the rate of unbinding of single protomers from DNA-bound polymers (k'_{off}), for protein-DNA interactions under simplifying conditions, providing additional information compared to the Mcghee & von Hippel model. Similarly to the Mcghee & von Hippel model, its disadvantage is that it cannot be easily generalised to analyse kinetic data derived from mixtures of different protein species with different ρ values binding to DNA [235], such as pre-formed RAD51 polymers of different lengths [12].

Both the simplified dimer model used to analyse the F86E RAD51 SPR data (Fig. 5.4A) and the Schwarz & Watanabe model assume protein molecules bind to DNA as monomers. Hence, if SPR data in chapter 5 using F86E RAD51 had been obtained at equilibrium, equation 6.4 could have been used in addition to the simplified dimer model to obtain estimates of the F86E RAD51 k'_{off} values on dT-50, dN-50, dA-50 and dN-50p. This could have provided further support for the correlation between DNA flexibility and lifetime of F86E RAD51 molecules on DNA (Fig. 5.6B, D).

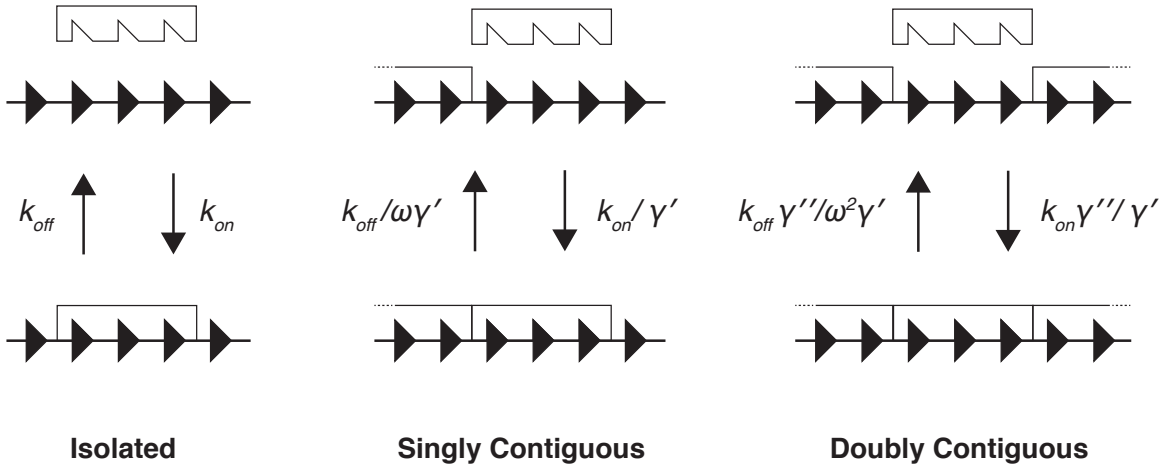


Figure 6.3: The Schwarz & Watanabe model. The binding of a protein molecule (here depicted with $\rho = 3$) to isolated DNA sites is characterised by a forward rate constant k_{on} and a dissociation rate constant k_{off}. Binding to singly contiguous sites is characterised by $\frac{k_{\text{on}}}{\gamma'}$ and $\frac{k_{\text{off}}}{\omega\gamma'}$, and binding to doubly contiguous sites is characterised by $\frac{k_{\text{on}}\gamma''}{\gamma'}$ and $\frac{k_{\text{off}}\gamma''}{\omega^2\gamma'}$. ω is the cooperativity parameter defined in [141]. γ' and γ'' are positive constants that allow to define differences between the association rate constants, e.g. if $\gamma' < 1$, $\frac{k_{\text{on}}}{\gamma'} > k_{\text{on}}$. Based on Fig. 6 and section 4 from [235].

6.4 Future Work

6.4.1 Validation of the DUET Model

To further test the DUET model it would be useful to measure RAD51 binding kinetics to dsDNA molecules of different flexibilities via SPR. Given that lower GC content increases dsDNA flexibility [241], one strategy would be to use dsDNA dN-50p molecules of lower or higher GC content. This would allow to verify whether dsDNA flexibility correlates with RAD51 polymerisation speed, similarly to what was shown for ssDNA with dT-50, dN-50 and dA-50 using both WT and F86E RAD51 (Figs. 4.3B-C, 4.5B-C, 5.6B-C, 5.8B-C).

In vivo, RPA is recruited to ssDNA and represents a barrier to RAD51 recruitment [7]. While *in vitro* RAD51 can remove RPA from ssDNA in the absence of other protein co-factors [38], whether ssDNA flexibility affects RPA binding kinetics, and whether RPA stiffens ssDNA and consequently alters RAD51 binding kinetics to ssDNA is currently unknown. *In vivo*, it was shown that poly-dA tracts flanking DNA replication initiation sites are not bound by RPA and are susceptible to DSBs [233], suggesting flexible ssDNA is required for RPA binding and consequent replication fork protection. How ssDNA flexibility affects RPA binding kinetics could be further analysed by measure RPA binding kinetics to dA-50, dN-50 and dT-50 via SPR and fitting the Langmuir model [195] (section 1.5.2) to these data. Additionally, another interesting avenue for future work would be to 1) measure the flexibility of RPA-coated ssDNA dT-50, dN-50 and dA-50 (assuming detectable RPA binding to dA-50) via SAXS, 2) assess the RAD51 binding kinetics on RPA-coated ssDNA dT-50, dN-50 and dA-50 using SPR, and 3) fit the n-mer model from chapter 3 to these SPR data. This would provide insight as to whether RPA potentially stiffens ssDNA and consequently alters RAD51 adsorption and/or polymer elongation kinetics *in vivo*.

BRCA2 is a key RAD51 binding partner and is thought to facilitate RAD51 binding to ssDNA via its BRC motifs and C-terminal TR2 domain [3, 4, 5, 6, 41, 42]. However, the impact of BRCA2 on RAD51 binding kinetics to ssDNA and dsDNA is not well characterised. For this reason, an interesting future experiment could be to 1) measure RAD51 binding kinetics to ssDNA dT-50, dN-50, dA-50 and dsDNA dN-50p in the presence of purified BRCA2 and 2) fit the ssDNA and dsDNA n-mer models from chapter 3 to these data to understand how DNA flexibility affects the ability of BRCA2 to bind DNA via its OB-folds and modify RAD51 polymerisation kinetics. Given that BRCA2 has been proposed to bind monomeric RAD51 via its BRC motifs [6], this approach would also be useful to understand whether BRCA2 reduces the average length of RAD51 polymers in solution and on DNA. This approach may then be further expanded to explore how multiple protein cofactors, such as BRCA2 in complex with PALB2 and BRCA1, synergistically modulate RAD51 polymerisation kinetics and how these synergistic effects may depend on DNA flexibility. In particular, an interesting experiment could be to assess how both BRCA2 and the Shu complex, which are proposed to bind the 3' and 5' ends of RAD51 presynaptic filaments respectively [44, 54], modulate RAD51 polymerisation kinetics.

Additionally, BRCA2 in complex with DSS1 is known to remove RPA and facilitate RAD51 polymerisation on ssDNA [40, 7]. Assuming that RPA does not stiffen DNA drammatically, a further intriguing set of experiments could aim to evaluate the effect of DNA flexibility on RAD51 binding kinetics to RPA-coated ssDNA dT50, dN50 and dA50 in the presence of the BRCA2-DSS1 complex. This would provide further insight into how the kinetics of the BRCA2-DSS1-dependent RPA-RAD51 exchange on ssDNA may be affected by the flexibility of the RPA-bound DNA.

6.4.2 Single Molecule Approaches

Experimental data on RAD51 binding kinetics to DNA in this thesis were generated using SPR. As mentioned in section 1.5, SPR enables highly reproducible characterisation of biomolecular interactions in real-time using small sample volumes of label-free analyte. However, SPR measures ensemble analyte binding kinetics to an immobilised ligand and does not provide information on single molecule binding kinetics. For this reason, the minimum RAD51 nucleation lengths on ssDNA and dsDNA and the fitted rate constants are based on population averages and not on single molecule observations. To further validate the observations from this study, it would be useful to carry out optical or magnetic tweezer experiments measuring RAD51 binding kinetics to single DNA molecules of variable flexibility [109, 11, 110, 12, 138]. Most likely, these analyses would have to be initially carried out using longer DNA molecules (> 100 base pairs / nucleotides) compared to the molecules used in this study, as DNA molecules shorter than 100 base pairs / nucleotides are often difficult to manipulate. A stochastic model based on, for example, the Gillespie algorithm [242], could be fit to these data to confirm or refine the RAD51 minimal nucleation lengths on ssDNA and dsDNA, as well as the structure of the polymerisation models and the values of the kinetic rate constants presented in this study.

Appendices

Appendix A

BioNetGen Code

A.1 ssDNA Dimer Model

```
begin parameters

kadf
kadr
kelf
kelr
ks

S_start
Rad51monomer
Rad51dimer

kobs1d kadf * Rad51dimer
kobs1m kadf * Rad51monomer

kobs2d kelf * Rad51dimer
kobs2m kelf * Rad51monomer

end parameters

begin molecule types

S_0()
S_1()
S_2()
S_3()
S_4()
S_5()
S_6()
```

```
S_7()
S_8()
S_9()
S_10()
S_11()
S_12()
S_13()
S_14()
S_15()
S_16()

end molecule types

begin seed species

S_0()        S_start

end seed species

begin reaction rules

#Dimer Reactions

S_0() -> S_2()                kobs1d
S_1() -> S_3()                kobs1d
S_2() -> S_4()                kobs2d
S_3() -> S_5()                kobs2d
S_4() -> S_6()                kobs2d
S_5() -> S_7()                kobs2d
S_6() -> S_8()                kobs2d
S_7() -> S_9()                kobs2d
S_8() -> S_10()                kobs2d
S_9() -> S_11()                kobs2d
S_10() -> S_12()        kobs2d
S_11() -> S_13()        kobs2d
S_12() -> S_14()        kobs2d
S_13() -> S_15()        kobs2d
S_14() -> S_16()        kobs2d

#Monomer Reactions

S_0() <-> S_1()                kobs1m, kadr
S_1() <-> S_2()                kobs1m, kadr
S_2() <-> S_3()                kobs1m, kadr
S_3() <-> S_4()                kobs2m, kelr
```

```
S_4() <-> S_5()                kobs2m, ks
S_5() <-> S_6()                kobs2m, ks
S_6() <-> S_7()                kobs2m, ks
S_7() <-> S_8()                kobs2m, ks
S_8() <-> S_9()                kobs2m, ks
S_9() <-> S_10()         kobs2m, ks
S_10() <-> S_11()         kobs2m, ks
S_11() <-> S_12()         kobs2m, ks
S_12() <-> S_13()         kobs2m, ks
S_13() <-> S_14()         kobs2m, ks
S_14() <-> S_15()         kobs2m, ks
S_15() <-> S_16()         kobs2m, ks

end reaction rules

begin observables

Molecules TotalRad51

S_1(),S_2(),S_3(),S_4(),S_5(),S_6(),S_7(),S_8(),
S_9(),S_10(),S_11(),S_12(),S_13(),S_14(),S_15(),S_16()

end observables

generate_network({overwrite=>1})
writeMfile({})
```

The BioNetGen code is divided into several sections. In the first section (`begin parameters`) all model parameters, including kinetic rate constants, are defined. In the second section (`begin molecule types`), all possible RAD51 polymer lengths on ssDNA are defined. In the third section (`begin seed species`), the initial condition for S_0 is defined ($S_0(0) =$`S_start` $= 1$). In the fourth section (`begin reaction rules`), all allowed RAD51 polymerisation reactions are defined with their associated kinetic rate constants. The final section (`begin observables`) defines the desired model output (in this case $\sum_1^{16} iS_i$).

A.2 dsDNA Dimer Model

```
begin parameters

kadf
kadr
kelf
ks

D_start
Rad51monomer
Rad51dimer

kobs1d kadf * Rad51dimer
kobs1m kadf * Rad51monomer

kobs2d kelf * Rad51dimer
kobs2m kelf * Rad51monomer

end parameters

begin molecule types

D_0()
D_1()
D_2()
D_3()
D_4()
D_5()
D_6()
D_7()
D_8()
D_9()
D_10()
D_11()
D_12()
D_13()
D_14()
D_15()
D_16()

end molecule types

begin seed species
```

```
D_0()        D_start

end seed species

begin reaction rules

#Dimer Reactions

D_0() -> D_2()                kobs1d
D_1() -> D_3()                kobs2d
D_2() -> D_4()                kobs2d
D_3() -> D_5()                kobs2d
D_4() -> D_6()                kobs2d
D_5() -> D_7()                kobs2d
D_6() -> D_8()                kobs2d
D_7() -> D_9()                kobs2d
D_8() -> D_10()                kobs2d
D_9() -> D_11()                kobs2d
D_10() -> D_12()        kobs2d
D_11() -> D_13()        kobs2d
D_12() -> D_14()        kobs2d
D_13() -> D_15()        kobs2d
D_14() -> D_16()        kobs2d

#Monomer Reactions

D_0() <-> D_1()                kobs1m, kadr
D_1() <-> D_2()                kobs2m, ks
D_2() <-> D_3()                kobs2m, ks
D_3() <-> D_4()                kobs2m, ks
D_4() <-> D_5()                kobs2m, ks
D_5() <-> D_6()                kobs2m, ks
D_6() <-> D_7()                kobs2m, ks
D_7() <-> D_8()                kobs2m, ks
D_8() <-> D_9()                kobs2m, ks
D_9() <-> D_10()        kobs2m, ks
D_10() <-> D_11()        kobs2m, ks
D_11() <-> D_12()        kobs2m, ks
D_12() <-> D_13()        kobs2m, ks
D_13() <-> D_14()        kobs2m, ks
D_14() <-> D_15()        kobs2m, ks
D_15() <-> D_16()        kobs2m, ks

end reaction rules
```

```
begin observables

Molecules TotalRad51

D_1(),D_2(),D_3(),D_4(),D_5(),D_6(),D_7(),D_8(),
D_9(),D_10(),D_11(),D_12(),D_13(),D_14(),D_15(),D_16()

end observables

generate_network({overwrite=>1})
writeMfile({})
```

In the first section (`begin parameters`) all model parameters, including kinetic rate constants, are defined. In the second section (`begin molecule types`), all possible RAD51 polymer lengths on dsDNA are defined. In the third section (`begin seed species`), the initial condition for D_0 is defined ($D_0(0)$ =`D_start` $= 1$). In the fourth section (`begin reaction rules`), all allowed RAD51 polymerisation reactions are defined with their associated kinetic rate constants. The final section (`begin observables`) defines the desired model output (in this case $\sum_1^{16} iD_i$).

A.3 n-mer Equilibrium Sub-Model

```
begin parameters

ka
kd
Rad51m

end parameters

begin molecule types

R_1()
R_2()
R_3()
R_4()
R_5()
R_6()
R_7()
R_8()
R_9()
R_10()
R_11()
R_12()
R_13()
R_14()
R_15()
R_16()

end molecule types

begin seed species

R_1()        Rad51m

end seed species

begin reaction rules

R_1() + R_1() <-> R_2()                ka, kd
R_1() + R_2() <-> R_3()                ka, kd
R_1() + R_3() <-> R_4()                ka, kd
R_1() + R_4() <-> R_5()                ka, kd
R_1() + R_5() <-> R_6()                ka, kd
```

```
R_1() + R_6() <-> R_7()              ka, kd
R_1() + R_7() <-> R_8()              ka, kd
R_1() + R_8() <-> R_9()              ka, kd
R_1() + R_9() <-> R_10()              ka, kd
R_1() + R_10() <-> R_11()              ka, kd
R_1() + R_11() <-> R_12()              ka, kd
R_1() + R_12() <-> R_13()              ka, kd
R_1() + R_13() <-> R_14()              ka, kd
R_1() + R_14() <-> R_15()              ka, kd
R_1() + R_15() <-> R_16()              ka, kd

R_2() + R_2() <-> R_4()              ka, kd
R_2() + R_3() <-> R_5()              ka, kd
R_2() + R_4() <-> R_6()              ka, kd
R_2() + R_5() <-> R_7()              ka, kd
R_2() + R_6() <-> R_8()              ka, kd
R_2() + R_7() <-> R_9()              ka, kd
R_2() + R_8() <-> R_10()              ka, kd
R_2() + R_9() <-> R_11()              ka, kd
R_2() + R_10() <-> R_12()              ka, kd
R_2() + R_11() <-> R_13()              ka, kd
R_2() + R_12() <-> R_14()              ka, kd
R_2() + R_13() <-> R_15()              ka, kd
R_2() + R_14() <-> R_16()              ka, kd

R_3() + R_3() <-> R_6()              ka, kd
R_3() + R_4() <-> R_7()              ka, kd
R_3() + R_5() <-> R_8()              ka, kd
R_3() + R_6() <-> R_9()              ka, kd
R_3() + R_7() <-> R_10()              ka, kd
R_3() + R_8() <-> R_11()              ka, kd
R_3() + R_9() <-> R_12()              ka, kd
R_3() + R_10() <-> R_13()              ka, kd
R_3() + R_11() <-> R_14()              ka, kd
R_3() + R_12() <-> R_15()              ka, kd
R_3() + R_13() <-> R_16()              ka, kd

R_4() + R_4() <-> R_8()              ka, kd
R_4() + R_5() <-> R_9()              ka, kd
R_4() + R_6() <-> R_10()              ka, kd
R_4() + R_7() <-> R_11()              ka, kd
R_4() + R_8() <-> R_12()              ka, kd
R_4() + R_9() <-> R_13()              ka, kd
R_4() + R_10() <-> R_14()              ka, kd
R_4() + R_11() <-> R_15()              ka, kd
```

```
R_4() + R_12() <-> R_16()                ka, kd

R_5() + R_5() <-> R_10()                ka, kd
R_5() + R_6() <-> R_11()                ka, kd
R_5() + R_7() <-> R_12()                ka, kd
R_5() + R_8() <-> R_13()                ka, kd
R_5() + R_9() <-> R_14()                ka, kd
R_5() + R_10() <-> R_15()                ka, kd
R_5() + R_11() <-> R_16()                ka, kd

R_6() + R_6() <-> R_12()                ka, kd
R_6() + R_7() <-> R_13()                ka, kd
R_6() + R_8() <-> R_14()                ka, kd
R_6() + R_9() <-> R_15()                ka, kd
R_6() + R_10() <-> R_16()                ka, kd

R_7() + R_7() <-> R_14()                ka, kd
R_7() + R_8() <-> R_15()                ka, kd
R_7() + R_9() <-> R_16()                ka, kd

R_8() + R_8() <-> R_16()                ka, kd

end reaction rules

begin observables

Molecules Rad51_16mers R_16()
Molecules Rad51_15mers R_15()
Molecules Rad51_14mers R_14()
Molecules Rad51_13mers R_13()
Molecules Rad51_12mers R_12()
Molecules Rad51_11mers R_11()
Molecules Rad51_10mers R_10()
Molecules Rad51_9mers R_9()
Molecules Rad51_Octamers R_8()
Molecules Rad51_Heptamers R_7()
Molecules Rad51_Hexamers R_6()
Molecules Rad51_Pentamers R_5()
Molecules Rad51_Tetramers R_4()
Molecules Rad51_Trimers R_3()
Molecules Rad51_Dimers R_2()
Molecules Rad51_Monomers R_1()

end observables
```

```
generate_network({overwrite=>1})
writeMfile({})
```

In the first section (`begin parameters`) all model parameters, including kinetic rate constants, are defined. In the second section (`begin molecule types`), all possible RAD51 polymer lengths in solution are defined. In the third section (`begin seed species`), the initial condition for R_1 is defined ($R_1(0) =$`Rad51m` $=$ RAD51 injection concentration). In the fourth section (`begin reaction rules`), all allowed RAD51 polymerisation reactions in solution are defined with their associated kinetic rate constants. The final section (`begin observables`) defines the desired model outputs (in this case nR_n for $n = 1, ..., 16$).

A.4 n-mer General ssDNA and dsDNA Kinetic Sub-Models

```
begin parameters

###CONCENTRATIONS FROM RAD51 EQUILIBRIUM SUB-MODEL
Rad51_monomer
Rad51_dimer
Rad51_trimer
Rad51_tetramer
Rad51_pentamer
Rad51_hexamer
Rad51_heptamer
Rad51_8mer
Rad51_9mer
Rad51_10mer
Rad51_11mer
Rad51_12mer
Rad51_13mer
Rad51_14mer
Rad51_15mer
Rad51_16mer
#####################################################

DNA_start

kp

kd_1
kd_2
kd_3
kd_4
kd_5
kd_6
kd_7
kd_8
kd_9
kd_10
kd_11
kd_12
kd_13
kd_14
kd_15
```

```
kd_16

ks

kadf_obs_1        kp * Rad51_monomer
kadf_obs_2        kp * Rad51_dimer
kadf_obs_3        kp * Rad51_trimer
kadf_obs_4        kp * Rad51_tetramer
kadf_obs_5        kp * Rad51_pentamer
kadf_obs_6        kp * Rad51_hexamer
kadf_obs_7        kp * Rad51_heptamer
kadf_obs_8        kp * Rad51_8mer
kadf_obs_9        kp * Rad51_9mer
kadf_obs_10        kp * Rad51_10mer
kadf_obs_11        kp * Rad51_11mer
kadf_obs_12        kp * Rad51_12mer
kadf_obs_13        kp * Rad51_13mer
kadf_obs_14        kp * Rad51_14mer
kadf_obs_15        kp * Rad51_15mer
kadf_obs_16        kp * Rad51_16mer

kelf_obs_1        kp * Rad51_monomer
kelf_obs_2        kp * Rad51_dimer
kelf_obs_3        kp * Rad51_trimer
kelf_obs_4        kp * Rad51_tetramer
kelf_obs_5        kp * Rad51_pentamer
kelf_obs_6        kp * Rad51_hexamer
kelf_obs_7        kp * Rad51_heptamer
kelf_obs_8        kp * Rad51_8mer
kelf_obs_9        kp * Rad51_9mer
kelf_obs_10        kp * Rad51_10mer
kelf_obs_11        kp * Rad51_11mer
kelf_obs_12        kp * Rad51_12mer
kelf_obs_13        kp * Rad51_13mer
kelf_obs_14        kp * Rad51_14mer
kelf_obs_15        kp * Rad51_15mer


end parameters

begin molecule types

DNA_0()
DNA_1()
DNA_2()
```

```
DNA_3()
DNA_4()
DNA_5()
DNA_6()
DNA_7()
DNA_8()
DNA_9()
DNA_10()
DNA_11()
DNA_12()
DNA_13()
DNA_14()
DNA_15()
DNA_16()

end molecule types

begin seed species

DNA_0()        DNA_start

end seed species

begin reaction rules

#ADSORPTION REACTIONS

DNA_0()        <-> DNA_1()                kadf_obs_1, kd_1
DNA_0() <-> DNA_2()               kadf_obs_2, kd_2
DNA_0() <-> DNA_3()               kadf_obs_3, kd_3
DNA_0() <-> DNA_4()               kadf_obs_4, kd_4
DNA_0() <-> DNA_5()               kadf_obs_5, kd_5
DNA_0() <-> DNA_6()               kadf_obs_6, kd_6
DNA_0() <-> DNA_7()               kadf_obs_7, kd_7
DNA_0() <-> DNA_8()               kadf_obs_8, kd_8
DNA_0() <-> DNA_9()               kadf_obs_9, kd_9
DNA_0() <-> DNA_10()               kadf_obs_10, kd_10
DNA_0() <-> DNA_11()               kadf_obs_11, kd_11
DNA_0() <-> DNA_12()               kadf_obs_12, kd_12
DNA_0() <-> DNA_13()               kadf_obs_13, kd_13
DNA_0() <-> DNA_14()               kadf_obs_14, kd_14
DNA_0() <-> DNA_15()               kadf_obs_15, kd_15
DNA_0() <-> DNA_16()               kadf_obs_16, kd_16

#MONOMER ELONGATION REACTIONS
```

```
DNA_1()        <-> DNA_2()              kelf_obs_1, ks
DNA_2()        <-> DNA_3()              kelf_obs_1, ks
DNA_3()        <-> DNA_4()              kelf_obs_1, ks
DNA_4() <-> DNA_5()              kelf_obs_1, ks
DNA_5() <-> DNA_6()              kelf_obs_1, ks
DNA_6() <-> DNA_7()              kelf_obs_1, ks
DNA_7() <-> DNA_8()              kelf_obs_1, ks
DNA_8() <-> DNA_9()              kelf_obs_1, ks
DNA_9() <-> DNA_10()              kelf_obs_1, ks
DNA_10() <-> DNA_11()              kelf_obs_1, ks
DNA_11() <-> DNA_12()              kelf_obs_1, ks
DNA_12() <-> DNA_13()              kelf_obs_1, ks
DNA_13() <-> DNA_14()              kelf_obs_1, ks
DNA_14() <-> DNA_15()              kelf_obs_1, ks
DNA_15() <-> DNA_16()              kelf_obs_1, ks

#DIMER ELONGATION REACTIONS

DNA_1() -> DNA_3()              kelf_obs_2
DNA_2() -> DNA_4()              kelf_obs_2
DNA_3() -> DNA_5()              kelf_obs_2
DNA_4() -> DNA_6()              kelf_obs_2
DNA_5() -> DNA_7()              kelf_obs_2
DNA_6() -> DNA_8()              kelf_obs_2
DNA_7() -> DNA_9()              kelf_obs_2
DNA_8() -> DNA_10()              kelf_obs_2
DNA_9() -> DNA_11()              kelf_obs_2
DNA_10() -> DNA_12()              kelf_obs_2
DNA_11() -> DNA_13()              kelf_obs_2
DNA_12() -> DNA_14()              kelf_obs_2
DNA_13() -> DNA_15()              kelf_obs_2
DNA_14() -> DNA_16()              kelf_obs_2

#TRIMER ELONGATION REACTIONS

DNA_1() -> DNA_4()              kelf_obs_3
DNA_2() -> DNA_5()              kelf_obs_3
DNA_3() -> DNA_6()              kelf_obs_3
DNA_4() -> DNA_7()              kelf_obs_3
DNA_5() -> DNA_8()              kelf_obs_3
DNA_6() -> DNA_9()              kelf_obs_3
DNA_7() -> DNA_10()              kelf_obs_3
DNA_8() -> DNA_11()              kelf_obs_3
DNA_9() -> DNA_12()              kelf_obs_3
```

```
DNA_10() -> DNA_13()                kelf_obs_3
DNA_11() -> DNA_14()                kelf_obs_3
DNA_12() -> DNA_15()                kelf_obs_3
DNA_13() -> DNA_16()                kelf_obs_3

#TETRAMER ELONGATION REACTIONS

DNA_1() -> DNA_5()                 kelf_obs_4
DNA_2() -> DNA_6()                 kelf_obs_4
DNA_3() -> DNA_7()                 kelf_obs_4
DNA_4() -> DNA_8()                 kelf_obs_4
DNA_5() -> DNA_9()                 kelf_obs_4
DNA_6() -> DNA_10()                 kelf_obs_4
DNA_7() -> DNA_11()                 kelf_obs_4
DNA_8() -> DNA_12()                 kelf_obs_4
DNA_9() -> DNA_13()                 kelf_obs_4
DNA_10() -> DNA_14()                 kelf_obs_4
DNA_11() -> DNA_15()                 kelf_obs_4
DNA_12() -> DNA_16()                 kelf_obs_4

#PENTAMER ELONGATION REACTIONS

DNA_1() -> DNA_6()                 kelf_obs_5
DNA_2() -> DNA_7()                 kelf_obs_5
DNA_3() -> DNA_8()                 kelf_obs_5
DNA_4() -> DNA_9()                 kelf_obs_5
DNA_5() -> DNA_10()                 kelf_obs_5
DNA_6() -> DNA_11()                 kelf_obs_5
DNA_7() -> DNA_12()                 kelf_obs_5
DNA_8() -> DNA_13()                 kelf_obs_5
DNA_9() -> DNA_14()                 kelf_obs_5
DNA_10() -> DNA_15()                 kelf_obs_5
DNA_11() -> DNA_16()                 kelf_obs_5

#HEXAMER ELONGATION REACTIONS

DNA_1() -> DNA_7()                 kelf_obs_6
DNA_2() -> DNA_8()                 kelf_obs_6
DNA_3() -> DNA_9()                 kelf_obs_6
DNA_4() -> DNA_10()                 kelf_obs_6
DNA_5() -> DNA_11()                 kelf_obs_6
DNA_6() -> DNA_12()                 kelf_obs_6
DNA_7() -> DNA_13()                 kelf_obs_6
DNA_8() -> DNA_14()                 kelf_obs_6
DNA_9() -> DNA_15()                 kelf_obs_6
```

```
DNA_10() -> DNA_16()                kelf_obs_6

#HEPTAMER ELONGATION REACTIONS

DNA_1() -> DNA_8()                 kelf_obs_7
DNA_2() -> DNA_9()                 kelf_obs_7
DNA_3() -> DNA_10()                 kelf_obs_7
DNA_4() -> DNA_11()                 kelf_obs_7
DNA_5() -> DNA_12()                 kelf_obs_7
DNA_6() -> DNA_13()                 kelf_obs_7
DNA_7() -> DNA_14()                 kelf_obs_7
DNA_8() -> DNA_15()                 kelf_obs_7
DNA_9() -> DNA_16()                 kelf_obs_7

#8-MER ELONGATION REACTIONS

DNA_1() -> DNA_9()                 kelf_obs_8
DNA_2() -> DNA_10()                 kelf_obs_8
DNA_3() -> DNA_11()                 kelf_obs_8
DNA_4() -> DNA_12()                 kelf_obs_8
DNA_5() -> DNA_13()                 kelf_obs_8
DNA_6() -> DNA_14()                 kelf_obs_8
DNA_7() -> DNA_15()                 kelf_obs_8
DNA_8() -> DNA_16()                 kelf_obs_8

#9-MER ELONGATION REACTIONS

DNA_1() -> DNA_10()                 kelf_obs_9
DNA_2() -> DNA_11()                 kelf_obs_9
DNA_3() -> DNA_12()                 kelf_obs_9
DNA_4() -> DNA_13()                 kelf_obs_9
DNA_5() -> DNA_14()                 kelf_obs_9
DNA_6() -> DNA_15()                 kelf_obs_9
DNA_7() -> DNA_16()                 kelf_obs_9

#10-MER ELONGATION REACTIONS

DNA_1() -> DNA_11()                 kelf_obs_10
DNA_2() -> DNA_12()                 kelf_obs_10
DNA_3() -> DNA_13()                 kelf_obs_10
DNA_4() -> DNA_14()                 kelf_obs_10
DNA_5() -> DNA_15()                 kelf_obs_10
DNA_6() -> DNA_16()                 kelf_obs_10

#11-MER ELONGATION REACTIONS
```

```
DNA_1() -> DNA_12()                kelf_obs_11
DNA_2() -> DNA_13()                kelf_obs_11
DNA_3() -> DNA_14()                kelf_obs_11
DNA_4() -> DNA_15()                kelf_obs_11
DNA_5() -> DNA_16()                kelf_obs_11

#12-MER ELONGATION REACTIONS

DNA_1() -> DNA_13()                kelf_obs_12
DNA_2() -> DNA_14()                kelf_obs_12
DNA_3() -> DNA_15()                kelf_obs_12
DNA_4() -> DNA_16()                kelf_obs_12

#13-MER ELONGATION REACTIONS

DNA_1() -> DNA_14()                kelf_obs_13
DNA_2() -> DNA_15()                kelf_obs_13
DNA_3() -> DNA_16()                kelf_obs_13

#14-MER ELONGATION REACTIONS

DNA_1() -> DNA_15()                kelf_obs_14
DNA_2() -> DNA_16()                kelf_obs_14

#15-MER ELONGATION REACTIONS

DNA_1() -> DNA_16()                kelf_obs_15

end reaction rules

begin observables

Molecules TotalRad51
 ↪  DNA_1(),DNA_2(),DNA_3(),DNA_4(),DNA_5(),DNA_6(),DNA_7(),DNA_8()
,DNA_9(),DNA_10(),DNA_11(),DNA_12(),DNA_13(),DNA_14(),DNA_15(),DNA_16()

end observables

generate_network({overwrite=>1})
writeMfile({})
```

In the first section (`begin parameters`) all model parameters, including kinetic rate constants, are defined. In the second section (`begin molecule types`), all possible RAD51 polymer lengths on ssDNA and dsDNA are defined. In the third section (`begin seed species`), the initial condition for S_0 and D_0 is defined ($S_0(0) = D_0(0) =$ `DNA_start` $= 1$). In the fourth section (`begin reaction rules`), all allowed RAD51 polymerisation reactions are defined with their associated kinetic rate constants. The final section (`begin observables`) defines the desired model output (in this case $\sum_1^{16} iD_i$).

Appendix B

Additional Figures

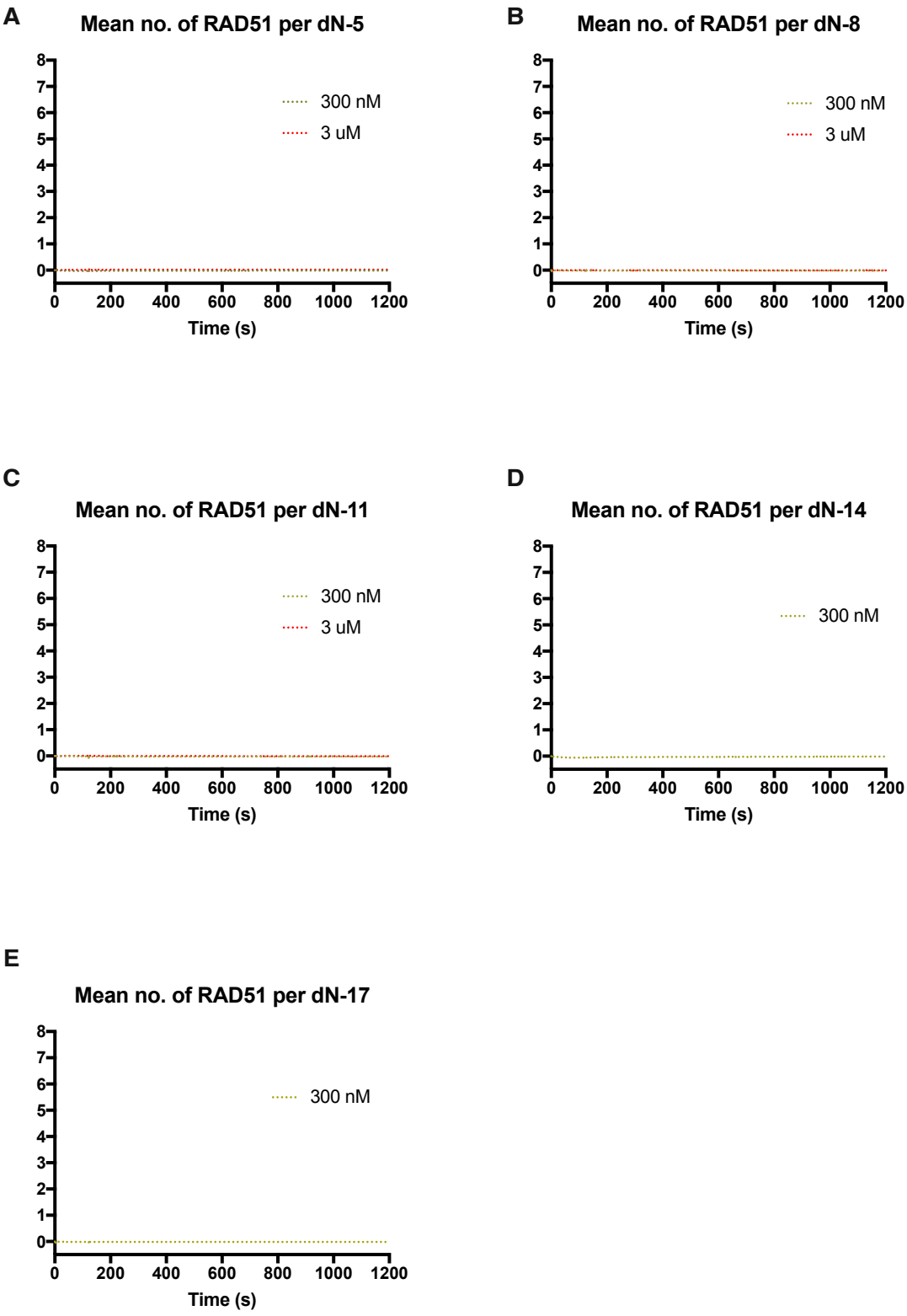


Figure B.1: WT RAD51 binding kinetics on dN-5 (**A**), dN-8 (**B**), dN-11 (**C**), dN-14 (**D**) and dN-17 (**E**). RAD51 binding was undetectable at 300 nM and 3 μM [RAD51] on dN-5, dN-8, dN-11, and undetectable at 300 nM [RAD51] on dN-14 and dN-17. These curves were not included in the model fitting (Figs. 3.2B, 3.7B). SPR data were collected and processed as described in sections 2.6 and 2.7.

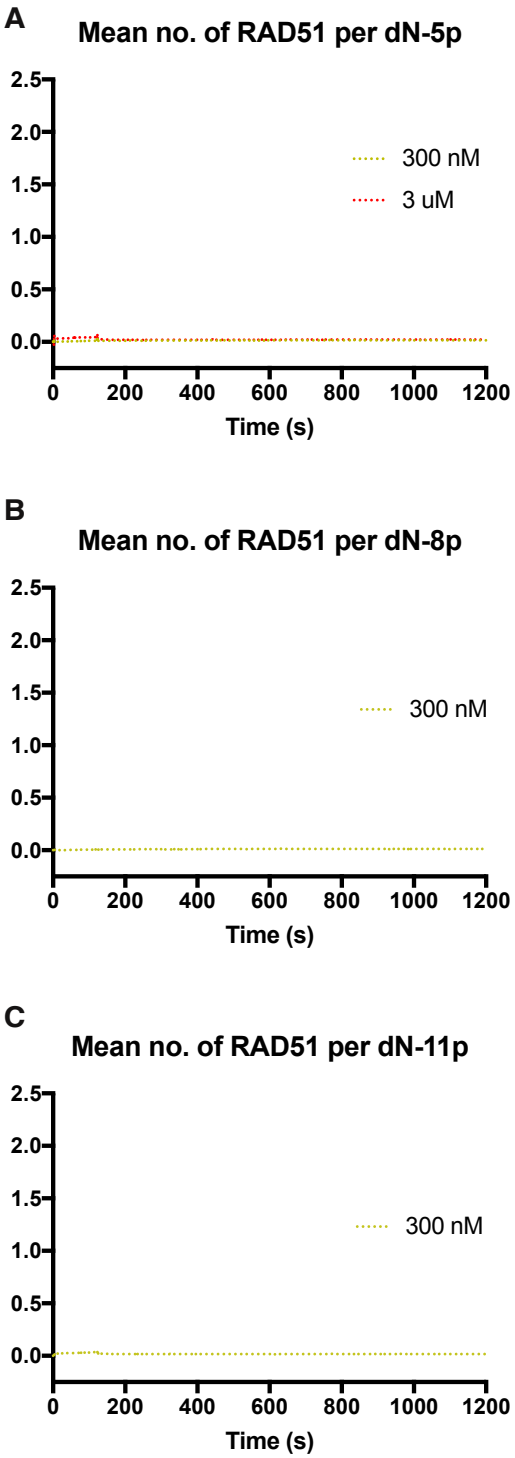


Figure B.2: WT RAD51 binding kinetics on dN-5p (**A**), dN-8p (**B**), and dN-11p (**C**). RAD51 binding was undetectable at 300 nM and 3 μM [RAD51] on dN-5p, and undetectable at 300 nM [RAD51] on dN-8p and dN-11p. These curves were not included in the model fitting (Fig. 3.2D, 3.7D). SPR data were collected and processed as described in sections 2.6 and 2.7.

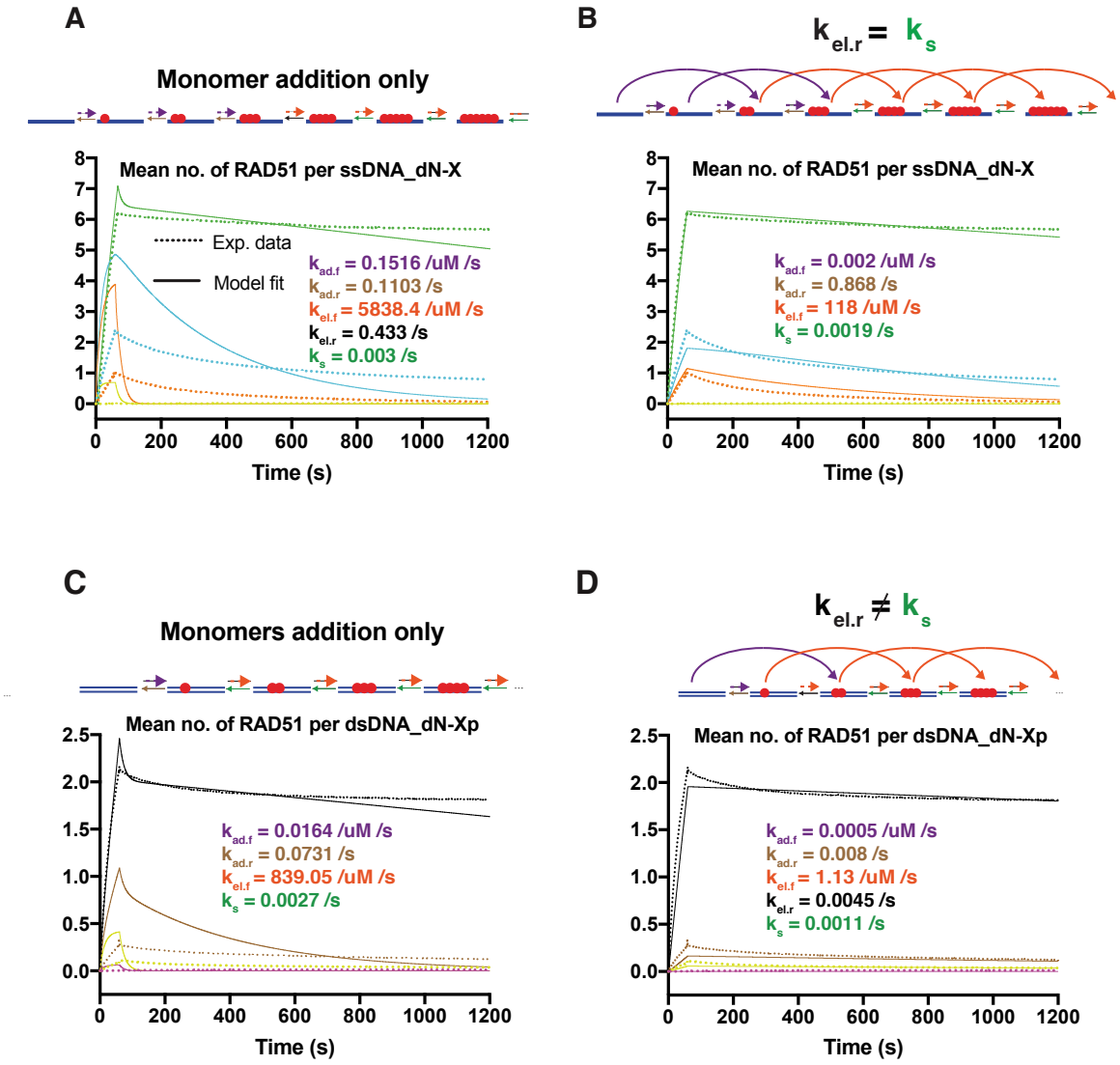


Figure B.3: ssDNA dN-X data fits for the ssDNA dimer models A - B, and dsDNA dN-Xp data fits for the dsDNA dimer models A and C. A kinetic representation of each alternative dimer model is shown, along with the associated parameter fits. **A**. ssDNA dimer model A describing RAD51 polymerisation only via monomer addition. **B**. ssDNA dimer model B ($\mathbf{k_{el.r} = k_s}$) describing RAD51 polymerisation only via four parameters ($\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_s}$). C. dsDNA dimer model A describing RAD51 polymerisation only via monomer addition. D.dsDNA dimer model C ($\mathbf{k_{el.r} \neq k_s}$) describing RAD51 polymerisation via five parameters ($\mathbf{k_{ad.f}}$, $\mathbf{k_{ad.r}}$, $\mathbf{k_{el.f}}$, $\mathbf{k_{el.r}}$, $\mathbf{k_s}$).

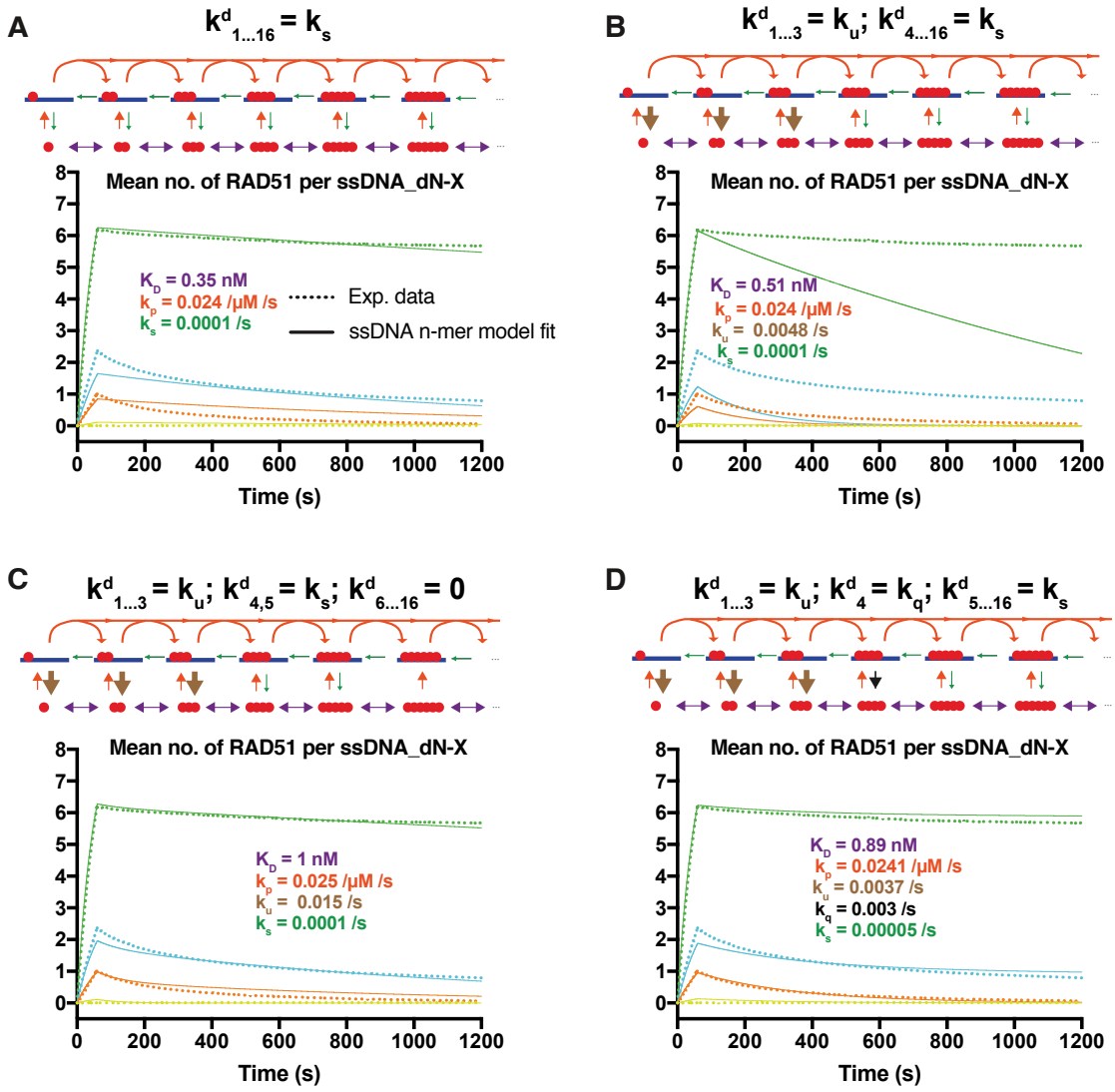


Figure B.4: ssDNA dN-X data fits for the ssDNA kinetic sub-models (KSMs) A - D. A kinetic representation of each KSM is shown above each plot, along with the associated simplifications compared to the general kinetic sub-model (Table 2.4) and the parameter fits. **A**. KSM A. **B**. KSM B. **C**. KSM C. **D**. KSM D.

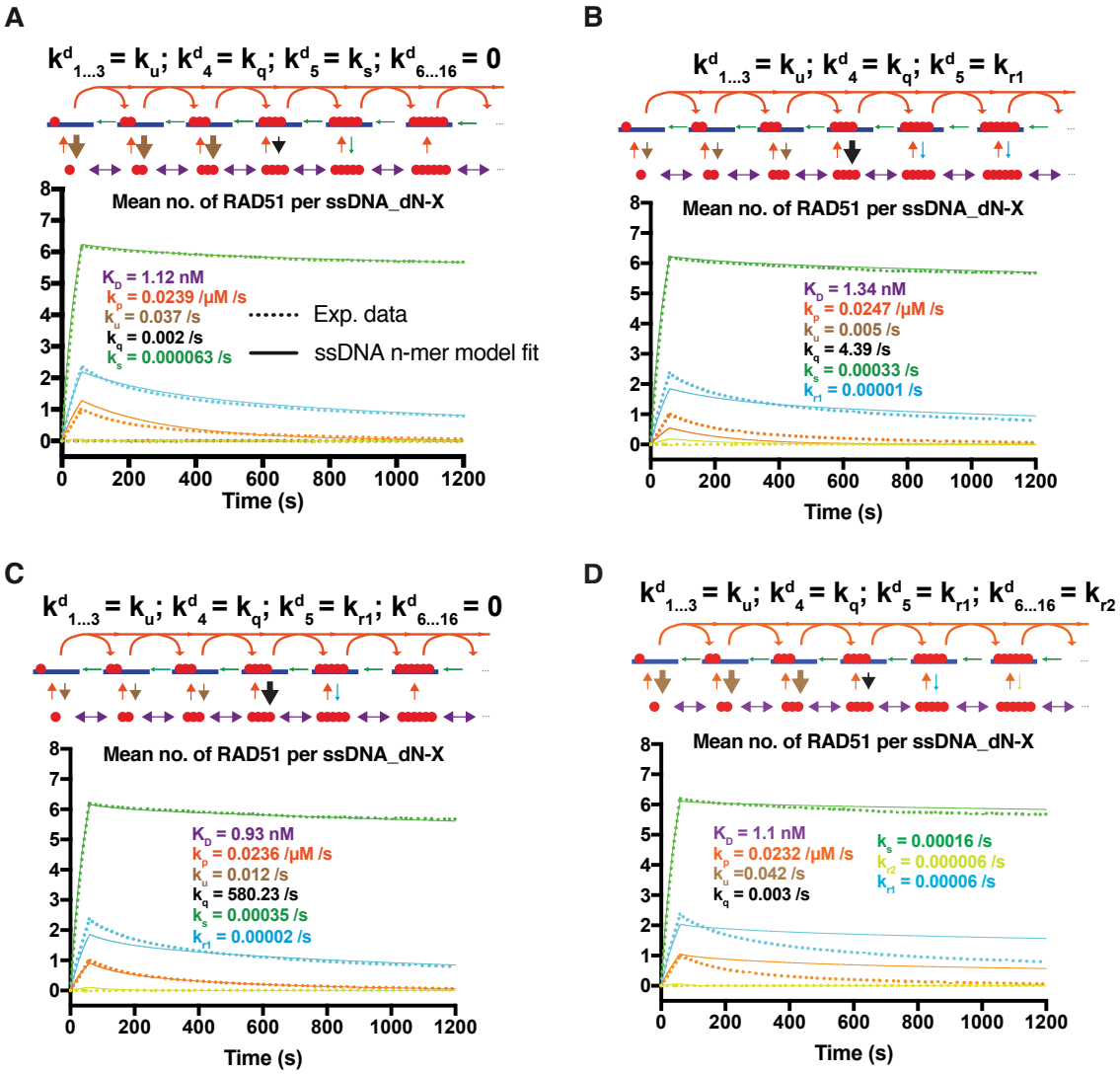


Figure B.5: ssDNA dN-X data fits for the ssDNA kinetic sub-models (KSMs) E, G, H, I. A kinetic representation of each KSM is shown above each plot, along with the associated simplifications compared to the general kinetic sub-model (Table 2.4) and the parameter fits. **A**. KSM E. **B**. KSM G. **C**. KSM H. **D**. KSM I.

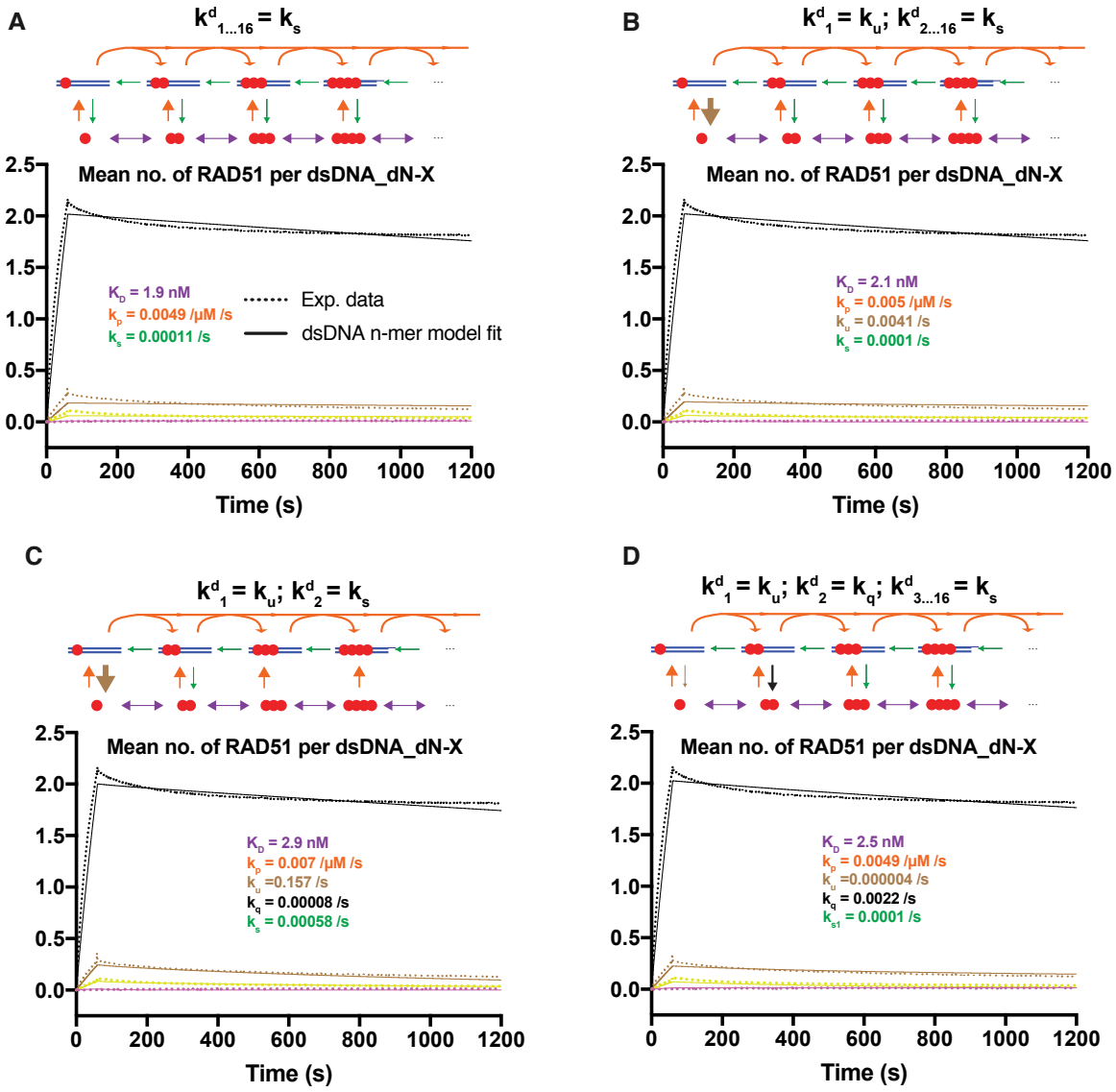


Figure B.6: dsDNA dN-Xp data fits for the dsDNA kinetic sub-models (KSMs) A, B, C, E. A kinetic representation of each KSM is shown above each plot, along with the associated simplifications compared to the general kinetic sub-model (Table 2.4) and the parameter fits. **A**. KSM A. **B**. KSM B. **C**. KSM C. **D**. KSM E.

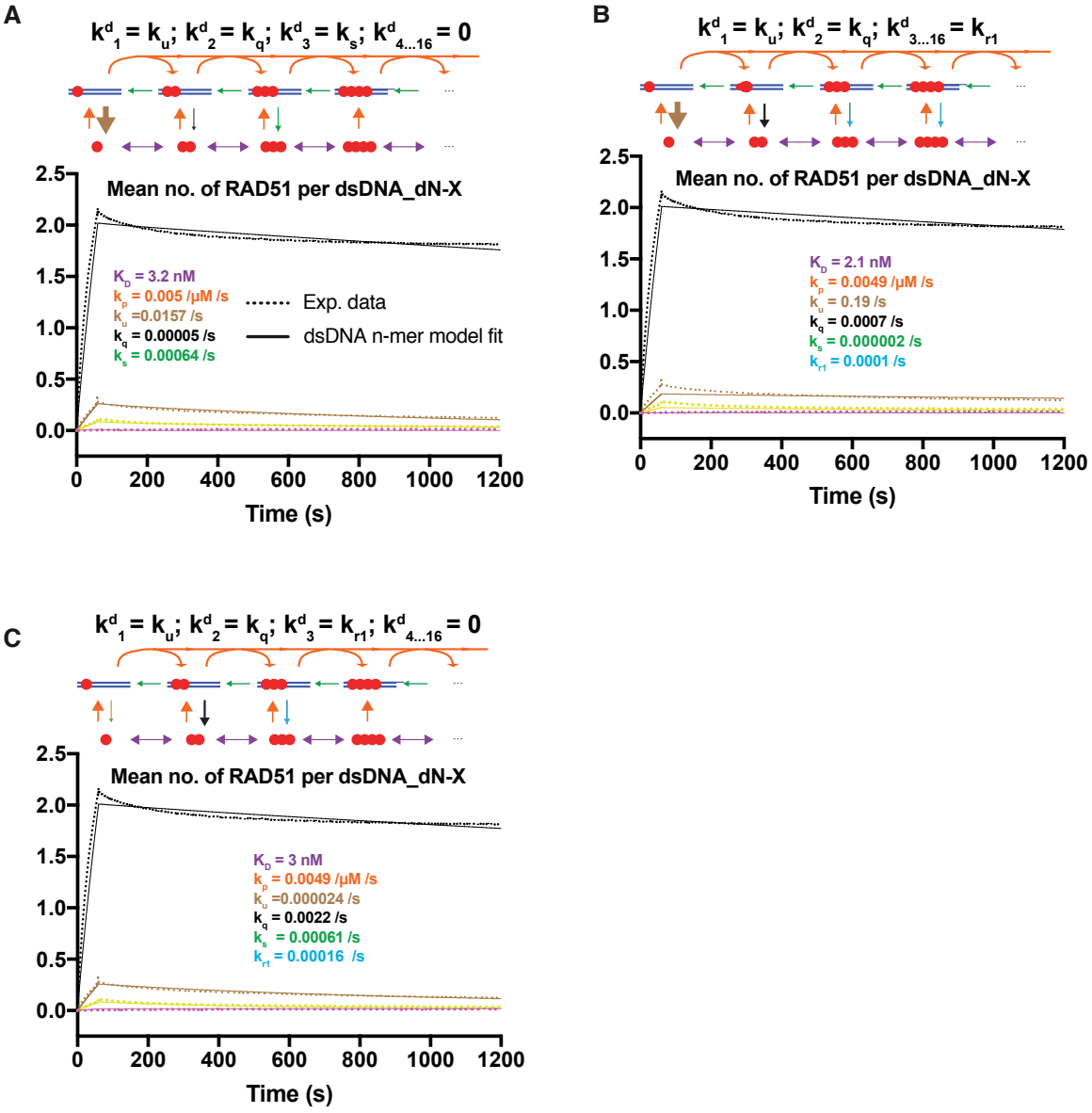


Figure B.7: dsDNA dN-Xp data fits for the dsDNA kinetic sub-models (KSMs) E - H. A kinetic representation of each KSM is shown above each plot, along with the associated simplifications compared to the general kinetic sub-model (Table 2.4) and the parameter fits. **A**. KSM E. **B**. KSM G. **C**. KSM H.

Bibliography

[1] S. P. Jackson and J. Bartek, “The DNA-damage response in human biology and disease,” *Nature*, vol. 461, no. 7267, pp. 1071–1078, 2010.

[2] R. Ceccaldi, B. Rondinelli, and A. D. D. Andrea, “Repair Pathway Choices and Consequences at the Double-Strand Break,” *Trends in Cell Biology*, vol. 26, no. 1, pp. 52–64, 2016.

[3] F. Esashi, V. E. Galkin, X. Yu, E. H. Egelman, and S. C. West, “Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2,” *Nature Structural and Molecular Biology*, vol. 14, no. 6, pp. 468–474, 2007.

[4] O. R. Davies and L. Pellegrini, “Interaction with the BRCA2 C-terminus Protects RAD51 – DNA Filaments from Disassembly by BRC Repeats,” *Nature Structural and Molecular Biology*, vol. 14, no. 6, pp. 475–483, 2007.

[5] R. B. Jensen, A. Carreira, and S. C. Kowalczykowski, “Purified human BRCA2 stimulates RAD51-mediated recombination,” *Nature*, vol. 467, no. 7316, pp. 678–683, 2010.

[6] A. Carreira and S. C. Kowalczykowski, “Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 26, 2011.

[7] W. Zhao, S. Vaithiyalingam, J. S. Filippo, W. J. Chazin, C. Wiese, and P. Sung, “Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry,” *Molecular Cell*, vol. 59, no. 2, pp. 176–187, 2015.

[8] R. Buisson, A.-M. Dion-Côté, Y. Coulombe, H. Launay, H. Cai, A. Z. Stasiak, A. Stasiak, B. Xia, and J.-Y. Masson, “Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination.,” *Nature structural & molecular biology*, vol. 17, no. 10, pp. 1247–54, 2010.

[9] T. Miyazaki, D. A. Bressan, M. Shinohara, J. E. Haber, and A. Shinohara, “In vivo assembly and disassembly of Rad51 and Rad52 complexes during double-strand break repair,” *EMBO Journal*, vol. 23, no. 4, pp. 939–949, 2004.

[10] C. Ma, Y. Kwon, P. Sung, and E. C. Greene, “Human RAD52 interactions with replication protein A and the RAD51 presynaptic complex,” *Journal of Biological Chemistry*, vol. 292, pp. 11702–11713, 2017.

[11] J. Miné, L. Disseau, M. Takahashi, G. Cappello, M. Dutreix, and J.-L. Viovy, “Real-time measurements of the nucleation, growth and dissociation of single Rad51 – DNA nucleoprotein filaments,” *Nucleic Acids Research*, vol. 35, no. 21, pp. 7171–7187, 2007.

[12] A. Candelli, J. T. Holthausen, M. Depken, I. Brouwer, M. A. M. Franker, M. Marchetti, I. Heller, S. Bernard, E. B. Garcin, M. Modesti, C. Wyman, G. J. L. Wuite, and E. J. G. Peterman, “Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 42, pp. 15090–5, 2014.

[13] A. Tubbs and A. Nussenzweig, “Endogenous DNA Damage as a Source of Genomic Instability in Cancer,” *Cell*, vol. 168, no. 4, pp. 644–656, 2017.

[14] M. Tarsounas and S. C. West, “Recombination at Mammalian Telomeres,” *Cell Cycle*, vol. 4, no. May, pp. 672–674, 2005.

[15] Y. Zhou, P. Caron, F. Toulouse, and T. T. Paull, “Quantitation of DNA double-strand break resection intermediates in human cells,” *Nucleic Acids Research*, vol. 42, no. 3, 2014.

[16] P. Huertas and S. P. Jackson, “Human CtIP Mediates Cell Cycle Control of DNA End Resection and Double Strand Break Repair,” *Journal of Biological Chemistry*, vol. 284, no. 14, pp. 9558–9565, 2009.

[17] J. Buis, T. Stoneham, E. Spehalski, and D. O. Ferguson, “Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2,” *Nature Structural and Molecular Biology*, vol. 19, no. 2, 2012.

[18] R. Anand, L. Ranjha, R. Anand, L. Ranjha, E. Cannavo, and P. Cejka, “Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Article Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection,” *Molecular Cell*, vol. 64, no. 5, pp. 940–950, 2016.

[19] L. Wohlbold, K. A. Merrick, S. De, R. Amat, J. H. Kim, J. J. Allen, C. Zhang, K. M. Shokat, J. H. J. Petrini, and R. P. Fisher, “Chemical Genetics Reveals a Specific Requirement for Cdk2 Activity in the DNA Damage Response and Identifies Nbs1 as a Cdk2 Substrate in Human Cells,” *PLoS Genetics*, vol. 8, no. 8, 2012.

[20] J. Falck, J. V. Forment, J. Coates, M. Mistrik, J. Lukas, J. Bartek, and S. P. Jackson, “CDK targeting of NBS1 promotes DNA-end resection, replication restart and homologous recombination,” *EMBO Reports*, vol. 13, no. 6, pp. 561–568, 2012.

[21] A. Sturzenegger, K. Burdova, R. Kanagaraj, M. Levikova, C. Pinto, P. Cejka, and P. Janscak, “DNA2 Cooperates with the WRN and BLM RecQ Helicases to Mediate Long-range DNA End Resection in Human Cells *,” *Journal of Biological Chemistry*, vol. 289, no. 39, pp. 27314–27326, 2014.

[22] C. Pinto, K. Kasaciunaite, R. Seidel, and P. Cejka, “Human DNA2 possesses a cryptic DNA unwinding activity that functionally integrates with BLM or WRN helicases,” *eLife*, vol. 5:e18574, pp. 1–24, 2016.

[23] A. W. Kijas, Y. C. Lim, E. Bolderson, K. Cerosaletti, M. Gatei, B. Jakob, F. Tobias, G. Taucher-scholz, N. Gueven, G. Oakley, P. Concannon, E. Wolvetang, K. K. Khanna, and L. Wiesm, “ATM-dependent phosphorylation of MRE11 controls extent of resection during homology directed repair by signalling through Exonuclease 1,” *Nucleic Acids Research*, vol. 43, no. 17, pp. 8352–8367, 2015.

[24] M. Steger, A. Tomas-loba, J. R. Chapman, P. Barral, J.-b. Vannier, A. A. Sartori, I. R. Adams, F. D. Batista, and S. J. Boulton, “Article RIF1 Is Essential for 53BP1-Dependent Nonhomologous End Joining and Suppression of DNA Double-Strand Break Resection,” *Molecular Cell*, vol. 49, pp. 858 – 871, 2013.

[25] H. Ghezraoui, C. Oliveira, J. R. Becker, K. Bilham, D. Moralli, C. Anzilotti, R. Fischer, M. Deobagkar-lele, M. Sanchiz-calvo, E. Fueyo-marcos, S. Bonham, B. M. Kessler, S. Rottenberg, R. J. Cornall, C. M. Green, and J. R. Chapman, “53BP1 cooperation with the REV7–shieldin complex underpins DNA structure-specific NHEJ,” *Nature*, vol. 560, pp. 122 – 127, 2018.

[26] H. Dev, T.-W. W. Chiang, C. Lescale, I. D. Krijger, A. G. Martin, D. Pilger, J. Coates, M. Sczaniecka-clift, W. Wei, M. Ostermaier, M. Herzog, J. Lam, A. Shea, M. Demir, Q. Wu, F. Yang, B. Fu, Z. Lai, G. Balmus, R. Belotserkovskaya, V. Serra, M. J. O. Connor, A. Bruna, P. Beli, L. Pellegrini, C. Caldas, L. Deriano, J. J. L. Jacobs, Y. Galanty, and S. P. Jackson, “Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null celss,” *Nature Cell Biology*, vol. 20, no. August, 2018.

[27] S. M. Noordermeer, S. Adam, D. Setiaputra, M. Barazas, S. J. Pettitt, A. K. Ling, M. Olivieri, A. Álvarez-quilón, N. Moatti, M. Zimmermann, S. Annunziato, D. B. Krastev, F. Song, I. Brandsma, J. Frankum, R. Brough, A. Sherker, S. Landry, R. K. Szilard, M. M. Munro, A. Mcewan, T. G. D. Rugy, Z.-y. Lin, T. Hart, J. Moffat, A.-c. Gingras, A. Martin, H. V. Attikum, J. Jonkers, C. J. Lord, S. Rottenberg, and D. Durocher, “The shieldin complex mediates 53BP1-dependent DNA repair,” *Nature*, vol. 560, pp. 117 – 121, 2018.

[28] S. Gao, S. Feng, S. Ning, J. Liu, H. Zhao, Y. Xu, J. Shang, K. Li, Q. Li, R. Guo, and D. Xu, “An OB-fold complex controls the repair pathways for DNA double-strand breaks,” *Nature Communications*, vol. 3925, pp. 1–10, 2018.

[29] S. F. Bunting, C. Elsa, N. Wong, H.-t. Chen, F. Polato, A. Gunn, A. Bothmer, S. F. Bunting, E. Calle, N. Feldhahn, O. Fernandez-capetillo, L. Cao, X. Xu, C.-x. Deng, T. Finkel, M. Nussenzweig, and J. M. Stark, “53BP1 Inhibits Homologous Recombination in Brca1 -Deficient Cells by Blocking Resection of DNA Breaks,” *Cell*, vol. 141, pp. 243–254, 2010.

[30] M. V. Botuyan, J. Lee, I. M. Ward, J.-e. Kim, R. James, J. Chen, and G. Mer, “Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair,” *Cell*, vol. 127, no. 7, pp. 1361–1373, 2006.

[31] A. Fradet-Turcotte, M. D. Canny, C. Escribano-Diaz, A. Orthwein, C. C. Y. Leung, H. Huang, S. M. Noordermeer, F. Sicheri, D. Durocher, M.-c. Landry, and J. Kitevski-leblanc, “53BP1 is a reader of the DNA-damage-induced H2A Lys15 ubiquitin mark,” *Nature*, vol. 499, pp. 50 – 56, 2013.

[32] C. Escribano-Diaz, A. Orthwein, A. Fradet-Turcotte, M. Xing, J. T. F. Young, A. Orthwein, C. Escribano-dı, M. A. Cook, A. P. Rosebrock, M. Munro, M. D. Canny, and D. Xu, “A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice,” *Molecular Cell*, vol. 49, pp. 872 – 883, 2013.

[33] L. Chen, C. J. Nievera, A. Y.-l. Lee, and X. Wu, “Cell Cycle-dependent Complex Formation of BRCA1-CtIP-MRN Is Important for DNA Double-strand Break Repair,” *Journal of Biological Chemistry*, vol. 283, no. 12, pp. 7713–7720, 2008.

[34] X. Yu and J. Chen, “DNA Damage-Induced Cell Cycle Checkpoint Control Requires CtIP , a Phosphorylation-Dependent Binding Partner of BRCA1 C-Terminal Domains,” *Molecular and Cellular Biology*, vol. 24, no. 21, pp. 9478–9486, 2004.

[35] Z. Shao, A. J. Davis, K. R. Fattah, S. So, J. Sun, K.-j. Lee, J. Yang, and D. J. Chen, “Persistently bound Ku at DNA ends attenuates DNA end resection and homologous recombination,” *DNA Repair*, vol. 11, no. 3, pp. 310–316, 2012.

[36] J. Sun, K.-j. Lee, A. J. Davis, and D. J. Chen, “Human Ku70 / 80 Protein Blocks Exonuclease 1-mediated DNA Resection in the Presence of Human Mre11 or Mre11 / Rad50 Protein Complex,” *Journal of Biological Chemistry*, vol. 287, no. 7, pp. 4936–4945, 2012.

[37] L. R. Myler, I. F. Gallardo, M. M. Soniat, Y. Kim, T. T. Paull, I. J. Finkelstein, L. R. Myler, I. F. Gallardo, M. M. Soniat, R. A. Deshpande, and X. B. Gonzalez, “Single-Molecule Imaging Reveals How Mre11- Short Article Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair,” *Molecular Cell*, vol. 67, no. 5, pp. 891–898.e4, 2017.

[38] C. J. Ma, B. Gibb, Y. Kwon, P. Sung, and E. C. Greene, “Protein dynamics of human RPA and RAD51 on ssDNA during assembly and disassembly of the RAD51 filament,” *Nucleic Acids Research*, vol. 45, no. 2, pp. 749–761, 2017.

[39] R. Chen, S. Subramanyam, A. H. Elcock, M. Spies, and M. S. Wold, “Dynamic binding of replication protein a is required for DNA repair,” *Nucleic Acids Research*, vol. 44, no. 12, pp. 5758–5772, 2016.

[40] J. Liu, T. Doty, B. Gibson, and W.-D. Heyer, “Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA,” *Nature Structural and Molecular Biology*, vol. 17, no. 10, pp. 1260–1262, 2010.

[41] T. Thorslund, M. J. Mcilwraith, S. A. Compton, S. Lekomtsev, M. Petronczki, J. D. Griffith, and S. C. West, “The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA,” *Nature Structural and Molecular Biology*, vol. 17, no. 10, pp. 1263–1265, 2010.

[42] A. Carreira, J. Hilario, I. Amitani, R. J. Baskin, M. K. K. Shivji, A. R. Venkitaraman, and S. C. Kowalczykowski, “The BRC Repeats of BRCA2 Modulate the DNA-Binding Selectivity of RAD51,” *Cell*, vol. 136, no. 6, pp. 1032–1043, 2009.

[43] P. Baumann and S. C. West, “The human Rad51 protein: polarity of strand transfer and stimulation by hRP-A,” *The EMBO Journal*, vol. 16, no. 17, pp. 5198–5206, 1997.

[44] T. Shahid, J. Soroka, E. H. Kong, L. Malivert, M. J. Mcilwraith, T. Pape, S. C. West, and X. Zhang, “Structure and mechanism of action of the BRCA2 breast cancer tumor suppressor,” *Nature*, vol. 21, no. 11, pp. 962–968, 2014.

[45] Y. Hashimoto, A. R. Chaudhuri, M. Lopes, and V. Costanzo, “Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis,” *Nature Structural & Molecular Biology*, vol. 17, no. 11, pp. 1305–1311, 2010.

[46] K. Zadorozhny, V. Sannino, O. Belan, J. Mlcouskova, M. Spirek, V. Costanzo, and L. Krejci, “Fanconi-Anemia-Associated Mutations Destabilize RAD51 Filaments and Impair Replication Fork Report Fanconi-Anemia-Associated Mutations Destabilize RAD51 Filaments and Impair Replication Fork Protection,” *Cell Reports*, vol. 21, pp. 333–340, 2017.

[47] M. Lomonosov, S. Anand, M. Sangrithi, R. Davies, and A. R. Venkitaraman, “Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein,” *Genes & Development*, vol. 17, pp. 3017–3022, 2003.

[48] K. Schlacher, N. Christ, N. Siaud, A. Egashira, H. Wu, and M. Jasin, “Double-Strand Break Repair-Independent Role for BRCA2 in Blocking Stalled Replication Fork Degradation by MRE11,” *Cell*, vol. 145, no. 4, pp. 529–542, 2011.

[49] R. Buisson and J.-Y. Masson, “PALB2 self-interaction controls homologous recombination,” *Nucleic Acids Research*, vol. 40, no. 20, pp. 10312–10323, 2012.

[50] K. A. Miller, D. M. Yoshikawa, I. R. Mcconnell, R. Clark, D. Schild, and J. S. Albala, “RAD51C Interacts with RAD51B and Is Central to a Larger Protein Complex in Vivo Exclusive of RAD51,” *Journal of Biological Chemistry*, vol. 277, no. 10, pp. 8406–8411, 2002.

[51] C. Wiese, D. W. Collins, J. S. Albala, L. H. Thompson, A. Kronenberg, and D. Schild, “Interactions involving the Rad51 paralogs Rad51C and XRCC3 in human cells,” *Nucleic Acids Research*, vol. 30, no. 4, pp. 1001–1008, 2002.

[52] T. Liu, L. Wan, Y. Wu, J. Chen, and J. Huang, “hSWS1-SWSAP1 Is an Evolutionarily Conserved Complex Required for Efficient Homologous Recombination Repair,” *Journal of Biological Chemistry*, vol. 286, no. 48, pp. 41758–41766, 2011.

[53] M. R. G. Taylor, K. R. Chaurasiya, D. Rueda, L. Krejci, S. J. Boulton, K. R. Chaurasiya, J. D. Ward, R. Carzaniga, X. Yu, E. H. Egelman, L. M. Collinson, D. Rueda, L. Krejci, and S. J. Boulton, “Rad51 Paralogs Remodel Presynaptic Rad51 Filaments to Stimulate Homologous Recombination,” *Cell*, pp. 271–286, 2015.

[54] M. R. G. Taylor, C. J. Ma, E. C. Greene, L. Krejci, S. J. Boulton, C. J. Ma, R. Carzaniga, T. Takaki, and L. M. Collinson, “A Polar and Nucleotide-Dependent Mechanism of Action for RAD51 Paralogs in RAD51 Filament Remodeling,” *Molecular Cell*, vol. 64, pp. 926–939, 2016.

[55] T. B. McClendon, M. R. Sullivan, K. A. Bernstein, and J. L. Yanowitz, “Promotion of Homologous Recombination by SWS-1 in Complex with RAD-51 Paralogs in,” *Genetics*, vol. 203, no. May, pp. 133–145, 2016.

[56] P. Baumann, F. E. Benson, and S. C. West, “Human Rad51 Protein Promotes ATP-Dependent Homologous Pairing and Strand Transfer Reactions In Vitro,” *Cell*, vol. 87, pp. 757–766, 1996.

[57] S. Schwendener, S. Raynard, S. Paliwal, A. Cheng, R. Kanagaraj, I. Shevelev, J. M. Stark, P. Sung, and P. Janscak, “Physical Interaction of RECQ5 Helicase with RAD51 Facilitates Its Anti-recombinase Activity,” *Journal of Biological Chemistry*, vol. 285, no. 21, pp. 15739–15745, 2010.

[58] J. Simandlova, J. Zagelbaum, M. J. Payne, W. Kit, I. Shevelev, K. Hanada, S. Chatterjee, D. A. Reid, Y. Liu, P. Janscak, E. Rothenberg, and I. D. Hickson, “FBH1 Helicase Disrupts RAD51 Filaments in Vitro and Modulates Homologous Recombination in Mammalian Cells *,” *Journal of Biological Chemistry*, vol. 288, no. 47, pp. 34168–34180, 2013.

[59] K. Ito, Y. Murayama, M. Takahashi, and H. Iwasaki, “Two three-strand intermediates are processed during Rad51-driven DNA strand exchange,” *Nature Structural & Molecular Biology*, vol. 25, no. January, 2018.

[60] J. Xu, L. Zhao, Y. Xu, W. Zhao, P. Sung, and H.-w. Wang, “Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange,” *Nature Structural and Molecular Biology*, vol. 24, no. 1, pp. 40–46, 2016.

[61] W. Zhao, J. B. Steinfeld, F. Liang, X. Chen, D. G. Maranon, C. J. Ma, Y. Kwon, T. Rao, W. Wang, C. Sheng, X. Song, Y. Deng, J. Jimenez-sainz, L. Lu, R. B. Jensen, Y. Xiong, G. M. Kupfer, C. Wiese, E. C. Greene, and P. Sung, “BRCA1–BARD1 promotes RAD51-mediated homologous DNA pairing,” *Nature*, vol. 550, no. 7676, pp. 360–365, 2017.

[62] L. J. Barber, J. L. Youds, J. D. Ward, M. J. Mcilwraith, N. J. O. Neil, M. I. R. Petalcorin, J. S. Martin, S. J. Collis, S. B. Cantor, M. Auclair, H. Tissenbaum, S. C. West, A. M. Rose, and S. J. Boulton, "RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination," *Cell*, vol. 135, no. 2, pp. 261–271, 2008.

[63] O. M. Mazina and A. V. Mazin, "Human Rad54 Protein Stimulates DNA Strand Exchange Activity of hRad51 Protein in the Presence of Ca2+," *Journal of Biological Chemistry*, vol. 279, no. 50, pp. 52042–52051, 2004.

[64] J. Spies, A. Waizenegger, O. Barton, M. Su, W. D. Wright, W.-d. Heyer, M. Lo, J. Spies, A. Waizenegger, O. Barton, M. Su, W. D. Wright, and W.-d. Heyer, "Nek1 Regulates Rad54 to Orchestrate Homologous Recombination and Replication Fork Stability," *Molecular Cell*, vol. 62, pp. 903–917, 2016.

[65] X. Li, C. M. Stith, P. M. Burgers, and W.-D. Heyer, "PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase delta," *Molecular Cell*, vol. 36, no. 4, pp. 704–713, 2010.

[66] X. Wang, G. Ira, A. Tercero, A. M. Holmes, J. F. X. Diffley, and J. E. Haber, "Role of DNA Replication Proteins in Double-Strand Break-Induced Recombination in Saccharomyces cerevisiae," *Molecular and Cellular Biology*, vol. 24, no. 16, pp. 6891–6899, 2004.

[67] X. J. Chen, "Mechanism of Homologous Recombination and Implications for Aging-Related Deletions in Mitochondrial DNA," *Microbiology and Molecular Biology Reviews*, vol. 77, no. 3, pp. 476–496, 2013.

[68] J. Li, D. L. Holzschu, and T. Sugiyama, "PCNA is efficiently loaded on the DNA recombination intermediate to modulate polymerase δ , η , and ζ activities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 7672 – 7677, 2013.

[69] M. Seki, F. Marini, and R. D. Wood, "POLQ (Pol q), a DNA polymerase and DNA-dependent ATPase in human cells," *Nucleic Acids Research*, vol. 31, no. 21, 2003.

[70] P. A. Mateos-Gomez, F. Gong, N. Nair, K. M. Miller, E. Lazzerini-denchi, and A. Sfeir, "Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination," *Nature*, vol. 518, no. 7538, pp. 254–257, 2014.

[71] G.-l. Moldovan, M. V. Madhavan, K. D. Mirchandani, R. M. Mccaffrey, P. Vinciguerra, and A. D. D. Andrea, "DNA Polymerase POLN Participates in Cross-Link Repair and Homologous Recombination," *Molecular and Cellular Biology*, vol. 30, no. 4, pp. 1088–1096, 2010.

[72] C. Biertumpfel, Y. Zhao, Y. Kondo, S. Ramon-Maiques, M. Gregory, J. Y. Lee, Y. Zhao, Y. Kondo, and S. Ramo, "Structure and mechanism of human DNA polymerase η," *Nature*, vol. 465, no. June, pp. 1–5, 2010.

[73] J. A. Newman, D. O. Christopher, H. Aitkenhead, J. A. Newman, C. D. O. Cooper, H. Aitkenhead, and O. Gileadi, “Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway,” *Structure*, vol. 23, no. 12, pp. 2319–2330, 2015.

[74] M. J. Mcilwraith, A. Vaisman, Y. Liu, E. Fanning, R. Woodgate, and S. C. West, “Human DNA Polymerase η Promotes DNA Synthesis from Strand Invasion Intermediates of Homologous Recombination,” *Molecular Cell*, vol. 20, pp. 783–792, 2005.

[75] J. D.-m. Goetz, T. A. Motycka, M. Han, M. Jasin, and A. E. Tomkinson, “Reduced repair of DNA double-strand breaks by homologous recombination in a DNA ligase I-deficient human cell line,” *DNA Repair*, vol. 4, pp. 649–654, 2005.

[76] L. S. Symington and L. S. Symington, “Mechanism and regulation of DNA end resection in eukaryotes Mechanism and regulation of DNA end resection in eukaryotes,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 9238, 2016.

[77] A. Xie, A. Kwok, and R. Scully, “Role of mammalian Mre11 in classical and alternative nonhomologous end joining,” *Nature Structural and Molecular Biology*, vol. 16, no. 8, pp. 814–818, 2009.

[78] K. K. Chiruvella, Z. Liang, and T. E. Wilson, “Repair of Double-Strand Breaks by End Joining,” *Cold Spring Harb Perspect Biol*, vol. 5:a012757, 2013.

[79] S. Jin and D. T. Weaver, “Double-strand break repair by Ku70 requires heterodimerization with Ku80 and DNA binding functions,” *EMBO Journal*, vol. 16, no. 22, pp. 6874–6885, 1997.

[80] P.-o. Mari, B. I. Florea, S. P. Persengiev, N. S. Verkaik, H. T. Bru, M. Modesti, G. Giglia-mari, K. Bezstarosti, J. A. A. Demmers, T. M. Luider, A. B. Houtsmuller, and D. C. V. Gent, “Dynamic assembly of end-joining complexes requires interaction between Ku70 / 80 and XRCC4,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 49, pp. 18597–18602, 2006.

[81] P. Sung and H. Klein, “Mechanism of homologous recombination : mediators and helicases take on regulatory functions,” *Nature Molecular Cell Biology*, vol. 7, pp. 739–750, 2006.

[82] M. C. Muñoz, C. Laulier, A. Gunn, A. Cheng, D. F. Robbiani, A. Nussenzweig, and J. M. Stark, “Ring Finger Nuclear Factor RNF168 Is Important for Defects in Homologous Recombination Caused by Loss of the Breast Cancer Susceptibility Factor BRCA1,” *Journal of Biological Chemistry*, vol. 287, no. 48, pp. 40618–40628, 2012.

[83] R. Bhargava, D. O. Onyango, and J. M. Stark, “Regulation of Single-Strand Annealing and its Role in Genome Maintenance,” *Trends in Genetics*, vol. 32, no. 9, pp. 566–575, 2016.

[84] T. A. Motycka, T. Bessho, S. M. Post, P. Sung, and A. E. Tomkinson, “Physical and Functional Interaction between the XPF / ERCC1 Endonuclease and hRad52,” *Journal of Biological Chemistry*, vol. 279, no. 14, pp. 13634–13639, 2004.

[85] J.-I. Ma, E. M. Kim, J. E. Haber, and S. E. Lee, “Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences,” *Molecular and Cellular Biology*, vol. 23, no. 23, pp. 8820–8828, 2003.

[86] D. Simsek and M. Jasin, “Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4 – ligase IV during chromosomal translocation formation,” *Nature Structural and Molecular Biology*, vol. 17, no. 4, pp. 410–416, 2010.

[87] Y. Zhang and M. Jasin, “An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway,” *Nature Structural and Molecular Biology*, vol. 18, no. 1, pp. 80–84, 2010.

[88] M. Wang, W. Wu, W. Wu, B. Rosidi, L. Zhang, H. Wang, and G. Iliakis, “PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways,” *Nucleic Acids Research*, vol. 34, no. 21, pp. 6170–6182, 2006.

[89] A. H. Bizard and I. D. Hickson, “The Dissolution of Double Holliday Junctions,” *Cold Spring Harb Perspect Biol*, pp. 1–14, 2014.

[90] A. B. Conway, T. W. Lynch, Y. Zhang, G. S. Fortin, C. W. Fung, L. S. Symington, and P. a. Rice, “Crystal structure of a Rad51 filament,” *Nature Structural and Molecular Biology*, vol. 11, no. 8, pp. 791–796, 2004.

[91] D. S. Shin, L. Pellegrini, D. S. Daniels, B. Yelent, L. Craig, D. Bates, D. S. Yu, M. K. Shivji, C. Hitomi, A. S. Arvai, N. Volkmann, H. Tsuruta, and T. L. Blundell, “Full-length archaeal Rad51 structure and mutants : mechanisms for RAD51 assembly and control by BRCA2,” *EMBO Journal*, vol. 22, no. 17, 2003.

[92] D. S. Yu, E. Sonoda, S. Takeda, C. L. Huang, L. Pellegrini, T. L. Blundell, and A. R. Venkitaraman, “Dynamic control of Rad51 recombinase by self-association and interaction with BRCA2,” *Molecular Cell*, vol. 12, no. 4, pp. 1029–1041, 2003.

[93] F. E. Benson, A. Stasiak, and S. C. West, “Purification and characterization of the human Rad51 protein, an analogue of E. coli RecA.,” *The EMBO journal*, vol. 13, no. 23, pp. 5764–71, 1994.

[94] D. Ristic, M. Modesti, T. V. D. Heijden, and J. V. Noort, “Human Rad51 filaments on double- and single-stranded DNA : correlating regular and irregular forms with recombination function,” *Nucleic Acids Research*, vol. 33, no. 10, pp. 3292–3302, 2005.

[95] P. Chi, S. V. Komen, M. G. Sehorn, S. Sigurdsson, and P. Sung, “Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function,” *DNA Repair*, vol. 5, pp. 381–391, 2006.

[96] P. Sung, “Catalysis of ATP-Dependent Homologous DNA Pairing and Strand Exchange by Yeast RAD51 Protein,” *Science*, vol. 265, no. August, pp. 2–5, 1994.

[97] P. Sung and D. L. Robberson, “DNA Strand Exchange Mediated by a RAD51-ssDNA Nucleoprotein Filament with Polarity Opposite to That of RecA,” *Cell*, vol. 82, pp. 453–461, 1995.

[98] S. Dyn, J. P. Menetski, A. Varghese, and S. C. Kowalczykowski, “Properties of the High-Affinity Single-Stranded DNA Binding State of the Escherichia coli RecA Protein,” *Biochemistry*, vol. 27, no. 1979, pp. 1205–1212, 1988.

[99] N. E. Freitags and K. Mcenteeq, “Site-directed Mutagenesis of the RecA Protein of Escherichia coli,” *Journal of Biological Chemistry*, vol. 266, pp. 7058–7066, 1991.

[100] Z. Zhang, D. Yoon, J. R. Laporte, and J. Chen, “Appropriate Initiation of the Strand Exchange Reaction Promoted by RecA Protein Requires ATP Hydrolysis,” *J. Mol. Biol.*, vol. 309, pp. 29 – 43, 2001.

[101] G. Tombline and R. Fishel, “Biochemical Characterization of the Human RAD51 Protein,” *Journal of Biological Chemistry*, vol. 277, no. 17, pp. 14417–14425, 2002.

[102] S. L. Brennersq, R. S. Mitchell, S. W. Morricaltii, S. K. Neuendorfll, B. C. Schutten, and M. M. Coxll, “recA Protein-promoted ATP Hydrolysis Occurs throughout recA Nucleoprotein Filaments,” *Journal of Biological Chemistry*, vol. 262, no. 9, pp. 4011–4016, 1987.

[103] J. M. Mason, K. Dusad, W. D. Wright, J. Grubb, B. Budke, W.-d. Heyer, P. P. Connell, R. R. Weichselbaum, and D. K. Bishop, “RAD54 family translocases counter genotoxic effects of RAD51 in human tumor cells,” *Nucleic Acids Research*, vol. 43, no. 6, pp. 3180–3196, 2015.

[104] J. Y. Lee, Z. Qi, and E. C. Greene, “ATP hydrolysis Promotes Duplex DNA Release by the RecA Presynaptic Complex,” *Journal of Biological Chemistry*, vol. 291, no. 42, pp. 22218–22230, 2016.

[105] S. Ragone, J. D. Maman, N. Furnham, and L. Pellegrini, “Structural basis for inhibition of homologous recombination by the RecX protein,” *The EMBO Journal*, vol. 27, pp. 2259 – 2269, 2008.

[106] D. V. Bugreev and A. V. Mazin, “Ca2+ activates human homologous recombination protein Rad51 by modulating its ATPase activity,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 27, 2004.

[107] K.-s. Shim, G. Tombline, C. D. Heinen, N. Charbonneau, C. Schmutte, and R. Fishel, “Magnesium influences the discrimination and release of ADP by human RAD51,” *DNA Repair*, vol. 5, pp. 704–717, 2006.

[108] L. H. Fornander, K. Frykholm, A. Reymer, A. Renodon-cornie, M. Takahashi, and B. Norde, “Ca 2 + improves organization of single-stranded DNA bases in human Rad51 filament , explaining stimulatory effect on gene recombination,” *Nucleic Acids Research*, vol. 40, no. 11, pp. 4904–4913, 2012.

[109] T. Van der Heijden, R. Seidel, M. Modesti, R. Kanaar, C. Wyman, and C. Dekker, “Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules,” *Nucleic Acids Research*, vol. 35, no. 17, pp. 5646–5657, 2007.

[110] J. Hilario, I. Amitani, R. J. Baskin, and S. C. Kowalczykowski, “Direct imaging of human Rad51 nucleoprotein dynamics on individual DNA molecules,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 2, pp. 361–368, 2009.

[111] Y. Liu, A. Z. Stasiak, J.-y. Masson, M. J. Mcilwraith, A. Stasiak, and S. C. West, “Conformational Changes Modulate the Activity of Human RAD51 Protein,” *J. Mol. Biol.*, vol. 337, pp. 817–827, 2004.

[112] X. Yu, S. A. Jacobs, S. C. West, T. Ogawa, and E. H. Egelman, “Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8419–8424, 2001.

[113] J. M. Short, Y. Liu, S. Chen, N. Soni, M. S. Madhusudhan, M. K. K. Shivji, and A. R. Venkitaraman, “High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy,” *Nucleic Acids Research*, vol. 44, no. 19, pp. 9017–9030, 2016.

[114] T. Ogawa, X. Yu, A. Shinohara, and E. H. Egelman, “Similarity of the Yeast RAD51 Filament to the Bacterial RecA Filament,” *Science*, vol. 259, no. April, 1993.

[115] R. M. Story, I. T. Weber, and T. A. Steitz, “The structure of the E. coli recA protein monomer and polymer,” *Nature*, vol. 355, pp. 318 – 324, 1992.

[116] E. A. Hewat, R. W. H. Ruigrok, and E. Dicapua, “Activation of recA protein : the pitch of the helical complex with single-stranded DNA,” *EMBO Journal*, vol. 10, no. 9, pp. 2695–2698, 1991.

[117] Z. Chen, H. Yang, and N. P. Pavletich, “Mechanism of homologous recombination from the RecA – ssDNA / dsDNA structures,” *Nature*, vol. 453, no. May, 2008.

[118] M. S. Vanloock, X. Yu, S. Yang, A. L. Lai, C. Low, M. J. Campbell, E. H. Egelman, and F. City, “ATP-Mediated Conformational Changes in the RecA Filament University of Virginia Health Sciences,” *Structure*, vol. 11, no. 03, pp. 187–196, 2003.

[119] Y. Wu, Y. He, I. A. Moya, X. Qian, and Y. Luo, “Crystal Structure of Archaeal Recombinase RadA : A Snapshot of Its Extended Conformation,” *Molecular Cell*, vol. 15, pp. 423–435, 2004.

[120] I. Brouwer, T. Moschetti, A. Candelli, E. B. Garcin, M. Modesti, L. Pellegrini, G. J. L. Wuite, and E. J. G. Peterman, “Two distinct conformational states define the interaction of human RAD51-ATP with single-stranded DNA,” *EMBO Journal*, pp. 1–13, 2018.

[121] R. B. Robertson, D. N. Moses, Y. Kwon, P. Chan, P. Chi, H. Klein, P. Sung, and E. C. Greene, “Structural transitions within human Rad51 nucleoprotein filaments,” *Proceedings of the National Academy of Sciences*, vol. 106, pp. 12688 – 12693, 2009.

[122] T. Nishinaka, Y. Doi, R. Hara, and E. Yashima, “Elastic Behavior of RecA-DNA Helical Filaments,” *J. Mol. Biol.*, vol. 370, pp. 837–845, 2007.

[123] A. Renodon-Cornière, Y. Takizawa, S. Conilleau, V. Tran, S. Iwai, H. Kurumizaka, and M. Takahashi, “Structural Analysis of the Human Rad51 Protein – DNA Complex Filament by Tryptophan Fluorescence Scanning Analysis : Transmission of Allosteric Effects between ATP Binding and DNA Binding,” *J. Mol. Biol.*, vol. 383, pp. 575–587, 2008.

[124] R. Amunugama and R. Fishel, “Subunit Interface Residues F129 and H294 of Human RAD51 Are Essential for Recombinase Function,” *PLoS ONE*, vol. 6, no. 8, pp. 1–8, 2011.

[125] D. Rhodes and A. Klug, “Helical periodicity of DNA determined by enzyme digestion,” *Nature*, vol. 286, pp. 573 – 578, 1980.

[126] S. Atwell, L. Disseau, A. Z. Stasiak, A. Stasiak, A. Renodon-cornie, and M. Curie, “Probing Rad51-DNA interactions by changing DNA twist,” *Nucleic Acids Research*, vol. 40, no. 22, pp. 11769–11776, 2012.

[127] V. E. Galkin, F. Esashi, X. Yu, S. Yang, S. C. West, and E. H. Egelman, “BRCA2 BRC motifs bind RAD51 – DNA filaments,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 24, pp. 8537–8542, 2005.

[128] J. T. Holthausen, M. T. J. V. Loenhout, H. Sanchez, and D. Ristic, “Effect of the BRCA2 CTRD domain on RAD51 filaments analyzed by an ensemble of single molecule techniques,” *Nucleic Acids Research*, vol. 39, no. 15, pp. 6558–6567, 2011.

[129] L. Pellegrini, D. S. Yu, T. Lo, S. Anand, M. Lee, T. L. Blundell, and A. R. Venkitaraman, “Insights into DNA recombination from the structure of a RAD51 – BRCA2 complex,” *Nature*, vol. 420, pp. 287–293, 2002.

[130] T. Shinohara, S. Ikawa, W. Iwasaki, T. Hiraki, T. Hikima, T. Mikawa, N. Arai, N. Kamiya, and T. Shibata, “Loop L1 governs the DNA-binding specificity and order for RecA-catalyzed reactions in homologous recombination and DNA repair,” *Nucleic Acids Research*, vol. 43, no. 2, pp. 973–986, 2015.

[131] G. Tombline, C. D. Heinen, K.-s. Shim, and R. Fishel, “Biochemical Characterization of the Human RAD51 Protein. III. Modulation of DNA Binding by Adenosine Nucleotides.,” *Journal of Biological Chemistry*, vol. 277, no. 17, pp. 14434–14442, 2002.

[132] S. Sigurdsson, K. Trujillo, B. Song, S. Stratton, and P. Sung, “Basis for Avid Homologous DNA Strand Exchange by Human Rad51 and RPA,” *Journal of Biological Chemistry*, vol. 276, no. 12, pp. 8798–8806, 2001.

[133] K.-s. Shim, C. Schmutte, K. Yoder, and R. Fishel, “Defining the salt effect on human RAD51 activities,” *DNA Repair*, vol. 5, pp. 718–730, 2006.

[134] Y. Wu, X. Qian, Y. He, I. A. Moya, and Y. Luo, “Crystal Structure of an ATPase-active Form of Rad51 Homolog from Methanococcus voltae,” *Journal of Biological Chemistry*, vol. 280, no. 1, pp. 722–728, 2005.

[135] R. Amunugama, Y. He, S. Willcox, R. A. Forties, K.-s. Shim, R. Bundschuh, Y. Luo, J. Griffith, and R. Fishel, “RAD51 Protein ATP Cap Regulates Nucleoprotein Filament Stability,” *Journal of Biological Chemistry*, vol. 287, no. 12, pp. 8724–8736, 2012.

[136] J. V. Mameren, M. Modesti, R. Kanaar, C. Wyman, G. J. L. Wuite, and E. J. G. Peterman, “Dissecting Elastic Heterogeneity along DNA Molecules Coated Partly with Rad51 Using Concurrent Fluorescence Microscopy and Optical Tweezers,” *Biophysical Journal*, vol. 91, no. 8, pp. L78–L80, 2006.

[137] M. Modesti, D. Ristic, T. V. D. Heijden, C. Dekker, and J. V. Mameren, “Fluorescent Human RAD51 Reveals Multiple Nucleation Sites and Filament Segments Tightly Associated along a Single DNA Molecule,” *Structure*, vol. 15, no. May, pp. 599–609, 2007.

[138] J. V. Mameren, M. Modesti, R. Kanaar, C. Wyman, E. J. G. Peterman, and G. J. L. Wuite, “Counting RAD51 proteins disassembling from nucleoprotein filaments under tension,” *Nature*, vol. 457, no. 7230, pp. 745–748, 2009.

[139] J. M. Cox, O. V. Tsodikov, and M. M. Cox, “Organized Unidirectional Waves of ATP Hydrolysis within a RecA Filament,” *PLoS Biology*, vol. 3, no. 2, 2005.

[140] J. K. D. Zutter and K. L. Knight, “The hRad51 and RecA Proteins Show Significant Differences in Cooperative Binding to Single- stranded DNA,” *J. Mol. Biol.*, vol. 52, pp. 769 – 780, 1999.

[141] J. D. Mcghee and P. H. V. O. N. Hippel, “Theoretical Aspects of DNA-Protein Interactions : Co-operative and Non-co-operative Binding of Large Ligands to a One-dimensional Homogeneous Lattice,” *J. Mol. Biol.*, vol. 86, pp. 469–489, 1974.

[142] S. C. Kowalczykowski, S. L. Paul, N. Lonberg, J. W. Newport, J. A. McSwiggen, and P. H. V. Hippel, “Cooperative and Noncooperative Binding of Protein Ligands to Nucleic Acid Lattices : Experimental Approaches to the Determination of Thermodynamic Parameters,” *Biochemistry*, vol. 25, no. 1986, pp. 1226–1240, 1986.

[143] B. A. A. Travers, “The structural basis of DNA flexibility,” *Phil. Trans. R. Soc. Lond.*, vol. 362, pp. 1423–1438, 2004.

[144] H. R. Drew, J. R. Weeks, and A. A. Travers, “Negative supercoiling induces spontaneous unwinding of a bacterial promoter,” *The EMBO journal*, vol. 4, no. 4, pp. 1025–1032, 1985.

[145] T. J. Richmond and C. A. Davey, “The structure of DNA in the nucleosome core,” *Nature*, vol. 423, pp. 145–150, 2003.

[146] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha, “Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy,” *Biophysical Journal*, vol. 86, no. 4, pp. 2530–2537, 2004.

[147] Q. Chi, G. Wang, and J. Jiang, “The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory,” *Physica A*, vol. 392, no. 5, pp. 1072–1079, 2013.

[148] J. Kang, J. Jung, and S. Keun, “Flexibility of single-stranded DNA measured by single-molecule FRET,” *Biophysical Chemistry*, vol. 195, pp. 49–52, 2014.

[149] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, “Ionic effects on the elasticity of single DNA molecules,” *Proceedings of the National Academy of Sciences*, vol. 94, no. 12, pp. 6185–6190, 1997.

[150] A. Brunet, C. Tardin, L. Salome, P. Rousseau, N. Destainville, and M. Manghi, “Dependence of DNA Persistence Length on Ionic Strength of Solutions with Monovalent and Divalent Salts : A Joint Theory − Experiment Study,” *Macromolecules*, vol. 48, pp. 3641 – 3652, 2015.

[151] L. Cruzeiro-Hansson and J. M. Goodfellow, “Flexibility and curvature of DNA duplexes containing O4-methylthymine : implications for DNA repair,” *Carcinogenesis*, vol. 15, no. 8, pp. 1525–1533, 1994.

[152] E. Seibert, J. B. A. Ross, and R. Osman, “Role of DNA Flexibility in Sequence-Dependent Activity of Uracil DNA,” *Biochemistry*, vol. 41, pp. 10976–10984, 2002.

[153] C. Cao, Y. L. Jiang, J. T. Stivers, and F. Song, “Dynamic opening of DNA during the enzymatic search for a damaged base,” *Nature Structural & Molecular Biology*, vol. 11, no. 12, pp. 1230–1236, 2004.

[154] M. T. Hess, U. Scwitter, M. Petretta, B. Giese, and H. Naegeli, “Bipartite substrate discrimination by human nucleotide,” *Proceedings of the National Academy of Sciences*, vol. 94, no. June, pp. 6664–6669, 1997.

[155] V. Mocquet, K. Kropachev, M. Kolbanovskiy, A. Tapias, Y. Cai, S. Broyde, N. E. Geacintov, and J.-m. Egly, “The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo [a] pyrenyl- DNA lesions,” *EMBO Journal*, vol. 26, no. 12, pp. 2923–2932, 2007.

[156] J.-e. Yeo, A. Khoo, and A. F. Fagbemi, “The Efficiencies of Damage Recognition and Excision Correlate with Duplex Destabilization Induced by Acetylamino fl uorene Adducts in Human Nucleotide Excision Repair,” *Chem. Res. Toxicol.*, vol. 25, pp. 2462 – 2468, 2012.

[157] H. Mu, K. Kropachev, L. Wang, L. Zhang, A. Kolbanovskiy, M. Kolbanovskiy, N. E. Geacintov, and S. Broyde, “Nucleotide excision repair of 2-acetylaminofluorene- molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies,” *Nucleic Acids Research*, vol. 40, no. 19, pp. 9675–9690, 2012.

[158] J.-H. Min and N. P. Pavletich, “Recognition of DNA damage by the Rad4 nucleotide excision repair protein,” *Nature*, vol. 449, pp. 570 – 576, 2007.

[159] M. Štros, “HMGB proteins : Interactions with DNA and chromatin,” *Biochimica et Biophysica Acta*, vol. 1799, pp. 101–113, 2010.

[160] E. D. Ross, P. R. Hardwidge, and L. J. M. Iii, “HMG Proteins and DNA Flexibility in Transcription Activation,” *Molecular and Cellular Biology*, vol. 21, no. 19, pp. 6598–6605, 2001.

[161] J. Zhang, M. J. Mccauley, L. J. M. Iii, M. C. Williams, and N. E. Israeloff, “Mechanism of DNA flexibility enhancement by HMGB proteins,” *Nucleic Acids Research*, vol. 37, no. 4, pp. 1107–1114, 2009.

[162] S. S. Lange, D. L. Mitchell, and K. M. Vasquez, “High mobility group protein B1 enhances DNA repair and chromatin modification after DNA damage,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 30, 2008.

[163] M. J. McCauley, E. M. Rueter, I. Rouzina, L. J. M. Iii, and M. C. Williams, “Single-molecule kinetics reveal microscopic mechanism by which High-Mobility Group B proteins alter DNA flexibility,” *Nucleic Acids Research*, vol. 41, no. 1, pp. 167–181, 2013.

[164] D. Murugesapillai, M. J. Mccauley, R. Huo, M. H. N. Holte, A. Stepanyants, L. J. Maher, N. E. Israeloff, and M. C. Williams, “DNA bridging and looping by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin,” *Nucleic Acids Research*, vol. 42, no. 14, pp. 8996–9004, 2014.

[165] J. Huang, B. Z. Deborah, J. T. Reardon, Lippard S J, and A. Sancar, “HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. October, pp. 10394–10398, 1994.

[166] E. Mitkova, I. Ugrinova, I. G. Pashev, and E. A. Pasheva, “The Inhibitory Effect of HMGB-1 Protein on the Repair of Cisplatin-Damaged DNA Is Accomplished through the Acidic Domain,” *Biochemistry*, vol. 44, pp. 5893–5898, 2005.

[167] S. S. Lange, M. C. Reddy, and K. M. Vasquez, “Human HMGB1 directly facilitates interactions between nucleotide excision repair proteins on triplex-directed psoralen interstrand crosslinks,” *DNA Repair*, vol. 8, pp. 865–872, 2009.

[168] A. N. Kriatchko, S. Bergeron, and P. C. Swanson, “HMG-box domain stimulation of RAG1/2 cleavage activity is metal ion dependent,” *BMC Molecular Biology*, vol. 2, 2008.

[169] I. Brouwer, H. Zhang, A. Candelli, D. Normanno, E. J. G. Peterman, G. J. L. Wuite, I. Brouwer, H. Zhang, A. Candelli, D. Normanno, E. J. G. Peterman, and G. J. L. Wuite, “Human RAD52 Captures and Holds DNA Strands, Increases DNA Flexibility, and Prevents Melting of Duplex DNA: Implications for DNA Recombination,” *Cell Reports*, vol. 18, no. 12, pp. 2845–2853, 2017.

[170] D. Boal, *Mechanics of the Cell*. Cambridge: Cambridge University Press, 2012.

[171] L. Jen-Jacobson, L. E. Engler, and L. A. Jacobson, “Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins,” *Structure*, vol. 8, no. 00, pp. 1015–1023, 2000.

[172] R. H. Ebright, Y. W. Ebright, and A. Gunasekera, “Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E. coli lac DNA site,” *Nucleic Acids Research*, vol. 17, no. 23, pp. 16–18, 1989.

[173] V. Petri, M. Hsieh, and M. Brenowitz, “Thermodynamic and Kinetic Characterization of the Binding of the TATA Binding Protein to the Adenovirus E4 Promoter,” *Biochemistry*, vol. 34, pp. 9977 – 9984, 1995.

[174] K. A. Sharp, “Polyelectrolyte Electrostatics : Salt Dependence, Entropic, and Enthalpic Contributions to Free Energy in the Nonlinear Poisson-Boltzmann Model,” *Biopolymers*, vol. 36, pp. 227–243, 1995.

[175] S. S. Ruth and M. T. Record Jr, “Coupling of Local Folding to Site-Specific Binding of Proteins to DNA,” *Science*, vol. 263, no. February, pp. 777 – 784, 1994.

[176] E. Merabet and K. A. Gary, “Calorimetric Analysis of cl Repressor Binding to DNA Operator Sitest,” *Biochemistry*, vol. 34, pp. 8554–8563, 1995.

[177] C. Berger, I. Jelesarov, and H. R. Bosshard, “Coupled Folding and Site-Specific Binding of the GCN4-bZIP Transcription Factor to the AP-1 and ATF / CREB DNA Sites Studied by Microcalorimetry,” *Biochemistry*, vol. 2960, no. 96, pp. 14984–14991, 1996.

[178] B. Liedberg, C. Nylander, and I. Lundstrom, “Surface plasmon resonance biosensing for gas detection and biosensing,” *Sensors and Actuators*, vol. 4, pp. 299–304, 1983.

[179] S. Sjolander and C. Urbaniczky, “Integrated Fluid Handling System for Biomolecular Interaction Analysis,” *Analytical Chemistry*, vol. 63, no. 20, pp. 2338–2345, 1991.

[180] M. Raghavan, “BIAcore : a microchip-based system for analyzing the formation of macromolecular complexes,” *Structure*, vol. 3, pp. 331–333, 1995.

[181] S. Hearty, P. Leonard, and R. O'Kennedy, *Measuring Antibody–Antigen Binding Kinetics Using Surface Plasmon Resonance. In: Chames P. (eds) Antibody Engineering. Methods in Molecular Biology (Methods and Protocols)*. Humana Press, Totowa, NJ, 2012.

[182] H. Chardin, K. Mercier, C. Frydman, and N. Vollmer, "Surface Plasmon Resonance imaging : A method to measure the af fi nity of the antibodies in allergy diagnosis," *Journal of Immunological Methods*, vol. 405, pp. 23–28, 2014.

[183] J. D. Steckbeck, I. Orlov, A. Chow, H. Grieser, K. Miller, J. Bruno, J. E. Robinson, R. C. Montelaro, and K. S. Cole, "Kinetic Rates of Antibody Binding Correlate with Neutralization Sensitivity of Variant Simian Immunodeficiency Virus Strains," *Journal of Virology*, vol. 79, no. 19, pp. 12311–12320, 2005.

[184] S. N. Khilko, M. T. Jelonek, M. Corr, L. F. Boyd, A. L. M. Bothwell, and D. H. Margulies, "Measuring interactions of MHC class I molecules using surface plasmon resonance," *Journal of Immunological Methods*, vol. 183, pp. 77–94, 1995.

[185] K. H. Piepenbrink, B. E. Gloor, K. M. Armstrong, and B. M. Baker, "Methods for quantifying T cell receptor binding affinities and thermodynamics," *Methods Enzymol.*, vol. 466, pp. 359 – 381, 2009.

[186] Z. Salamon, Y. Wang, J. L. Soulages, M. F. Brown, and G. Tollin, "Surface Plasmon Resonance Spectroscopy Studies of Membrane Proteins : Transducin Binding and Activation by Rhodopsin Monitored in Thin Membrane Films," *Biophysical Journal*, vol. 71, no. 1, pp. 283–294, 1996.

[187] Z. Salamon, H. A. Macleod, and G. Tollin, "Surface plasmon resonance spectroscopy as a tool for investigating the biochemical and biophysical properties of membrane protein systems . I : Theoretical principles," *Biochimica et Biophysica Acta*, vol. 1331, pp. 117–129, 1997.

[188] V. I. Silin, E. A. Karlik, K. D. Ridge, and D. J. Vanderah, "Development of surface-based assays for transmembrane proteins : Selective immobilization of functional CCR5 , a G protein-coupled receptor," *Analytical Biochemistry*, vol. 349, pp. 247–253, 2006.

[189] A. Otto, "Excitation of Nonradiative Surface Plasmon Waves in Silver by the Method of Frustrated Total Reflection," *Z. Phys.*, vol. 410, pp. 398–410, 1968.

[190] E. Kretschmann and H. Raether, "Radiative Decay of Non Radiative Surface Plasmons Excited by Light," *Z. Naturforsch., Teil A*, pp. 2135 – 2136, 1968.

[191] P. A. V. D. Merwe, "Surface Plasmon Resonance - General Principles of Experiments," pp. 1–50, 2003.

[192] G. Healthcare, "Biacore ™ T200 Instrument Handbook," 2013.

[193] S. Lofas and B. Johnsson, “A Novel Hydrogel Matrix on Gold Surfaces in Surface Plasmon Resonance Sensors for Fast and Efficient Covalent Immobilization of Ligands,” *J. Chem. Soc., Chem. Commun.*, pp. 1526–1528, 1990.

[194] E. Stenberg, B. PErsson, H. Roos, and C. Urbaniczky, “Quantitative Determination of Surface Concentration of Protein with Surface Plasmon Resonance Using Radiolabeled Proteins,” *Journal of Colloid and Interface Science*, vol. 143, no. 2, pp. 513 – 526, 1991.

[195] D. J. O’Shannessy, M. Brigham-Burke, K. K. Soneson, P. Hensley, and I. Brooks, “Determination of Rate and Equilibrium Binding Constants for Macromolecular Interactions Using Surface Plasmon Resonance: Use of Nonlinear Least Squares Analysis Methods,” *Analytical Chemistry*, vol. 212, pp. 457 – 468, 1993.

[196] B. P. A. V. D. Merwe, D. L. Bodian, S. Daenke, P. Linsley, and S. J. Davis, “CD80 (B7-1) Binds Both CD28 and CTLA-4 with a Low Affinity and Very Fast Kinetics,” *J. Exp. Med.*, vol. 185, no. 3, 1997.

[197] P. Schuck, “Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 26, pp. 541 – 566, 1997.

[198] B. Goldstein, D. Coombs, X. He, A. R. Pineda, and C. Wofsy, “The influence of transport on the kinetics of binding to surface receptors : application to cells and BIAcore,” *Journal of Molecular Recognition*, vol. 12, pp. 293–299, 1999.

[199] S. Felder, M. Zhou, P. Hu, J. Urena, A. Ullrich, M. Chaudhuri, M. White, S. E. Shoelson, and J. Schlessingerl, “SH2 Domains Exhibit High-Affinity Binding to Tyrosine- Phosphorylated Peptides Yet Also Exhibit Rapid Dissociation and Exchange,” *Molecular and Cellular Biology*, vol. 13, no. 3, pp. 1449–1455, 1993.

[200] P. G. Stockley and B. Persson, *Surface Plasmon Resonance Assays of DNA-Protein Interactions. In: Leblanc B., Moss T. (eds) DNA-Protein Interactions. Methods in Molecular Biology™ (Methods and Protocols)*. Humana Press, 2009.

[201] M. Buckle, R. M. Williams, M. Negroni, and H. Buc, “Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance,” *Proceedings of the National Academy of Sciences*, vol. 93, no. January, pp. 889–894, 1996.

[202] H. Wang, A. Ramakrishnan, S. Fletcher, and E. V. Prochownik, “A quantitative , surface plasmon resonance-based approach to evaluating DNA binding by the c-Myc oncoprotein and its disruption by small molecule inhibitors,” *Journal of Biological Methods*, vol. 2, no. 2, 2015.

[203] I. D. Parsons, P. G. Stockley, A. Mekhalfia, and G. M. Blackburn, “Probing the molecular mechanism of action of co-repressor in the E . coli methionine repressor-operator complex using surface plasmon resonance (SPR),” *Nucleic Acids Research*, vol. 23, no. 2, pp. 211–216, 1995.

[204] J. Goyette, C. S. Salas, N. Coker-gordon, M. Bridge, S. A. Isaacson, J. Allard, and O. Dushek, “Biophysical assay for tethered signaling reactions reveals tether-controlled activity for the phosphatase SHP-1,” *Sci. Adv.*, vol. 3, p. e1601692, 2017.

[205] M. Dutreix, “(GT)n Repetitive Tracts Affect Several Stages of RecA-promoted Recombination,” *J. Mol. Biol.*, vol. 113, pp. 105–113, 1997.

[206] S. Subramanyam, W. T. Jones, M. Spies, and M. A. Spies, “Contributions of the RAD51 N-terminal domain to BRCA2-RAD51 interaction,” *Nucleic Acids Research*, vol. 41, no. 19, pp. 9020–9032, 2013.

[207] M. Durand, K. Chevrie, M. Chassignol, N. T. Thuong, and J. C. Maurizot, “Circular dichroism studies of an oligodeoxyribonucleotide containing a hairpin loop made of a hexaethylene glycol chain: conformation and stability,” *Nucleic Acids Research*, vol. 18, no. 21, pp. 6353–6359, 1990.

[208] A. H. El-Sagheer, R. Kumar, S. Findlow, J. M. Werner, A. N. Lane, and T. Brown, “A Very Stable Cyclic DNA Miniduplex with Just Two Base Pairs,” *ChemBioChem*, vol. 9, pp. 50–52, 2008.

[209] D. G. Myszka, “Improving biosensor analysis,” *J. Mol. Recognit.*, vol. 12, pp. 279–284, 2000.

[210] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, “Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems,” *J. R. Soc. Interface*, vol. 6, no. July 2008, pp. 187–202, 2009.

[211] P. Del Moral, M. Arnaud, and D. Ajay, “An adaptive sequential Monte Carlo method for approximate Bayesian computation,” *Stat. Comput.*, no. 1, 2011.

[212] C. C. Drovandi and A. N. Pettitt, “Estimation of Parameters for Macroparasite Population Evolution Using Approximate Bayesian Computation,” *Biometrics*, vol. 67, no. March, pp. 225–233, 2011.

[213] D. Silk, S. Filippi, and M. P. H. Stumpf, “Optimizing threshold-schedules for sequential approximate Bayesian computation : applications to molecular systems Bayesian computation,” *Statistical Applications in Genetics and Molecular Biology*, vol. 12, no. 5, pp. 603–618, 2013.

[214] M. A. Beaumont, J.-m. Cornuet, J.-m. Marin, C. P. Robert, J.-m. Marin, and C. P. Robert, “Adaptive approximate Bayesian computation,” *Biometrika*, vol. 96, no. 4, pp. 983–990, 2009.

[215] S. Filippi, C. P. Barnes, J. Cornebise, and M. P. H. Stumpf, “On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo,” *Statistical Applications in Genetics and Molecular Biology*, vol. 12, no. 1, pp. 87–107, 2013.

[216] L. A. Harris, J. S. Hogg, J. A. P. Sekar, S. Gupta, I. Korsunsky, A. Arora, D. Barua, R. P. Sheehan, and J. R. Faeder, “Systems biology BioNetGen 2.2 : advances in rule-based modeling,” *Bioinformatics*, vol. 32, no. July 2016, pp. 3366–3368, 2016.

[217] S. Subramanyam, M. Ismail, I. Bhattacharya, and M. Spies, “Tyrosine phosphorylation stimulates activity of human RAD51 recombinase through altered nucleoprotein filament dynamics,” *Proceedings of the National Academy of Sciences*, 2016.

[218] P. R. Edwards and R. J. Leatherbarrow, “Determination of Association Rate Constants by an Optical Biosensor Using Initial Rate Analysis,” *Analytical Biochemistry*, vol. 6, no. 246, pp. 1–6, 1997.

[219] R. Chen and M. S. W. Ã, “Prospects & Overviews Replication protein A: Single-stranded DNA’s first responder,” *BioEssays*, vol. 36, pp. 1156–1161, 2014.

[220] G. Schay, B. Borka, L. Kernya, E. Bulyaki, J. Kardos, M. Fekete, and J. Fidy, “Without Binding ATP, Human Rad51 Does Not Form Helical Filaments on ssDNA,” *J. Phys. Chem.*, vol. 120, pp. 2165 – 2178, 2016.

[221] D. Murugesapillai, M. J. Mccauley, L. J. M. Iii, and M. C. Williams, “Single-molecule studies of high-mobility group B architectural DNA bending proteins,” *Biophys Rev*, vol. 9, pp. 17–40, 2017.

[222] J. B. Mills, E. Vacano, and P. J. Hagerman, “Flexibility of Single-stranded DNA: Use of Gapped Duplex Helices to Determine the Persistence Lengths of Poly(dT) and Poly(dA),” *J. Mol. Biol.*, vol. 285, pp. 245 – 257, 1999.

[223] A. Y. L. Sim, J. Lipfert, D. Herschlag, and S. Doniach, “Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle x-ray scattering,” *Physical Review E*, vol. 86, pp. 1–5, 2012.

[224] C. Ke, M. Humeniuk, H. S-Gracz, and P. E. Marszalek, “Direct Measurements of Base Stacking Interactions in DNA by Single-Molecule Atomic-Force Spectroscopy,” *Physical Review Letters*, vol. 99, p. 018302, 2007.

[225] D. B. McIntosh, G. Duggan, Q. Gouil, and O. A. Saleh, “Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking,” *Biophysical Journal*, vol. 106, no. 3, pp. 659–666, 2014.

[226] J. Fan and N. P. Pavletich, “Structure and conformational change of a replication protein A heterotrimer bound to ssDNA,” *Genes & Development*, vol. 26, pp. 2337–2347, 2012.

[227] Y. U. E. Zou, Y. Liu, X. Wu, and S. M. Shell, "Functions of Human Replication Protein A (RPA): From DNA Replication to DNA Damage and Stress Responses," *Journal of Cellular Physiology*, vol. 273, no. January, pp. 267–273, 2006.

[228] Y. Kim, J. H. Geiger, S. Hahn, and P. B. Sigler, "Crystal structure of a yeast TBP/TATA-box complex," *Nature*, vol. 365, pp. 512 – 520, 1993.

[229] M. M. Vilenchik and A. G. Knudson, "Endogenous DNA double-strand breaks : Production, fidelity of repair, and induction of cancer," *Proceedings of the National Academy of Sciences*, vol. 100, no. 22, pp. 12871–12876, 2003.

[230] M. Reuter, A. Zelensky, I. Smal, E. Meijering, W. A. V. Cappellen, H. M. D. Gruiter, G. J. V. Belle, M. E. V. Royen, A. B. Houtsmuller, J. Essers, R. Kanaar, and C. Wyman, "BRCA2 diffuses as oligomeric clusters with RAD51 and changes mobility after DNA damage in live cells," *Journal of Cell Biology*, vol. 207, no. 5, pp. 599–613, 2014.

[231] T. Neumaier, J. Swenson, C. Pham, A. Polyzos, A. T. Lo, P. Yang, and J. Dyball, "Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells," *Proceedings of the National Academy of Sciences*, vol. 109, no. 2, 2012.

[232] F. Aymard, M. Aguirrebengoa, E. Guillou, B. M. Javierre, B. Bugler, V. Rocher, J. S. Iacovoni, A. Biernacka, M. Skrzypczak, K. Ginalski, M. Rowicka, P. Fraser, and G. Legube, "Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes," *Nature Structural and Molecular Biology*, vol. 24, no. 4, pp. 353–361, 2017.

[233] A. Tubbs, S. Sridharan, N. V. Wietmarschen, P. D. Aplan, O. Hyrien, A. Tubbs, S. Sridharan, N. V. Wietmarschen, Y. Maman, E. Callen, and A. Stanlie, "Dual Roles of Poly (dA:dT) Tracts in Replication Initiation and Fork Collapse," *Cell*, vol. 174, pp. 1127–1142, 2018.

[234] C. Forrey, J. F. Douglas, and M. K. Gilson, "The Fundamental Role of Flexibility on the Strength of Molecular Binding," *Soft Matter*, vol. 8, pp. 6385–6392, 2012.

[235] G. Schwarz and F. Watanabe, "Thermodynamics and Kinetics of Cooperative Protein-Nucleic Acid Binding I. General Aspects of Analysis of Data," *J. Mol. Biol.*, vol. 163, pp. 467 – 484, 1983.

[236] J. Sturm, "Binding of Ligands to a One-Dimensional Heterogeneous Lattice. I. General Model for the Calculation of Binding Isotherms," *Biopolymers*, vol. 20, no. 1, pp. 753–763, 1981.

[237] W. Bujalowski, T. M. Lohman, and C. F. Anderson, "On the Cooperative Binding of Large Ligands to a One-Dimensional Homogeneous Lattice: The Generalized Three-State Lattice Model," *Biopolymers*, vol. 28, pp. 1637–1643, 1989.

[238] M. Takahashi, “Analysis of DNA-RecA Protein Interactions Involving the Protein Self-association Reaction,” *Journal of Biological Chemistry*, vol. 264, pp. 288–295, 1989.

[239] Y.-d. Chen, “A General Secular Equation for Cooperative Binding of n-mer Ligands to a One-Dimensional lattice,” *Biopolymers*, vol. 30, pp. 1113–1121, 1990.

[240] F. Watanabe and G. Schwarz, “Thermodynamics and Kinetics of Acid Binding II . Studies on the Binding Between Protamine and Calf Thymus DNA,” *J. Mol. Biol.*, vol. 163, pp. 485–498, 1983.

[241] R. P. C. Driessen, G. Sitters, N. Laurens, G. F. Moolenaar, G. J. L. Wuite, N. Goosen, and R. T. Dame, “Effect of Temperature on the Intrinsic Flexibility of DNA and Its Interaction with Architectural Proteins,” *Biochemistry*, 2014.

[242] D. T. Gillespie, “A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions,” *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.