Structure and mechanism of the BRCA2 tumour suppressor

Taha Shahid

31 October 2014

Thesis submitted in fulfilment for the degree of
Doctor of Philosophy in Structural Molecular Biology (PhD)

Department of Life Sciences
Imperial College London
Copyright

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Abstract

Mutations in the human BRCA2 gene are a leading cause of susceptibility to breast, ovarian and prostate cancers. Its protein equivalent BRCA2, is the key mediator in repair of double-stranded DNA breaks and interstrand crosslinks – the most dangerous forms of DNA damage, via RAD51 effected homologous recombination. Thereby, preferably the sister chromatid, or homologous chromosome is utilised to drive the repair. Described here is the first structural and related mechanistic characterisation of full-length (384 kDa) BRCA2, and its complex with RAD51 and DNA. Electron microscopy reconstruction reveals that BRCA2 exists as a dimer in head-to-tail conformation, which binds two oppositely directed sets of RAD51 molecules. Single-stranded DNA (derived from exonucleaseolytic processing of double-strand breaks) binds BRCA2 along its long axis, such that one set of RAD51 monomers always binds thereon regardless of its binding polarity. These then self-assemble into RAD51–ssDNA nucleoprotein filaments that catalyse the main reaction, having been nucleated and primed for growth by BRCA2. In complementary work, it is shown that BRCA2 increases the frequency of RAD51–ssDNA filament formation, but does not impact filament length itself. It is also shown that BRCA2 initiates monodirectional (3’→5’) extension of the RAD51 filament on single-stranded DNA from nucleation sites (thence at the 3’ end). Furthermore, BRCA2 initiates such nucleations at multiple sites along individual ssDNA molecules, suggesting that these extend and merge into a single nucleoprotein filament, which would be most capable of catalysing the subsequent homologous pairing and strand exchange reactions. Together, the structures and biochemical data define the molecular mechanism via which BRCA2 orchestrates the formation of RAD51 nucleoprotein filaments, which constitute the central reaction intermediate in the repair of DNA double-strand breaks.
Attestation

I certify that the experiments and analyses reported hereby, except when explicitly identified, were performed by the author. For the reader’s facility, the author’s specific contributions are set out below:

• The BRCA2 and BRCA2–RAD51 reconstructions (data collection, processing and analyses).
• BRCA2 and BRCA2–RAD51 antibody complexes (data collection, processing and analyses).
• Gapped DNA complexes with BRCA2 (data collection, processing and analyses).
• Tilt-series and tilt-pair structural validations and handedness studies (methods in §2.5.4.10 and §2.5.4.10.1, data collection, processing and analyses).
• The BRCA2–RAD51–ssDNA complex (processing and analysis, §2.5.4.13 and §3.2.11.2.1).
• Statistical calculations from scanning TEM data.

All figures are original/the author’s unless given due mention. All prose/exposition is entirely the author’s own.
Acknowledgements

I wish to acknowledge a number of individuals who between them made my PhD possible.

Prof Xiaodong Zhang, for above all giving me the opportunity to work on, and entrusting me with a great project.

My family, without whose love and support this PhD would not be possible. I am hugely indebted for the many sacrifices, of not only a financial nature, which they have made in way of my higher education. And Tianyi Zhang, without whom this thesis would not be possible. I am grateful to them for unwavering faith in my abilities, and dedication to my works. Thus, it is to them that I dedicate my PhD thesis.

My good friends, Dr Simon Parker and Robert Mooney, for always believing in the best of me.

I also thank:

Tianyi Zhang for implementing the figures in Inkscape and MacPyMOL.

Dr Eric Kong for his diligence as a co-worker on the project.

Drs Daniel Bose, Caroline Ewens, and Tillmann Pape for initiating me into image processing for single particle electron microscopy. And Dr Tillmann Pape for his warm and thorough mentoring in electron microscopy.

Dr Ric Aramayo for his friendly help and teaching when asked.
Publication from this thesis

# Table of Contents

Copyright ........................................................................................................................................ i
Abstract .......................................................................................................................................... ii
Attestation ......................................................................................................................................... iii
Acknowledgements ......................................................................................................................... iv
Publication from this thesis ............................................................................................................... v
Table of Contents ........................................................................................................................... vi
List of Figures .................................................................................................................................... ix

1 Introduction ................................................................................................................................. 1
  1.1 BRCA2 and the cancer link ....................................................................................................... 1
    1.1.1 Other breast cancer susceptibility genes ........................................................................... 1
  1.2 DNA damage ............................................................................................................................ 2
    1.2.1 DNA damage and cancer .............................................................................................. 3
    1.2.2 DNA double-stranded breaks ....................................................................................... 4
  1.3 DNA repair ............................................................................................................................... 5
    1.3.1 DNA double-stranded break repair .............................................................................. 6
      1.3.1.1 Non-homologous end-joining .................................................................................. 7
      1.3.1.2 Homologous recombination ................................................................................. 9
  1.4 The RAD51 recombinase ......................................................................................................... 17
    1.4.1 Cell viability and chromosome stability .......................................................................... 17
    1.4.2 RAD51 orthologues ....................................................................................................... 19
    1.4.3 RAD51 filament structure ............................................................................................ 20
  1.5 The BRCA2 tumour suppressor ............................................................................................... 23
    1.5.1 BRCA2 and cell proliferation ....................................................................................... 25
    1.5.2 BRCA2 and genomic stability ....................................................................................... 27
    1.5.3 Colocalisation of BRCA2 and RAD51 ......................................................................... 29
    1.5.4 Structure of BRC4 – RAD51 (ATPase domain) ............................................................ 30
    1.5.5 Two classes of BRC repeats ......................................................................................... 35
    1.5.6 Structure of BRCA2DBD – DSS1 – ssDNA ................................................................... 36
  1.6 In vitro full-length BRCA2 ..................................................................................................... 40
    1.6.1 Overview of thesis ....................................................................................................... 43

2 Methods ....................................................................................................................................... 45
  2.1 Protein expression and purification ....................................................................................... 45
    2.1.1 BRCA2 .......................................................................................................................... 45
    2.1.2 BRCA2-RAD51 ............................................................................................................ 45
2.2 DNAs ........................................................................................................46
  2.2.1 DNA gold-labelling ........................................................................46
2.3 Protein-DNA complexes ........................................................................47
  2.3.1 BRCA2-ssDNA ..............................................................................47
  2.3.2 BRCA2-RAD51-ssDNA ..................................................................48
    2.3.2.1 Protein gold-labelling .............................................................48
2.4 Antibody labelling ..................................................................................49
2.5 Negative-stain electron microscopy ........................................................49
  2.5.1 Specimen preparation .......................................................................49
  2.5.2 Data acquisition ...............................................................................50
  2.5.3 Data preprocessing ..........................................................................51
    2.5.3.1 Coarsening .............................................................................51
    2.5.3.2 Particle selection ....................................................................52
    2.5.3.3 Filtering, masking and normalisation .......................................52
  2.5.4 Image processing (IMAGIC V) ..........................................................53
    2.5.4.1 Reference-free alignment (centring) .........................................54
    2.5.4.2 Multivariate statistical analysis (MSA) ....................................55
    2.5.4.3 Hierarchical ascendant classification (HAC) ..........................56
    2.5.4.4 Class averaging .......................................................................57
    2.5.4.5 Multi-reference alignment (MRA) ..........................................58
    2.5.4.6 Euler angle assignment ............................................................60
    2.5.4.7 Three-dimensional reconstruction ...........................................62
    2.5.4.8 Scanning transmission electron microscopy (STEM) ............66
    2.5.4.9 Resolution ...............................................................................67
    2.5.4.10 Structural validation ...............................................................69
    2.5.4.11 Antibody complexes .............................................................74
    2.5.4.12 gDNA complexes ..................................................................76
    2.5.4.13 BRCA2–RAD51+ssDNA .........................................................76
3 Results ........................................................................................................78
  3.1 Purification ............................................................................................78
  3.2 Negative stain electron microscopy .......................................................79
    3.2.1 Micrographs and particles .............................................................79
    3.2.2 MSA and classification .................................................................81
    3.2.3 Angular assignment and particle alignment ..................................84
    3.2.4 Structures of BRCA2 and BRCA2–RAD51 ..................................87
    3.2.5 Resolution .....................................................................................92
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.6 Structural validation</td>
<td>94</td>
</tr>
<tr>
<td>3.2.7 Structural domains</td>
<td>97</td>
</tr>
<tr>
<td>3.2.8 Fitting, conformational change in BRCA2 on RAD51 binding, and</td>
<td>100</td>
</tr>
<tr>
<td>handedness</td>
<td></td>
</tr>
<tr>
<td>3.2.9 STEM mass measurements</td>
<td>101</td>
</tr>
<tr>
<td>3.2.10 BRCA2–ssDNA</td>
<td>103</td>
</tr>
<tr>
<td>3.2.11 BRCA2–RAD51–ssDNA dynamics (West laboratory)</td>
<td>105</td>
</tr>
<tr>
<td>3.2.11.1 Filament frequency and length</td>
<td>105</td>
</tr>
<tr>
<td>3.2.11.2 Filamentisation polarity</td>
<td>106</td>
</tr>
<tr>
<td>3.2.11.3 Multinucleation</td>
<td>109</td>
</tr>
<tr>
<td>4 Discussion</td>
<td>111</td>
</tr>
<tr>
<td>4.1 BRCA2 DNA-binding domain and RPA</td>
<td>111</td>
</tr>
<tr>
<td>4.2 Dimerisation and its significance</td>
<td>114</td>
</tr>
<tr>
<td>4.3 Polarity of filament growth and strand transfer</td>
<td>116</td>
</tr>
<tr>
<td>References</td>
<td>120</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. Double-stranded DNA break repair by non-homologous end-joining. ..................8
Figure 2. Double-stranded DNA break repair by homologous recombination. ..................13
Figure 3. Resolution of recombinational repair products. ................................................16
Figure 4. Karyotype analysis of gamma-irradiated Rad51-deficient cells. .........................18
Figure 5. RAD51 protein sequence and orthologues. ......................................................20
Figure 6. Crystal structure of a Rad51 filament. ..............................................................23
Figure 7. Human BRCA2 protein sequence and major binding partners. .........................25
Figure 8. Crystal structure of a Rad51 filament. ..............................................................27
Figure 9. Crystal structure of BRCA2 DNA/DSS1-binding domain. ..............................39
Figure 10. EM of BRCA2 and BRCA2–RAD51. ............................................................81
Figure 11. Eigenimages, class averages and belonging particles. .....................................84
Figure 12. Angular orientation distributions. .................................................................87
Figure 13. 3D reconstruction of BRCA2 and BRCA2–RAD51 complex. .........................91
Figure 14. Resolution of reconstructions. ......................................................................93
Figure 15. Validation of structures. ..............................................................................97
Figure 16. Antibody complexes and identification of structural domains. ......................99
Figure 17. Rearrangement of BRCA2 upon RAD51 binding. ........................................101
Figure 18. STEM mass determination and statistical analysis. .......................................102
Figure 19. 2D analysis of BRCA2–gDNA complexes. ..................................................104
Figure 20. Nucleation of RAD51–ssDNA filaments by BRCA2. ..................................105
Figure 21. Polarity of RAD51–ssDNA filament growth. ..............................................107
Figure 22. Multiple RAD51 filament nucleations along ssDNA. ...................................109
Figure 23. Comparison of BRCA2DBD–ssDNA and RPA–ssDNA crystal structures. ....113
Figure 24. Mechanism of action of BRCA2. .................................................................115
Figure 25. Further mechanistic details. .......................................................................118
1 Introduction

1.1 BRCA2 and the cancer link

An estimated 1.7 million women are diagnosed with, and 0.5 million die of breast cancer worldwide each year\(^1\). As such, breast cancer constitutes the most common malignancy among women, with a lifetime risk in excess of 10\(^%\)\(^2\).

Whilst most breast cancers are sporadic, some are the result of inherited predisposition, primarily due to mutations in one of the two tumour-suppressor breast cancer susceptibility genes, \textit{BRCA1} and \textit{BRCA2} (breast cancer 1 and 2, early onset)\(^2\).

Women born with mutations in either gene are at a 10- to 20-fold risk relative to the general population of developing breast (and ovarian) cancer\(^3\). In fact, the lifetime risk of breast cancer among female mutation carriers now exceeds 80\(^%\)\(^2\).

And while both \textit{BRCA1} and \textit{BRCA2} mutation carriers display markedly increased incidences of breast and ovarian cancers, the latter also exhibit increased risks of male breast, prostate and pancreatic cancers\(^4\). Furthermore, other cancers (e.g. laryngeal cancer), may also be elevated in carriers of \textit{BRCA2} mutations\(^5\).

Risk variants of \textit{BRCA2} result in inactivation of the encoded protein, predominantly by causing premature protein truncation (for instance as a consequence of disruption to the open reading frame of the transcriptional unit), or nonsense-mediated RNA decay\(^3\).

Although \textit{BRCA1} and \textit{BRCA2} act as recessive genes at the cellular level, cancer susceptibility is transmitted as an autosomal dominant trait in families harbouring mutations. This is because heterozygous mutations get converted to homozygosity in the cancers that they cause, typically by loss of the wild-type allele\(^3\).

These findings bear huge clinical ramifications for affected women. Options facing these women include medical surveillance (screening starting at a very early age), chemoprevention, and often pre-emptive prophylactic mastectomy and risk-reducing removal of the ovaries\(^2,6\).

The genetic basis for inherited predisposition towards breast cancer has been intensively studied since the last decade of the 20\(^\text{th}\) century, with several years of biological research having firmly implicated \textit{BRCA1} and \textit{BRCA2} in the repair of DNA double-strand breaks\(^3\).

1.1.1 Other breast cancer susceptibility genes

Three distinct classes of breast cancer susceptibility genes with different magnitudes of risk have been recognised\(^3\).
**BRCA1** and **BRCA2** constitute the two major high-penetrance alleles. Mutations in **TP53** (tumour protein p53) confer a high risk for breast and other cancers, but these mutations are very rare and account for a much smaller fraction of the familial risk.\(^3\)

A group of moderate-penetrance susceptibility alleles has been identified based principally on the participation of the encoded proteins in pathways that incorporate BRCA1 and BRCA2. This includes **ATM** and **CHEK1** (checkpoint kinases involved in DNA repair that phosphorylate BRCA1 and p53), **BRIP1** (a binding partner of BRCA1 implicated in some of its DNA repair activities), and **PALB2** (partner and localiser of BRCA2 – a high affinity BRCA2 binding protein). Mutations in these genes confer a two- to four-fold relative risk of developing breast cancer.\(^3\)

Finally, a cohort of low-penetrance alleles has been recognised. Mutant genes of this type extend a relative breast cancer risk of up to ~1.25-fold for heterozygotes and 1.65-fold for homozygotes. Examples include **CASP8**, **FGFR2**, **TOX3**, **MAP3K1**, and **LSP1**.\(^3\)

### 1.2 DNA damage

Life’s fundamental objective is the preservation, and conveyance of its genetic material, intact and unchanged to the next generation.

This has to be accomplished in spite of the perpetual damage being inflicted upon DNA, both by exogenous agents such as ionising radiation and genotoxic chemicals, as well as endogenous sources like free radicals generated as by-products of normal cellular activities. Furthermore, DNA itself being a dynamic structure with high energy is prone to a degree of physicochemical instability that leaves the potential for such and spontaneous damages to occur\(^7\).

In fact, each of the ~10 trillion cells in the human body incurs tens of thousands of DNA lesions per day\(^8\).

A variety of lesions arise in DNA at significant rates *in vivo*, with the most common types and their respective instances in a single cell per day being as follows: 20,000 single-stranded breaks; 10,000 base-hydrolysis events; 5,000 alkylating lesions; 2,000 oxidising lesions; and 10–20 double-strand breaks\(^9\).

Most ubiquitous of the environmental DNA-damaging agents is ultraviolet light. UV exposure in strong sunlight can induce ~100,000 lesions per exposed cell per hour. Some ionising damage also results from the decay of naturally occurring radioactive substances, e.g. uranium and its radioactive decay product radon gas. Presently, the most widespread DNA-damaging chemicals are those contained in tobacco products. However, they can also be found in certain foods, such as heterocyclic amines in overcooked meats.\(^8\)
DNA damage arises via physiological processes as well, e.g. base mismatches that occasionally get introduced during DNA replication. A further thousands of base lesions per cell per day are caused by hydrolytic reactions and non-enzymatic methylations. DNA lesions are also created by reactive oxygen compounds, arising for instance as by-products of oxidative respiration. White blood cells too yield reactive oxygen and nitrogen compounds at sites of infections and inflammation. Such agents can attack DNA, leading to adducts that impair base pairing and/or block DNA replication and transcription, cause base loss or single-stranded DNA breaks. And when two DNA single-strand breaks arise in close proximity to each other, or the DNA replication apparatus encounters a single-strand break (or certain other types of DNA lesion), double-strand breaks are formed.8

These DNA lesions besides blocking genome replication and transcription, if mis- or dis-repaired can lead to mutations and wider-scale genomic aberrations, which threaten cell and organism viability – and underpin, and are the hallmark of cancer8.

1.2.1 DNA damage and cancer

Cancer and DNA damage are intimately related, and genomic instability is an elemental driver and feature of cancer. That damage to DNA is the root cause of all cancer has been established since early in the last decade of the 20th century.10 Both exogenous and endogenous carcinogens act by creating DNA damage and inducing mutations8.

Two major kinds of genomic instability are associated with cancer cells. The mutational instability phenotype is typified by point mutations or small deletions, while the chromosomal instability phenotype is characterised by gross rearrangement of chromosomes.11

Mutations may promote carcinogenesis by the activation of proto-oncogenes, or the inactivation of tumour suppressor genes8,10.

Proto-oncogenes are normal genes that operate as positive growth regulators, involved in promoting the proliferation and differentiation of normal cells. A variety of them are implicated in various crucial steps of a cell’s growth, and change in a proto-oncogene’s sequence or in the amount of protein its produces can interfere with its normal role in cell growth regulation. Runaway cellular growth may ensue, leading to the formation of a cancerous tumour.12,13

An oncogene is an altered or mutated version of a proto-oncogene. The latter can get transformed into the former in three ways. Point mutations can arise spontaneously or due to endogenous or environmental influences, resulting in the production of an altered protein that cannot be properly regulated. A second means of oncogenesis takes place via the process of chromosomal translocation, whereby a segment of a chromosome breaks off and attaches
to another chromosome. Should the dislocated chromosome contain a proto-oncogene, it could be removed from its regulatory controls and be continually produced. This excess production of protein molecules stimulates the cellular process under their control, upsetting the balance of the mechanisms of cell growth. The third route entails an amplification of the number of copies of a proto-oncogene, which also results in an overproduction of the protein – and its accompanying effects.13

Tumour suppressor genes, also called anti-oncogenes, are also a class of normal genes. They encode proteins that regulate cell growth by inhibiting the proliferation of cells, which too is critical for regular cell development and differentiation; repair DNA mistakes and damage; and instruct cells when to die (the process of apoptosis, also known as programmed cell death). Hence, tumour suppressor genes also act to curb the uncontrolled growth of cancer cells. Cancer-causing alterations in these genes are acquired by very similar methods to proto-oncogenes, and too act to disrupt the balance between restraint and activation of the molecular processes that regulate a cell’s life cycle – contributing to the uncontrolled cell growth characteristic of cancerous tumours.12,14

Normally, cells detect DNA injury by blocked transcription, replication or specialised sensors, trigger cell-cycle arrests at specific checkpoints in the G1, S, G2 and M phases, and repair the lesions before they get converted into permanent mutations10. If the damage is too significant, cell death can result, which can happen in two different ways. Cells may enter a state of irreversible growth arrest (replicative cell death), or opt for the ultimate mode of rescue by initiating apoptosis at the expense of the whole cell11. However, if the mutation is not repaired, and the cell does not die, a person may develop cancer, which is much more likely if the mutation is in a tumour suppressor gene12.

1.2.2 DNA double-stranded breaks

Of all kinds of DNA damage, the DNA double-strand break constitutes the most dangerous form of lesion that can occur in DNA, and its repair one that poses the greatest challenge to cells15. The former because a double-strand break involves the risk of losing a whole part of a chromosome and potentially hundreds or thousands of genes alongside, and the latter as unlike in most other sorts of damage both strands of the double-helix are affected such that there is no immediately available template in the form of an intact complementary strand to use for correcting the lesion. In addition, to properly heal such a break the cell has to know which ends of the break belong together, an especially difficult task given the size of larger genomes such as those of humans10. As a result, DNA double-strand breaks are potent inducers of the chromosomal instability phenotype11.
Despite not occurring as frequently as other lesions, DNA double-strand breaks are the most lethal of all the types of DNA damage. While cells can adapt to low levels of unrepaired or irreparable lesions of other kinds, a single residual DNA double-strand break can kill a cell if it inactivates an essential gene or triggers apoptosis.

DNA double-strand breaks could be induced by any of the factors mentioned before or from mechanical stress on the chromosomes, however provided the comparative stability of double-stranded DNA they would unlikely be caused as a direct consequence of the usual physical and chemical carcinogens – a notion perhaps reflected in the relatively low frequency of these breaks. Instead, double-strand breaks are far more likely to be caused as a consequence of the fragility of the DNA replication machinery itself, for instance when a replication fork traverses a template containing a single-strand break, it will be converted into a double-strand break on one of the sister chromatids in the process, with it possibly likewise being the case for certain other forms of unrepaired lesion – particularly interstrand crosslinks.

Frequent chromosomal aberrations induced by double-stranded DNA breaks include the loss, gain, and exchange of chromosome segments. Loss of a chromosome fragment can inactivate tumour suppressor genes (e.g. by loss of heterozygosity), whereas amplification can activate proto-oncogenes, should the respective regions encode such genes – thus leading to carcinogenesis. Chromosomal translocations, which often occur in a balanced (reciprocal) fashion, can lead to the dysregulation of gene expression or fusion of two genes that then acquire oncogenic potential – and are thought to emerge from double-stranded breaks on the two participating chromosomes. Moreover, inactivated tumour suppressor genes and activated oncogenes, ensue DNA replication stress with double-stranded break formation, fuelling chromosomal instability continuously.

DNA double-stranded breaks also pose problems, and are especially toxic, if they occur during mitosis – as intact chromosomes are imperative for correct chromosome segregation during cell division. Proper genomic replication is obstructed by double-stranded breaks, and should broken chromosomes be carried through mitosis, the acentric chromosome fragments would not partition equally between daughter cells.

### 1.3 DNA repair

In view of the diversity of DNA damage, evolution has invested significantly in implementing a tapestry of sophisticated repair mechanisms that each target a subset of different lesions – and between them counteract the deleterious effect of the various types of damage inflicted on an organism’s genomic information. The main kinds of repair pathways and their roles, in increasing order of the complexity of the problem are: base-excision repair –
targeting small chemical alterations of bases; nucleotide-excision repair – dealing with helix-distorting damage to bases; mismatch repair – correcting mispaired bases; interstrand cross-link and double-strand break repairs respectively\textsuperscript{10}.

While some DNA repair systems entail a direct enzyme-mediated reversal of the lesion (e.g. O\textsuperscript{6}-methylguanine methyltransferase, which removes non-native O\textsuperscript{6}-methyl groups from guanine residues), most progress by a sequence of catalytic reactions mediated by multiple proteins\textsuperscript{8,10}. In base-excision repair, the damaged base is usually recognised by a DNA glycosylase enzyme that catalyses its removal, before nuclease, polymerase and ligase proteins finish the repair\textsuperscript{8}. The nucleotide-excision repair pathway, acts by way of one of two sub-pathways, which differ in the manner of lesion detection: transcription-coupled repair that targets transcription-blocking lesions, and global genome nucleotide-excision repair\textsuperscript{8}. As a key aspect of nucleotide-excision repair, the damage is excised as a several base oligonucleotide, which gets operated upon by DNA polymerases and accompanying factors, before ensuing ligation\textsuperscript{8}. With mismatch repair, the recognition of a mismatched base pair, or an insertion/deletion loop, prompts a single-stranded incision that is then acted upon by nuclease, polymerase and ligase enzymes\textsuperscript{8} (cf. base-excision repair). Common to all these repair processes, is a ‘cut-and-patch’-style reaction, whereby the lesion (with or without flanking sequences), is extracted and the consequential single-strand gap filled-in, by making use of the undamaged complementary strand as template\textsuperscript{10}.

Double-strand break repair on the other hand, is more problematic as both strands are affected. Two chief, distinct and complementary mechanisms have evolved to tackle this problem. These are non-homologous end-joining, and homologous recombination\textsuperscript{10}.

\subsection{DNA double-stranded break repair}

The importance and difficulty of repairing double-stranded DNA breaks is highlighted by the fact that all eukaryotic cells have developed several pathways for dealing with this type of injury. The two major repair modes, homologous recombination and non-homologous end-joining, differ in their requirement for a homologous DNA template and fidelity of the repair reaction. While the former requires such a template and guarantees accurate repair, the latter does not.\textsuperscript{11}

The relative contribution of these two double-strand break repair processes is likely to differ depending on the stage of the cell cycle\textsuperscript{11}. Homologous recombination is most efficient, and seems to dominate in the S and G2 phases, whereby the DNA is replicated providing a pristine second copy of the sequence (sister chromatid), for aligning the breaks and mediating faithful repair\textsuperscript{10,11}. By contrast, the less accurate end-joining is more relevant in G1, when such a copy is unavailable\textsuperscript{10}.
1.3.1.1 Non-homologous end-joining

Non-homologous end-joining occurs without a template, and involves simply the re-ligation of the broken termini in what is a very efficient splicing reaction\(^\text{19}\) – see Fig. 1. However, therein lies a potential problem, in that in the event of a DNA double-strand break the exposed ends of the break are vulnerable to attack from other enzymes within the cell and there exists a non-trivial possibility of DNA at the break-site losing several bases – that this pathway links together with little or no regard for what the true sequence should be.

Since end-joining is typically accompanied by a modification of the sequence information at the break site, this pathway is inherently mutagenic\(^\text{11}\). However, so long as the modification is in a non-coding region of the DNA – it should constitute a ‘silent mutation’.

A core set of enzymes mediate repair via this process. They include the 70K and 80K subunits of the KU heterodimer (KU70/KU80); the catalytic subunit of DNA-dependent protein kinase (DNA-PK\(_{\text{CS}}\)); X-ray repair cross-complementing protein 4 (XRCC4); and DNA ligase IV (LIG4).\(^\text{20}\)

Upon double-strand break formation, a KU heterodimer binds each of the broken DNA ends\(^\text{20}\). Once bound, KU attracts DNA-PK\(_{\text{CS}}\)\(^\text{11}\), which performs a regulatory role in the process involving its protein kinase activity\(^\text{20}\) – leading to the recruitment of end-processing enzymes, polymerases and DNA ligase IV\(^\text{8}\). The two KU heterodimers then associate to form a bridging complex that enlists further proteins, and acts as an alignment factor to promote end-joining\(^\text{20}\). Polymerases \(\mu\) and \(\lambda\) fill in gaps, while nucleases such as Artemis trim excess ends as necessary to create flush-ended DNA – before an XLF-XRCC4-DNA ligase IV complex finishes the repair\(^\text{8,20-23}\).

Cells in G1 have only the homologous chromosome, and not the sister chromatid for re-combinational repair. This may be difficult to find in the complex genome, and could even be detrimental as a template for repair – by virtue of causing homozygosity for recessive mutations\(^\text{10}\). Hence, in certain circumstances, the non-homologous end-joining pathway may indeed be advantageous.

1.3.1.1.1 Alternative end-joining

Another, less well known and characterised KU-independent variant of non-homologous end-joining called microhomology-mediated end-joining (or alternative end-joining) also exists, but always results in sequence deletions\(^\text{8}\).

The defining feature of this repair mechanism is the use of extensive (~5-25) base pairs of microhomologous sequences for the alignment of broken ends prior to annealing, therefore resulting in large deletions flanking the original break\(^\text{23}\).
By contrast, non-homologous end-joining exploits short (<5) nucleotide microhomologies, leading to if anything small deletions or insertions in the repair product.23

**Figure 1.** Double-stranded DNA break repair by non-homologous end-joining. After double-strand break formation, the KU–DNA-PKcs complex is involved in the initial recognition of the break and the juxtaposition of the DNA ends.11 Thereby, the signalling function is thought to be performed by the DNA-PKcs component (purple), whilst the KU70/KU80 heterodimer (dark/pale orange) binds with avidity to the DNA termini – as contacts between the two subunits lead to the formation of a highly charged pore through which DNA passes.19 KU can also promote end-alignment, likely by confining DNA movement to a singular path.19 The ends may be processed by other proteins, such as Artemis (pale blue) or Pol μ and λ (yellow), which results in the removal or addition of a few base-pairs if internal microhomologies are used for annealing before sealing.8,10,11 This is followed by an end-to-end ligation by the XLF-XRCC4-DNA ligase IV (dark blue, green, & red respectively) complex that terminates the process.8,11
resulting in any sequences lost at the break-site being fixed as a mutation or deletion in the rejoined product.\cite{footnote}

1.3.1.2 Homologous recombination

Homologous recombination is a most important mechanism for the repair of DNA double-strand breaks and inter-strand crosslinks, preventing the demise of stalled or broken replication forks, and several other dimensions of chromosome maintenance.\cite{footnote} Accordingly, it is imperative for genomic integrity, as impairment in repairing such complex DNA damage and resolving DNA replication stress, contributes to chromosomal instability and leads to tumour aetiology.\cite{footnote} Therefore, dysfunction or deficiency in this process, brought about by mutations in proteins that play a mediator role therein – yields the cancer phenotype.\cite{footnote}

The homologous recombination pathway acts to restore double-stranded breaks in DNA, by employing an intact and undamaged homologous DNA molecule as template, which could either be the sister chromatid (preferably) or homologous chromosome – thus entailing an error-free reaction.\cite{footnote}

Furthermore, faithful duplication of the genome is intimately linked to homologous recombination, as imperfections in the DNA template often present a strong impediment to the replication machinery, and cause an arrest and breakdown of the replication fork.\cite{footnote} As such, the issuing DNA intermediates get acted upon by recombinatory factors that then reinstitute a functional replication fork.\cite{footnote}

Interstrand crosslinks it has been established are first converted into double-strand breaks in a replication-dependent manner.\cite{footnote} Although potentially dangerous, the latter are needed to remove the former and facilitate subsequent homology-directed recombinational repair.\cite{footnote} Homologous recombination mutants are extremely sensitive to interstrand crosslink-inducing drugs, suggesting it is the main pathway for double-strand break-associated interstrand crosslink repair.\cite{footnote} Moreover, human cells defective for non-homologous end-joining components are insensitive to interstrand crosslinking agents, showing end-joining is not required for their repair.\cite{footnote}

All organisms ranging from bacteria to mammals utilise the highly conserved family of RecA and RAD51-like recombinases to catalyse the DNA-DNA interactions required for homologous recombination.\cite{footnote} It so happens that in humans, the BRCA2 protein mediates the central enzyme RAD51 that carries out this process.\cite{footnote} And while mutations in BRCA2 do not remove the whole pathway, they seriously impair its efficiency.\cite{footnote}
1.3.1.2.1 End-resection

The initial cellular response is carried out by the MRE112-RAD502-NBS1 (MRN or N/M/R) complex11, which senses the double-strand break – and rapidly relocates to the damage site28.

Thereby, the MRE11 dimer can facilitate short-range synapsis of the two broken termini that were once contiguous28. Each MRE11 molecule binds a single copy of RAD5028, which possesses two coiled-coil regions that act as flexible ‘arms’ that protrude away from the DNA28,29. Through a zinc-‘hook’ located at the distal terminus of their coiled-coil arms, the two RAD50 molecules are able to dimerise with RAD50 subunits from an MRN complex assembled on the sister chromatid (or homologous chromosome) – thus serving as long-range tethers between the two28,29 (see Fig. 2i). NBS1 (Nijmegen breakage syndrome 1) or Nibrin, by virtue of its nuclear localisation signal and binding to MRE11, is responsible for recruiting the MRE112-RAD502 core complex from the cytoplasm into the nucleus28.

Then, the CtIP protein binds MRN, and catalyses or confers upon it 5’→ 3’ exonuclease activity to create 3’-OH ending single-stranded DNA tails at either side of the break – in a process termed end-resection28 (Fig. 2ii). Following initial nucleolytic processing by MRN/CtIP of ~100 nucleotides, further resection is exercised by ExoI in conjunction with BLM (Bloom syndrome) helicase – that results in ~2-3 kilobase stretches of single-stranded DNA28 (Fig. 2iii).

1.3.1.2.2 Presynapsis

These 3’ overhangs are first bound and coated by replication protein A (RPA)24,25. RPA is an abundant heterotrimeric nuclear complex (composed of 70, 32 and 14 kDa subunits), which is implicated in every DNA metabolic process involving single-stranded DNA25,30–32. Its role in homologous recombination is a complex one, whereby it serves to stimulate recombination by eliminating secondary structure in the single-strand tails25,32. However, an amount that is required to saturate the available single-stranded DNA (i.e. ~25 nucleotides per RPA molecule), also acts to suppress the nucleation of RAD51 filaments thereon24,32 – that are to form the centrepiece of this repair reaction.

This inhibitory effect is overcome by RAD51 cofactors, together dubbed mediator enzymes25,32. Recombination mediators share common properties in that they are all capable of interacting with their cognate recombinase, and preferentially bind single-stranded over double-stranded DNA32. In some cases, an interaction with RPA is also noticed32. A large body of evidence points to such a function for BRCA2 amongst humans. Specifically, BRCA2 binds single-stranded but not double-stranded DNA, physically interacts with the RAD51 recombinase, and is necessary for the formation of DNA damage-induced nuclear RAD51
foci on resected double-strand breaks. Only catalytic quantities are essential to negate RPA impediment, probably as the addition of recombinase molecules to a nascent filament on single-stranded DNA (i.e. filament growth), is adequate to displace RPA from DNA. Additionally, unlike its prokaryotic orthologue RecA, which exhibits a kinetic delay in binding double-stranded compared with single-stranded DNA; RAD51 displays only little specificity for the latter over the former. It is hereby again that BRCA2 enacts its part by differentially targeting RAD51 to single-stranded DNA (see Fig. 2iv).

Another BRCA2-associated protein is PALB2 or FANCN (Fanconi anaemia complementation group N), which promotes its proper localisation and stability in key nuclear structures, e.g. the chromatin and nuclear matrix. Stability is endowed in part by preventing BRCA2 from getting stranded in the nuclear soluble fraction, whereby it would be inherently less stable due partly to the effects of proteasome-mediated degradation. But compatible with the existence of at least two nuclear localisation signals within BRCA2, PALB2 is not needed for its entry into the nucleus.

RAD52, whose orthologue in yeast is the primary recombination mediator therein, seems only to have subsidiary role as such amongst vertebrates and humans. It appears to provide an independent and secondary pathway, albeit less efficient, for RAD51-effectuated homology-directed repair – especially in the absence of BRCA2. The simultaneous loss however, of RAD52 and BRCA2, results in synthetic lethality.

1.3.1.2.3 Synapsis

The formation of RAD51 upon single-stranded DNA nucleoprotein filament, called the ‘presynaptic filament’ – is the first rate-limiting stage in homologous recombination. This is a slow process as it is susceptible to disruption by the single-stranded DNA binding protein RPA, and so requires the intervention of mediator proteins such as BRCA2 for enhancement. However, once assembled it has the remarkable ability to capture a duplex DNA molecule and search for homology therein.

It is believed that this occurs via random collisions between the presynaptic filament and the duplex molecule. Segments of the duplex are bound and reiteratively tested until homology is located. Once found, the presynaptic filament forms initially a ‘paranemic’, and then a ‘plectonemic’ joint with the duplex DNA molecule. In the former, an internal region of the single-stranded DNA is paired with the duplex molecule through canonical Watson-Crick hydrogen bonds, but the two are not topologically intertwined. This entails predominantly the formation of A-T base pairs between the recombining DNA molecules. The triple-stranded, paranemically linked nucleoprotein intermediate is referred to as the ‘synaptic complex’. 
Though transient, it facilitates the free DNA end being held in homologous registry with the duplex, to institute the establishment of a plectonemic joint – whereby the participating DNA strands are bound by Watson-Crick hydrogen bonds, and are topologically linked\textsuperscript{32}. Thereby, the would-be non-template strand is extruded, and the invading single-stranded DNA forms a joint molecule with the complementary strand of the host undamaged DNA, leading to the generation of the so-termed displacement (D)-loop structure (see Fig. 2v). This emergent plectonemic joint/D-loop can be lengthened by the DNA strand exchange being catalyzed by the presynaptic filament\textsuperscript{32}.

Several accessory factors also assist this phase of the process. HOP2-MND1 complex affords strong amplification of the recombinase activity of RAD51 in a bipartite manner\textsuperscript{32}. Firstly, it stabilises the presynaptic filament; and secondly works with the latter to capture the duplex DNA molecule and aid the assembly of the synaptic complex\textsuperscript{32}. RAD54 also interacts with RAD51, and the two mutually augment each other’s enzymatic activities, including the facilitation of homology search, and promotion of strand exchange/D-loop formation\textsuperscript{32}. RAD54B, a parologue of RAD54, has similar biochemical activities as the latter. Expressly, RAD51 supplements the activity of RAD54B, and the latter stimulates D-loop formation by the former\textsuperscript{32}. There exist five RAD51 paralogues as well, namely RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3. They associate as two major functionally distinct complexes \textit{in vivo}, i.e. RAD51B-RAD51C-RAD51D-XRCC2 (BCDX2) and RAD51C-XRCC3 (CX3)\textsuperscript{36,37}. The former helps recruit RAD51 to damage sites and stabilise the presynaptic complex, whereas the latter exhibits strand exchange/homologous pairing activity\textsuperscript{36,38}. Both of these complexes are epistatic with BRCA\textsuperscript{2,36,37}, and synthetically lethal with RAD52 inactivation\textsuperscript{36}.

Cohesins, four-subunit complexes that entrap DNA fibres within their ring-shaped structures, promote use of the sister chromatid as template, preventing inaccurate repair that might lead to loss of heterozygosity by recombination between homologous chromosomes\textsuperscript{10,39,40} (Fig. 2i).
Figure 2. Double-stranded DNA break repair by homologous recombination. (i) The damage (red DNA) is recognised by MRE11-RAD50-NBS1 (MRN) complex. MRE11 dimers (orange) can promote short-range synopsis of the broken ends, while RAD50 (blue) helps tether the sister or homologous chromosomes together (black DNA) – see text, §1.3.1.2.1. Cohesins (green) may further promote interactions between sister chromatids, i.e. as opposed to homologous chromosomes\(^\text{10}\). (ii) MRN acts too as a signaliser, recruiting CtIP (yellow) that through its 5’→3’ exonuclease activity exposes 3’ ssDNA tails at either end of the break\(^\text{28}\). (iii) Further resection is performed by ExoI (pale purple) with BLM helicase\(^\text{28}\), and the resulting single-stranded DNA coated by RPA (purple)\(^\text{27}\). (iv) These RPA-bound 3’ overhangs provide a substrate for BRCA2 (pink), which laden with a cargo of RAD51 molecules (pale yellow) loads them onto the ssDNA. The nucleated RAD51 then self-assembles about the ssDNA into a filament, subsequently displacing both RPA and BRCA2\(^\text{27}\). BRCA2 also associates with PALB2 (pale blue), but this is not required for its recombination mediator activity in vitro\(^\text{33}\). (v) The RAD51-ssDNA complex is able to catalyse invasion of the ssDNA into the intact duplex DNA, to first form a triple-stranded nucleoprotein filament and search for homology therein\(^\text{24}\). (vi) Once located, the complementary strand is excluded to form a displacement (D-) loop and the template strand used to repair the damage via polymerase and ligase action, with theuptaken ssDNA strand serving as a primer\(^\text{11,24,32}\). RAD52, RAD51 paralogues, RAD54, RAD54B, and the HOP2-MND1 complex assist at various stages of the process\(^\text{32}\).
1.3.1.2.4 Synthesis and ligation

The capture of homologous duplex DNA and formation of the synaptic complex is thus the second rate-limiting stage of homologous recombination, and is overcome by the HOP2-MND1 complex\(^{32}\). Subsequent homology search and aligning of sequences, followed by strand exchange and D-loop formation, presents a 3’ end that is used as a primer for DNA polymerisation\(^41\). This results in transfer of information from the unbroken donor to the broken recipient, called ‘gene conversion’\(^{41}\). Vigorous DNA synthesis from the invading single-stranded DNA is carried out by DNA polymerase eta (POL\(\eta\)), which exhibits specific D-loop recognition\(^{42,43}\). RAD51 synergises with POL\(\eta\), and facilitates primer extension by the latter by elongation of the D-loop bubble\(^{42}\). Continuity of the strands is then established by DNA ligase I (LIG1)\(^{21,26}\) – see Fig. 2vi.

1.3.1.2.5 End-product resolution

Afterwards, the joint molecules so formed may proceed down one of two routes. They can either dissociate, with the broken chromosome being rejoined through synthesis-dependent strand annealing; or engage the other resected broken end, creating Holliday junctions to then be dissolved or resolved by helicases/topoisomerases and structure-specific endonucleases (i.e. resolvases) respectively\(^44\). The former involves essentially a migrating D-loop that never captures the second end. Rather, the invading strand gets displaced and anneals with this end\(^32\) (see Fig. 3a). Hereby, RAD54 due to its branch migration activity, causes dissociation of the D-loop\(^44\).

Alternatively, the second resected double-strand break end can be captured by annealing to the extended D-loop, generating cruciform DNA structures called Holliday junctions that link the recombining duplexes\(^32,45\). These have to be removed, as the cell would otherwise be unable to undergo division if the repaired pair of chromosomes remain intertwined in this topology, and cannot be separated to opposite poles of the mitotic spindle. In somatic cells, the primary mechanism is dissolution of the junctions via the BLM-TOPOIII\(\alpha\)-RM1-RM2 complex\(^45\) (Fig. 3b). This entails convergent migration of the two Holliday junctions promoted by BLM helicase, creating hemicatenane that is dissociated by topoisomerase action of TOPOIII\(\alpha\)\(^45\).

Both the above processes culminate in non-crossover recombination products, whereby an exchange of arms between participant chromosomes does not take place. This is indeed the preferred outcome of homology-directed repair\(^46\), as it prevents any loss of heterozygosity that may drive carcinogenesis.

However, Holliday junction resolvases provide backup for the two above pathways by cleaving persistent junctions\(^45\). Depending on whether cleavage occurs across the inside or
outside pair of strands, the result is a non-crossover or crossover product respectively (Fig. 3c). The two are similar or identical operations owing to four-way symmetry of these structures. Despite the potential for resolvases to induce genome-distabilising rearrangements, they serve to safeguard cell viability – as residual or unresolved Holliday junctions present a blockade to chromosome segregation during mitosis\textsuperscript{45}. Human cells devoid of resolvase components exhibit defects such as anaphase bridges, and high frequencies of lagging chromosomes and cell mortality\textsuperscript{45,47}. Such phenotypes likely arise from accumulation of unresolved Holliday junctions, giving rise to chromosomal nondisjunction at mitosis. Similarly, cells from BLM-mutated individuals display elevated levels of chromosome exchanges, deficiencies in their segregation and genomic instability\textsuperscript{45}. There exist three Holliday junction resolvases in humans, i.e. GEN1, SLX1-SLX4 and MUS81-EME1. These constitute two different pathways for resolution, one mediated by GEN1, and another by SLX1-SLX4-MUS81-EME1 (SLX-MUS)\textsuperscript{45}. To prevent undue crossovers, their activities are limited during S phase, due to regulation either temporally (in case of MUS81-EME1) or by nuclear exclusion (GEN1)\textsuperscript{48}. For instance, human cells promote an association between SLX1-SLX4 and MUS81-EME1 during prometaphase in order to stimulate Holliday junction resolution, as SLX-MUS is a more potent resolvase than either of its components\textsuperscript{45}.

Mitotic homologous recombination is most frequently associated with non-crossovers. However, in meiosis it effects exchange of information between maternal and paternal alleles within gamete precursor cells after deliberate initiation of double-strand breaks by SPO11, generating genetic diversity among progeny derived from common parents. DMC1, a meiosis-specific recombinase, is found in almost all eukaryotes including humans, and is active only during meiosis – where its acts in concert with RAD51. HOP2-MND1 complex and RAD54B function with DMC1 in the same way as RAD51. And likewise, BRCA2 also interacts with DMC1, whose ablation leads to meiotic recombination defects.\textsuperscript{32}

In general however, mitotic recombination factors are better characterised than meiosis-specific ones.
Figure 3. Resolution of recombinational repair products. The homologous recombination process results in the formation of DNA joints connecting the recombining DNA molecules that need then to be separated. This can be achieved through synthesis-dependent strand annealing (a), whereby the D-loop is unwound and the freed ssDNA strand anneals with the complementary ssDNA from the other side of the double-stranded break. The reaction is concluded via gap-filling DNA synthesis and ligation, and exclusively non-crossover products are formed. Or else, the second break end can be captured and result in the formation of Holliday junctions (b, c). They are preferably then dissolved by an outward migration of the junctions mediated by BLM-TOPIIα complex (b), which too leads solely to non-crossovers – that are indeed the preferred outcome of HR. However, should these fail, the HJs may be resolved via endonucleolytic cleavage of either the inner (black) or outer (grey) pair of bases (by MUS81-EME1 and/or SLX1-SLX4, or GEN1)45, resulting in a non-crossover (black arrow) and crossover (grey) respectively (c). Probability of the two is thence equally divided.
1.4 The RAD51 recombinase

The 339-amino acid (~37 kDa) human RAD51 protein is the eukaryotic homologue of the prokaryotic RecA, the archaeal RadA, and the meiosis-specific eukaryotic recombinase DMC1 – and is the central recombinase enzyme amongst eukaryotes. It polymerises as filaments onto DNA and promotes strand invasion and exchange between sister chromatids or homologous chromosomes.

1.4.1 Cell viability and chromosome stability

In vertebrate organisms, the loss of RAD51 function engenders cell inviability. For instance, the mouse Rad51−/− genotype is an embryonic lethal. In one experiment, mice heterozygous for the Rad51 null mutation were generated using gene targeted embryonic stem cells, whereby a small deletion had been introduced corresponding to a region carrying the ATP-binding motif of the protein that is essential for its recombinogenic functions. Nine Rad51+/− mice (6 males & 3 females) were intercrossed, and their offspring genotyped. Thereby, Rad51+/− and Rad51−/− mice were observed in the ratio 1:2 respectively. Tissue samples from heterozygous animals revealed one-half the amount of RAD51 protein compared to wild-type, but such mice themselves appeared normal and were fertile. On the other hand, of the (148) offspring examined no Rad51−/− homozygote was found, strongly indicating that the homozygous embryos suffer growth cessation during their development. Analysis of the embryos’ early stages showed that only a marginal fraction of the Rad51−/− genotype was viable, i.e. 1 out of 109 embryos at the 4- to 8-cell stage and 2 out of 65 at the blastocyst stage – exceedingly low ratios in light of Mendelian genetics. These rare cases of relative development of Rad51−/− embryos could have been being sustained by residual RAD51 of maternal origin. Furthermore, reduction in levels of RAD51 by antisense transfection led to severe growth retardation in the m5S mouse cell line.

In a second study, mouse Rad51 was similarly disrupted by embryonic stem cell gene targeting, in this case having introduced an out-of-frame deletion in the gene. Rad51−/− mutant embryos arrested in development early after implantation. The decreased cellular proliferation was followed by programmed cell death. Such cells accumulated severe chromosome loss by the end of their lives. They also exhibited extreme sensitivity to gamma radiation. Again, no Rad51 null mutant mice were observed.

In a later experiment, RAD51−/− clones of the chicken B cell DT40 line were created, carrying a human RAD51 transgene under the control of a tetracycline (tet)-repressible promoter. The tet (i.e. HsRAD51+) cells possessed proliferative properties indistinguishable from wild-type DT40 cells, and the two displayed much the same response pattern to gamma
irradiation. Subsequently, endogenous (non-induced) RAD51 repair foci were disrupted by an inhibition of the transgene’s expression, via addition of tetracycline to the medium. The cells ceased proliferating and arrested at the G2/M boundary, having accumulated large numbers of cytologically visible chromosome breaks – see Fig. 4, before dying. In contrast, chromosomal aberrations were scarcely detectable in both tet’ clones and wild-type cells.

Such studies highlight the vital role of RAD51 orthologues in repairing spontaneously occurring and exogenous chromosome breaks in proliferating higher eukaryotic cells, and hence their very viability.

Figure 4. Representative karyotype analysis of 4-Gy-irradiated Rad51-deficient (chicken) DT40 cells, after incubation in a tetracycline (tet’ ) medium for 21 hours (see text, §1.4.1). Macrochromosomes 1–5 and Z are indicated. Isochromatid gaps and breaks are identified by small and large arrows respectively. Reproduced from Sonoda et al. (EMBO J., 1998).
1.4.2 RAD51 orthologues

Human RAD51 (and mouse Rad51), specify proteins that show a high degree of amino-acid similarity to the yeast Rad51 protein (83%), DMC1 being less so (64%). They are all also homologous to the E. Coli RecA protein, specifically from residues 33 to 240 of the latter (see Fig. 5). This common region constitutes the ‘homologous core’ of these proteins, and harbours the ATP-binding and hydrolysis domains. As such, the mammalian and RecA proteins share 51% similarity between their homologous cores. The RAD51 homologues are ~90 amino acids shorter in the C-terminal region, and 63 to 121 residues longer at the N-terminus. These extended N-termini comprise ~95 amino-acid sequences that are common amongst the eukaryotic proteins. Human and mouse RAD51, differ in fact by only 4 amino acids (hence 99% identical); while chicken and human RAD51 proteins are 95% identical.

The level of homology between human, yeast and bacterial recombinases is suggestive of the marked similarity between their respective homologous recombination pathways. Yeast and mammalian RAD51-mutant cells exhibit similar phenotypes, i.e. both display loss of chromosomes, impaired proliferation, and hypersensitivity to DNA-damaging agents. However, E. Coli recA and yeast RAD51 mutations do not impart cell lethality. This reflects vertebrate genomes being hundreds of times larger than yeast’s, incurring double-strand breaks as much more frequently – and thus a proportionally greater requirement for their repair. The sensitivity however is the same when the two’s relative genomic sizes are accounted for.

Analysis of DNA-PKcs and KU80-deficient cells reveals that they do have similarities to mouse RAD51-deficient cells. All are highly sensitive to ionising radiation, and KU80-deficient cells are prematurely senescent. Ku80−/− mice show growth retardation, and adults are 40-60% the normal size. But nonetheless KU80 and DNA-PKcs-deficient mice and cells are both viable, whereas MmRAD51-deficient cells are not. These observations illustrate that the consequences of removing the homologous recombination pathway for repair of DNA double-strand breaks, are much greater than removal of the nonhomologous pathway.
Figure 5. RAD51 protein sequence and orthologues (human and mouse RAD51, ScRad51 and E. Coli RecA). All recombinases share a core ATPase domain. The eukaryotic proteins share a common ~95 a.a. N-terminal region, while RecA has a ~90 a.a. C-terminal extension. Each of these binds DNA. ScRad51 harbour’s a further 55 a.a. N-terminal extension that is unique to yeast. Homologous domains are like-coloured. Type A (ATP binding) and Type B (hydrolysis) Walker motifs are indicated. The depicted locations and sizes of the sections are accurate and exactly to scale.

1.4.3 RAD51 filament structure

All recombinases can occur as either rings or right-handed nucleoprotein filaments. Rings form most often in the absence of DNA, and usually also nucleotide-cofactors. They may be hexameric (RecA), heptameric (RAD51) or octameric (DMC1). For example, RadA (the closely related archaeal homologue of Rad51), forms heptameric rings with the ATP-binding sites occluded. As such, rings constitute an inactive state of these proteins, but might still be physiologically relevant as ‘storage forms’ that prevent futile cycling of ATP. A flexible segment connecting the N-terminal extension of Rad51 and its ATPase core, comprises a single β-strand (i.e. β0) that serves as the primary interaction motif in polymeric assemblies – by packing onto the edge (β1) of the neighbouring protomer’s central β-sheet, and forming a β-zipper (see Fig. 6b). This interdomain linker protrudes from the rest of the molecule and provides the flexibility necessary for changes in quaternary structure. Analogously, in the RecA filament, such an intermolecular β-sheet is formed by β1 from one subunit contacting β3 from the adjacent monomer. Recombination-competency follows from the filamentous form, and these filaments exhibit marked conformational variability; in their helical pitch (sometimes even within different regions of the same filament), and extension of the enclosed DNA – in response to the type of cofactor and DNA bound. However, inactive filaments (those formed with ADP or without DNA) generally display lower pitches (~65–85 Å), than active ones (~90–130 Å).

The crystal structure of a ScRad51 filament formed in the presence of single-stranded DNA and ATPγS, revealed the nucleotide-binding pocket of one protomer (Walker A) to be in direct contact with the ATPase domain of the next (Walker B), and a pitch of 130 Å (Fig. 6a, c) – compatible with previously obtained electron microscopy structures of active
forms of recombinase filaments. An excess 79-residue N-terminal extension unique to yeast had been truncated, corresponding to the point where the NMR structure of human RAD51’s N-terminus becomes well-ordered. This truncated variant was nonetheless fully capable of DNA binding and strand exchange in vitro. Another important feature of the structure is the pseudo six-fold helical (P61) symmetry, the true symmetry being only three-fold (P31) – as alternating subunit-subunit interfaces are slightly different, due to slight rigid body movements between the N- and C-terminal (ATPase) domains. Inter-subunit interactions between ATPase domains thus alternate between two arrangements, which vary substantially at the nucleotide-binding pocket. In one configuration, the His352-bearing helix from the ATPase site of the adjacent monomer lies close to the phosphate-binding loop of the first, whereas in the other it is occluded by Phe187 of the first and shifted away from this pocket (Fig. 6c). Such a scheme also happens to be consistent with the observation that RecA-DNA filaments in solution show two different affinities for ATP. While it is possible that the histidine, which is highly conserved across all eukaryotic recombinases, directly catalyses ATP hydrolysis; it likely plays a part in allosteric regulation of the system, by sensing nucleotide binding and coupling hydrolysis with the filament’s overall conformation and binding to DNA. Stabilisation of this helix might promote the active state of the filament, by incorporating it as part of the protein-protein interface. The helix is disordered in non-filamentous forms of RadA and HsRAD51, implying it is fairly flexible, and that its localisation is dependent on interactions in the filament structure. In support of His352’s contribution to filament assembly, the H352A mutant is significantly defected in single-stranded DNA binding.

NMR and mutational analysis of human RAD51’s N-terminus, HsRAD51(1-114), established its role in single- and double-stranded DNA binding, specifically residues 61–69 that constitute a positively charged surface. The N-termini fold into compact all-helical domains, which form lobes lining the upper part of the filament, and point into its helical groove – with amino acids 61–69 located at the top of these lobes (as seen in the NMR structure fitted into an electron microscopy reconstruction of human RAD51 filament on single-stranded DNA).

DNA is believed to be bound in a relatively stretched (often inhomogeneously) and untwisted manner within recombinase filaments. And despite its presence, single-stranded DNA itself was not visualised in the aforementioned crystal structure, even though it mimics the DNA-bound conformation. This is likely due to the DNA being too poorly ordered to be seen, for instance because different protomers could be binding DNA at different stoichiometries. Study of DNA binding by ScRad51, in presence of ATP and Mg2+, indicated that its binds both single- and double-stranded DNA over a wide range of pH at 4 ± 1 nucleotides.
(or base pairs) per monomer. An average value of 4 nt per protomer, corresponds to an extension of ~120 Å per turn of DNA helix – and is hence consistent with the observed structure. It remains possible nevertheless, that DNA was not bound in the filament, or was digested by trace nuclease contamination. ATPγS was probably hydrolysed during the crystals’ growth, and clear density for only one large ion was seen at the nucleotide-binding site – most likely a sulphate (present in the medium) mimicking the binding of phosphate. There is ample space however, for an ATP molecule to be modelled into the phosphate-binding loop with just minor adjustments to the nearby structure.

Synthesis of the above results is clearly implicative of an allosteric communication between the ATPase- and DNA-binding sites to affect nucleoprotein filament structure and DNA strand exchange. It should be noted though, that while ATP-binding is necessary for the recombination function of Rad51 (as well as RadA and RecA), hydrolysis is required only by Dmc1.
Figure 6. Crystal structure of an active (ATPγS and ssDNA bound) configuration of the ScRad51 filament49, with an N-terminal truncation corresponding to the HsRAD51 N-terminus. (a) The filament has a pitch of 130 Å, and shows alternating monomer-monomer interfaces to be slightly different (yellow and green), such that the filament possesses pseudo 6-fold but only an exact 3-fold screw symmetry. A sulphate (black balls) can be seen mimicking the binding of phosphate in the ATPase site, which is situated directly at the interface between neighbouring subunits (arrow). N-terminal domains line the upper segment of the filament (circled). ATPγS and ssDNA themselves are not visualised due to being hydrolysed and disordered respectively. (b) The Rad51 filament polymerises via stacking of a β-strand from the linker between the N-terminal and ATPase domains of one monomer (green), against the central sheet of the adjacent monomer (yellow). Interactions between the ATPase domains comprise the His352-bearing helix, which is disordered in nonfilamentous forms, but is stabilised thereby. The histidine itself may or may not be involved in ATP hydrolysis. (c) In one variant of the intersubunit interface, His352 from one monomer is positioned directly over the ATPase site of the preceding monomer (left). In the second form, the His352 is shifted away from the ATPase site, and occluded therefrom by Phe187 of the previous monomer (right). Adapted from Conway et al. (Nat. Struct. Mol. Biol., 2004).

1.5 The BRCA2 tumour suppressor

Bacterial cells respond in an exceptionally affirmative fashion to the event of a double-stranded break in DNA, to the extent where within 30 minutes of its occurrence, RecA constitutes by far the most abundant protein in the cell. This luxury is not afforded to human and mammalian cells with their considerably larger genomes, which rely instead on harnessing the recombination protein that is present already in their nuclei, and relocalising it to the sites of double-stranded break damage where the repair reactions then take place – the so-called repair foci.65

BRCA2 orthologues constitute the primary recombination mediator in higher eukaryotes, providing a rapid mechanism for RAD51 loading that has seemingly relegated RAD54 to a more minor role therein over the course of evolution66. These orthologues vary greatly in size, ranging from the 3,418 a.a. (384.073 kDa) human protein – see Fig. 7, to the 383-residue Caenorhabditis elegans counterpart called BRC-232.

Human BRCA2 harbours a succession of eight repeated sequences of ~35 a.a. in length termed BRC motifs with variably sized linker regions, interspersed within a ~1100 residue roughly middle portion of the protein (a.a. 1009 – 2083) – and confer its ability to interact with RAD5167. Their number is highly conserved between mammalian species67, but greatly
diverse in non-mammals, e.g. *Arabidopsis thaliana* Brca2 and *C. elegans* BRC-2 possess four BRC repeats and one such repeat respectively.\(^\text{32}\)

The ~700 residue DNA-binding domain is located in the C-terminal region of BRCA2, i.e. amino acids 2,482 – 3,184, and is comprised of five subdomains. These include an alpha helical domain, followed by three oligonucleotide/oligosachharide-binding (OB) folds, with an insertion of a ‘tower’ subdomain in the second OB fold. Situated between the helical and first OB fold lies a pocket that receives a small highly acidic polypeptide called DSS1 (regarding which more later).\(^\text{68}\)

BRCA2 also harbours an additional 36 a.a. C-terminal RAD51 binding (CTRB) domain between residues 3,270 and 3,305 that is unrelated to the BRC repeats\(^\text{69}\) (see §1.5.1). Unlike the latter however, it interacts solely with RAD51 polymers and appears to stabilise nucleo-protein filaments already established on single-stranded DNA, via preventing their dissociation by the BRC repeats through their binding of individual monomers\(^\text{69}\). Phenotype of mice homozygous for a mutation in the CTRB domain is characterised by increased cancer incidence, reduce viability, and hypersensitivity to DNA crosslinking agents, but through a relatively moderate impairment of homology-directed repair\(^\text{32,69}\). Therefore, while it makes a significant contribution toward the mediator function of BRCA2, it appears to be comparatively dispensable vis-à-vis the BRC motifs\(^\text{32}\). However, evidence shows Ser3291 therein is phosphorylated by cyclin-dependent kinases (CDKs) at the G\(_2\)-M transition that eliminates the CTRB domain’s interaction with RAD51, and dephosphorylated upon DNA damage\(^\text{70}\). It has thus been suggested it provides a regulatory ‘switch’ that can either promote or suppress RAD51-mediated repair of DNA double-strand breaks, by disfavouring or favouring the disruption of RAD51-ssDNA filaments by BRC repeats respectively\(^\text{69}\). DMC1 too can bind this domain, albeit with weaker affinity\(^\text{32}\). It is though only found among vertebrate species\(^\text{69}\).

None of the BRC repeats however, appears to have affinity for DMC1. In fact, the main DMC1-interacting site has been mapped to a 26 a.a. region (residues 2386–2411) called the PhePP motif, which is highly conserved amongst mammalian orthologues.\(^\text{32}\)

Toward its C-terminus, BRCA2 harbours two nuclear localisation signal (NLS) sequences, with one flanking either side of the CTRB domain. The first is constituted by 3266-Lys-Lys-Arg-Arg-3269, and the other by 3313-Lys-Lys-Lys-3315.\(^\text{70}\) In fact, the smallest known cancer-associated deletion in *BRCA2* (that is also the most C-terminal of such mutations), substitutes a stop codon for Thr3195 (T3195X), thus ablating the final 7% (224 amino acids) of BRCA2 and rendering it cytoplasmic. The Lys3326Ter polymorphism however, eliminates 92 residues from its C-terminus without consequence.\(^\text{71}\)
The N-terminal region of BRCA2 binds PALB2, through the former’s amino acids 21 via 39. This contains a minimal ~10 residue motif (a.a. 26–36) that is structured in the complex and includes a short α-helix. PALB2, a 1186 residue protein, interacts with BRCA2 through the former’s C-terminal WD40 domain (a.a. ~836–1186), a sequence of repeated motifs commonly involved in protein-protein interactions that folds into a circularised β-propeller. In this case, it assumes a seven-bladed structure, and BRCA2(21–39) binds into a pocket constituted by the tips of the fourth and fifth blades. The interaction buries ~1,100 Å² of surface area, typical of transient regulatory bindings.\(^{72}\)

**Figure 7.** BRCA2 protein sequence, and major binding partners. The N-terminal region harbours the PALB2-binding domain (a.a. 21–39)\(^2\). The middle portion (a.a. 1,009–2,083) contains eight 35 a.a. BRC repeats that interact with RAD51\(^5\). After is a DMC1-binding PhePP motif (a.a. 2,386–2,411)\(^32\). The C-terminal region comprises the DNA binding domain (a.a. 2,482 – 3,184)\(^68\). This in turn is composed of five subdomains, the alpha helical domain (αD), and three oligonucleotide/oligosaccharide-binding folds (OB1, OB2 and OB3); an additional tower (T) domain constitutes an insertion in OB2. A small area between αD and OB1 also binds DSS1. Toward the C-terminus lies an independent RAD51-binding (CTRB) domain (a.a. 3,270–3,270)\(^69\). This is closely flanked by a nuclear localisation sequence (NLS) on either side (a.a. 3,266–3,269 and 3,313–3,315)\(^70\). The relative positions, lengths, and spacings of the various domains are exact and precisely to scale.

### 1.5.1 BRCA2 and Cell Proliferation

In an early study\(^73\), the Brca2 locus in mice was mutated using embryonic stem cell technology, in order to delete amino acids 626–1,437. Southern blotting identified clones carrying the mutant gene, and the cell lines were injected into blastocysts. Chimaeric mice born mostly transmitted the Brca2\(^-\) allele to their progeny. Heterozygous mice were healthy and fertile (and did not show signs of tumourigenesis until at least 8 months of age). Litters from Brca2\(^+/\) intercross matings did not yield any Brca2\(^-\) offspring, suggesting Brca2 is crucial for embryonic development.

Embryos from varying stages of development were histologically examined to determine the time and cause of their lethality. All (16) embryos were normal at 6½ days of develop-
ment. However, by 7½ days, ~25% of the (45) embryos exhibited a mutant phenotype and overall growth arrest. Given they resembled embryos from the previous day, implies the developmental block occurred after 6½ days of gestation. An inability to derive homozygous mutant embryonic stem cell lines, further suggested that the Brca2 gene is requisite for embryonic cells’ survival and proliferation. Since Brca2 null mutant embryos do initiate formation of morphological features, it is implicit that the development arrest is not due to a differentiation defect but lack of proliferation. This occurs at the time when the embryo undergoes its most rapid phase of cell division.

Appearance of the mutant phenotype in Brca2-deficient embryos was tallied with the gene’s expression via in situ hybridisation. No transcripts are detectable at 6½ days. But at 7½ days, Brca2 is upregulated and transcribed throughout the embryo. These transcripts are also particularly prevalent in tissues with high mitotic indices, and downregulated in post-division cells. Using a yeast two-hybrid screen, MmRad51 was shown to interact with Brca2 (in this case a C-terminal region comprising residues 3,138–3,232). The interaction was further affirmed in mammalian cells. A minimal 36-residue segment (a.a. 3,196–3,232) was found to display a strong association, of which removing 13 amino acids (residues 3,196–3,232) eliminated the interaction. 72, 95 and 100% identity exists between the 94, 36 and 13 amino-acid fragments therein and the human homologue respectively, suggestive of their functional significance. Expression patterns of Brca2 and MmRad51 were also nearly identical, and homozygous mutants of the two genes arrest growth at similar stages. This region then came to be regarded as the CTRB domain.

Brca2−/− and control (Brca2+/+ and Brca2+/−) embryos at day 3½ were gauged for γ-radiation sensitivity. Following administration of 400 Rads, inner cell mass of the latter was minimally affected, but completely abolished in the former. Furthermore, while trophoblasts in the controls were only marginally reduced, there was a significant decrease amongst Brca2−/− embryos. MmRad51 null embryos however, appear more sensitive to γ-irradiation (based on the response of trophoblast cells). Such assays indicate that replicative growth arrest and death are in fact a secondary consequence of a defect in DNA damage repair.

Above conclusions thus also help explain the high penetrance of early onset breast cancers amongst human carriers of a BRCA2 mutation. In mammary epithelial cells having lost BRCA2 function, RAD51-effectuated double-strand break repair is presumably compromised, which destabilises the genome. Some RAD51 activity continues, but ultimately BRCA2’s mediator function is found wanting. Higher severity of MmRad51 compared to Brca2 mutant phenotype (though very similar), too is consistent with this scenario.
But while MmBrca2 deficiency induces embryonic lethality, BRCA2-deficient breast epithelial cells proliferate and become tumours. One resolution of this apparent paradox is that developmental arrest is cell-type specific. Embryonic cells are hypersensitive to unrepaird double-strand breaks and hence inviable. Mammary epithelial cells on the other hand can survive with an unstable genome (perhaps due to a hormonal response\(^7\)), but accumulate genome alterations that contribute to malignancy. Mouse mutants of tumour suppressor genes also differ in the extent to which their phenotypes resemble the human syndromes.

### 1.5.2 BRCA2 and genomic stability

A mouse strain with a targeted mutation in exon 11 of Brca2 was used to probe its role in DNA repair, so-designed to model a mutation linked with ovarian cancer\(^7\). This was predicted to diminish the encoded protein to half its size at amino-acid residue 1492. Homozygous mutant cells expressed transcripts from the intact 5’ region of the gene and a partial protein was also synthesised. Hereby, the Brca2\(^{-}\) allele endowed a stunted development phenotype, potentially accordant with an incomplete Brca2 enzyme. Though the Brca2\(^{-}\) genotype caused significant growth retardation, such embryos were often viable till birth. Some even survived into adulthood, but succumbed to thymic lymphomas. The exceptional embryonic survivability of this strain was exploited to establish a Brca2\(^{-}\) cell line.

These were then used as a system to test for cellular and genome abnormalities that would delineate Brca2’s role in the DNA damage response. Cells in culture exhibited a proliferation impediment that worsened with successive passages of the cell cycle. Elevated arrest at the G1 and G2/M phases and an impaired ability to enter S phase, was accompanied by an increased expression of p21 and the pro-apoptotic p53 protein. Normal cells too arrest at the two checkpoints upon \(\gamma\)-irradiation. Whereas the p53- and p21-dependent G1 checkpoint prevents replication of damaged templates, the G2 checkpoint blocks segregation of damaged chromosomes. Which of the apoptotic or cell-cycle arrest responses predominates depends on the cell-type, e.g. lymphoid progenitor cells typically initiate rapid apoptosis. Such cells were extracted from foetal livers of Brca2\(^{-}\), Brca2\(^{+}\) and Brca2\(^{+/+}\) embryos. Homozygous mutant cells showed a similar apoptotic response to the controls when exposed to \(\gamma\)-radiation, even down to 10 rads. Apoptosis and checkpoint activation were thence intact. Thus Brca2’s involvement is indicated to be in DNA repair, independent (or at least distinct) from cell cycle control and apoptotic aspects of the damage response.

Mutant cells also exhibited increased sensitivity to various genotoxins, especially ionizing radiation and methyl-methanesulphonate (MMS), which induce DNA double-stranded breaks and interstrand crosslinks respectively. This suggested that Brca2’s function is in particular relevant to the repair of these lesions.
Moreover, Brca2-deficient cells displayed an excessive occurrence of spontaneous (uninduced) chromosomal aberrations. They included chromatid and chromosome breaks, and frequently tri- and quadriradial chromosomes. The latter are reflective of chromatid exchanges that take place in mitotic recombination during cell division. Apart from structural abnormalities, aberrations in chromosome number (polyploidy) was observed as well. There was a significant increase in the number of cells with greater than four times the DNA content of controls. As such these data are further evidential of Brca2’s specific function in DNA double-strand break repair.

Again, the divergent phenotypes of neoplastic transformation in Brca2−/− mice hereby, and proliferative arrest beggars explanation. Reconciliation ensues from a model whereby secondary mutations (including in proto-oncogenes) that grant evasion from cell cycle checkpoints constitute an important precondition for tumourigenesis. Hence, the very spontaneous genetic instability rendered in cells lacking functional Brca2, which is the cause of reduced proliferation – might also precipitate these secondary changes.

A second study sought to clarify the nature of genome aberrations in such Brca2−/− cells. The effect of Brca2 truncation on spontaneous gross chromosomal rearrangements was explored. Integrity of the chromosomes’ structure in Brca2-deficient cells was examined by the means of spectral karyotyping. Metaphase spreads derived from Brca2−/− lymphocytes (following 20 passages in culture) showed sundry structural anomalies amongst the chromosomes, including multiple translocations between two different (non-homologous) chromosomes and even acentric fragments with parts from three – see Fig. 8. These kinds of aberrancies usually arise from mutagenic processing of DNA double-strand breaks, whereby several broken DNA ends bearing little or no interhomology are indiscriminately ligated together. All (12) Brca2−/− metaphases surveyed exhibited a similar pattern of aberrant chromosomes, which were not observed in controls. Also, their preponderantly random and inconsistent nature between mutant cells indicated that the underlying DNA damage took place continuously.

Tumourigenic mutations in humans are frequently accrued as a consequence of gross chromosomal rearrangements, comprising translocations (genetic exchange between non-homologous chromosomes), as well as inversions, large deletions, and partial or complete loss of chromosomes. Results hereby indicate that BRCA2 inactivation amplifies the occurrence of such rearrangements, triggering the additional genetic changes necessary for neoplastic transformation by loss of heterozygosity at functionally important genomic loci. This homozygosity then underpins the loss of tumour suppressor genes and/or activation of proto-oncogenes that is characteristic of cancer.
Brca2 hence suppresses tumour formation by the repair of spontaneous (and induced) double-strand breaks, and prevention of gross chromosomal rearrangements and loss of heterozygosity.

**Figure 8.** Spectral karyotyping analysis of chromosomes from Brca2<sup>−/−</sup> lymphocyte cells after 20 passages in culture reveals a high frequency of spontaneous random aberrations. A representative metaphase spread shows a dislocated chromosome fragment (panel 4), myriad translocations between two chromosomes as in t(X;16), t(12;16), t(13;14), and t(X;16), rearrangement amongst three chromosomes such as 13, 16 and 5 (panel 5), and potentially even more. Reproduced from Yu et al. (Genes Dev., 2000).

### 1.5.3 Colocalisation of BRCA2 and RAD51

In reaction to DNA double-stranded breaks, the RAD51 protein relocates within the nucleus into distinct repair foci, which are visualisable by confocal microscopy<sup>65</sup>. Formation of RAD51 foci is dependent upon BRCA2, as well as the RAD51 paralogues, RAD52, RAD54, and the HOP2-MND1 complex. They are accompanied also by RPA. Localisation kinetics vary between components, indicating that their assembly is a carefully choreographed process, involving the sequential recruitment of one protein based upon another<sup>65</sup>. These interactions moderate RAD51’s ability to bind DNA.
Given that Rad51 and Brca2 interact in mice, one might expect that the two be found at the same sites upon induction of DNA double-strand breaks.

In an experiment\(^6\), an asynchronous culture of HeLa cells was analysed for the presence of RAD51 by immunofluorescent staining. After treatment with 10 Gy of ionising radiation and being allowed to settle for 3 hours, the number of cells exhibiting RAD51 foci increased manifold. Similarly, when irradiated cells (with a recovery period of one hour) were assessed for the presence of RAD51 and BRCA2 foci, the two were found to exactly coincide (see Fig. 9).

RAD51 likewise colocalised with RAD54, and RPA\(^6\). Accumulation of recombination mediator proteins such as BRCA2 supports the notion that these foci represent the sites of double-strand break repair by homologous recombination. The latter signals the presence of DNA intermediates (in this case undergoing repair).

Collectively, the above findings signify that BRCA2 directs the activity of RAD51 not only at molecular scale, but also at the microscopic (focus formation) level.

*Figure 9. BRCA2 (green) and RAD51 (red) foci in HeLa cells were visualised through confocal microscopy 1 hour after exposure to 10 Gy of ionising radiation using rabbit anti-BRCA2 and mouse anti-RAD51 antibodies and immunofluorescence staining, and observed to precisely colocalise (denoted by yellow foci in the merged images). A similar pattern was observed with RAD54, and partially so with RPA. Reproduced from Tarsounas et al. (Phil. Trans. R. Soc. B, 2004).*

### 1.5.4 Structure of BRC4 – RAD51 (ATPase domain)

One of the two major (partial) x-ray crystal structures of BRCA2 determined to date, is that of the fourth BRC repeat (BRC4) in complex with the ATPase core domain of RAD51\(^7\).

The structure depicts the BRC repeat binding RAD51 via mimicry of the latter’s polymerisation motif\(^8\) (see Fig. 10a, b). BRC4 remains in contact with RAD51 over an unbroken...
stretch of 28 residues (Leu1521 – Glue1548). Nine amino acids (Phe1524 – Val1532) form a β-hairpin that stacks up against β-strand 3 (β3) of RAD51, furthering the latter’s central β-sheet by an additional two short antiparallel strands – and ‘zipping’ the two molecules together akin to interprotomer contacts in Rad51 filament’s crystal structure (§1.4.3, Fig. 6).

For illustration, the sequence of human RAD51’s oligomerisation motif, 85-GFTTATE-91, bears high similarity to the BRC consensus, GFXTASG (where X denotes a generic amino acid).

Following the hairpin, the BRC motif folds around helix 4 (α4) of RAD51 via a small linker (Lys1533 – Ala1535), which carries into an amphipathic α-helix (Lys1536 – Val1542). Remainder of the residues at the C-terminal end (Lys1543 – Glu1548) make an irregular coil, with parts of a 310-helix traversing helices α4 and α5 of RAD51 (Fig. 10c).

The RAD51-BRC4 complex buries 2,026 Å² in surface area, comprising mostly of hydrophobic interactions (Fig. 10d, e), but also some polar contacts. Three main hydrophobic points of contact keep the BRC motif in close proximity to RAD51, i.e. Phe1524, Ala1527, and Leu1545/Phe1546 (further abbreviated henceforth). F1524 lies in the β-sheet contacting RAD51, its aromatic ring situated deep within a hydrophobic pocket of the latter (constituted by the side chains of M158, I160, A190, A192, L203, and A207). Similarly, A1527’s β-carbon is placed inside a small pocket in RAD51, formed by the side chains of F166, P168, L171, L186, and V189. L1545 and P1546 assume a wedge lodged between the helices α4 and α5 of RAD51; surrounded by residues L204, Y205 and S208 from the former, and R254, L255, G258 and F259 of the latter. Binding between the two molecules is further strengthened by hydrophobic contacts between Ile1534 in the linker region of BRC4, and the hydrogen-bonded Ser1538, Leu1539 and Val 1542 in the α-helix.

A few polar interactions also exist (Fig. 10f). Part of the interaction that keeps BRC4 lined up alongside β3 of RAD51, is maintained by the β-hairpin via a set of three successive antiparallel backbone-to-backbone hydrogen bonds, linking the sequence 1525-HTA-1527 of the former to 189-VAY-191 of the latter. RAD51’s D187 receives a hydrogen bond from S1528 of BRC4 (the third location in the hairpin loop), and also interacts electrostatically with K1530 therein. E213 of RAD51 is hydrogen-bonded to BRC4’s S1538, the main-chain nitrogen of A1535, and the carbonyl group of K1533 (through a water molecule). The S1538 is kept poised for interaction by A1535 of BRC4 and RAD51’s V212. Lastly, R250 from RAD51 is ionically paired with G1548 towards the C-terminus of BRC4.

Six out of eight BRC motifs in BRCA2 when expressed in vitro could interact directly with recombinant RAD51. While BRC5 and 6 were devoid of binding, BRC3 and 4 were particularly efficient. This is explained by interactions involving residues not conserved
across BRC repeats, granting enhanced or diminished stability to the binding. For instance, L1521 and L1522 of BRC4 form hydrophobic contacts with the side chains of F195 and H199 from RAD51, and the backbone carbon of L1522 further accepts a hydrogen bond from H199. Also, H1525 forms a quasihydrophobic pocket through packing against the aliphatic regions of T1520 and L1531. This lends extra stability to the β-hairpin loop, and accounts for BRC4’s higher affinity for RAD51. Disparity amongst the BRC repeats’ affinities for RAD51, has potentially important implications for the latter’s regulation by BRCA2 (see §1.5.5).

Curiously, and in apparent contradiction with BRCA2’s function as a recombination mediator, RAD51 nucleoprotein filament formation is suppressed when incubated with BRC peptides in vitro. Similarly, RAD51 focus formation is suppressed upon overexpression of BRC repeats in vivo. It so appears BRCA2 prevents self-association of RAD51 monomers into inactive filaments without bound DNA or rings (by sequestering and targeting it to single-stranded DNA). BRCA2 could hence constitute as ‘cage’, whereupon by interacting with RAD51 through multiple BRC motifs, it orders their spatial distribution such that they are optimally primed for loading onto DNA. 

Owing to a propensity for RAD51 to assume various self-assemblies, BRC4 had been covalently linked to RAD51 and purified as a fusion protein.

Amongst carcinogenic BRCA2 mutations, are multiple deleterious point mutations affecting BRC repeats, which the structure helps explain. Presumably, these disrupt some of the aforementioned interactions, and weaken RAD51 binding. Thereby, function of the remaining seven repeats seems unable to rescue the phenotype. This could be because the 8 BRC repeats work synergistically as a RAD51-binding module, whose overall topology is pivotal for its function. Changes that lessen the affinity of any of the repeats for RAD51, could hence interfere with BRCA2’s activity by perturbing the spatial distribution of recombinase molecules bound within it. 

For instance, the T1526A mutation weakens BRC4’s affinity for RAD51. Cancer-causing mutations affecting equivalent positions in other BRC repeats are also found, e.g. T1011R in BRC1, S1221P (BRC2), and T1980I in BRC7. It was evident from the structure that T1526 receives a hydrogen bond from the backbone nitrogen of K1530, and is important for the correct and stable configuration of the β-hairpin loop. It also donates a hydrogen bond to S1528’s hydroxyl function, keeping it poised for interaction with D187 of RAD51. Hence, by disrupting the above interactions, a mutation in the threonine (or equivalent) would compromise the binding between BRC repeats and RAD51.
Another common breast cancer-linked mutation alters Gly1529 (fourth residue in the hairpin), to arginine. This is likely also to adversely impact its conformation and diminish RAD51-binding.

Notably, both the above mutations occur in the β-hairpin region, the most critical fold for BRC motifs’ association with RAD51.
Figure 10. Crystal structure of BRC4 in complex with RecA-homology/ATPase domain of RAD51\(^9\). (a) BRC4 peptide (green) binds RAD51 (magenta helices, blue sheets, and yellow N-terminus) by mimicking the \(\beta\)-strand polymerisation motif of the latter (c.f. §1.4.3 and Fig. 5b). This entails an interaction of the BRCA2 sequence 1524-FHTA-1527 with RAD51 through antiparallel \(\beta\)-strand pairing, as seen in a close-up of the interface (b). Phe1524 and Ala1527 therein mediate hydrophobic contacts with RAD51. (c) A topological description of the complex, in the same colour scheme as previously. Detailed view of hydrophobic interactions involving the \(\alpha\)-helical region (d), and the \(\beta\)-hairpin region (e), in a coil (green BRCA2) upon surface (grey RAD51) format. Only interacting side chains are depicted. (e) Polar interactions. RAD51 in cartoon representation, and BRC4 as a coil (green). Relevant side chains shown in stick mode. Dotted lines denote hydrogen bonds. Implemented in MacPyMOL (and Inkscape) using PDB-1N0W.

1.5.5 Two classes of BRC repeats

Using a semiquantitative yeast two-hybrid assay, a fragment of BRCA2 containing BRC1–4 was found to bind RAD51 with an affinity comparable to a segment comprising all BRC repeats\(^8\). Similarly, BRC4’s interaction was shown to be 3\(\times\) stronger than that of BRC1\(^8\).

Recently, quantification of binding affinity of each BRC motif toward RAD51 and RAD51-ssDNA (using GST-tagged BRC repeats and purified RAD51), revealed two evolutionary distinct classes therein\(^6\). Repeats 1–4 showed strongest interaction with RAD51 in accordance with previous experiments and bound in an approximately 1:1 stoichiometry, whereas 1–5 bound very weakly. And while BRC 1–4 blocked ATP hydrolysis by RAD51 (that causes dissociation of filaments), BRC5–8 did not have this effect. However, the latter exhibited high affinity toward RAD51-ssDNA complexes. Furthermore, BRC repeats 1–4 prevented RAD51-dsDNA complex formation, but 5–8 did not harbour this activity.

Such data are implicative of a scenario whereby BRCA2 binds to free-form RAD51 through BRC1–4, with the high-affinity binding mode of this group catalysing the rate-limiting nucleation step, i.e. by delivering RAD51 to single-stranded DNA, locally suppressing ATPase activity, and preventing futile association with dsDNA. The second group of BRC5–8 could then bind RAD51 molecules emerging as filaments from the BRCA2-RAD51 nucleus on ssDNA, stabilising the RAD51-ssDNA association and guiding nascent filament formation. Once the filament begins to extend beyond the footprint of BRCA2, the latter would dissociate having facilitated both filament nucleation and growth.
Thereby, the two categories of repeats appear to cooperate in a complementary and reinforcing (but only partially redundant) manner, in order to maximise the probability of presynaptic filament formation.

1.5.6 Structure of BRCA2DBD – DSS1 – ssDNA

Another important structure is that of the BRCA2 DNA/DSS1-binding domain (DBD), in complex with DSS1 (deleted in split hand/split foot protein 1), and single-stranded DNA.68

Following BRC repeats in BRCA2, is a ~1000 residue C-terminal region that harbours the DNA/DSS1-binding domain, and is its most conserved segment amongst vertebrates (68% average identity versus 42% for the whole protein), as well as putative orthologues in lower eukaryotes. It is the site of 27% of tumour-derived missense mutations in breast cancer.82 DSS1 is a 70-residue acidic polypeptide, which is found associated with virtually all BRCA2 in human cell lines, and is seemingly important for its stability.83

The BRCA2 DBD consists of five sub-domains (see Fig. 11a). The first is made of 190 amino acids and comprises mostly α-helices (helical domain, αD). After are three structurally homologous domains that possess the ~110 residue oligonucleotide/oligosaccharide-binding (OB) fold, i.e. OB1, OB2, and OB3. In the crystal, these are arranged linearly, packed together into an extended 115 Å long structure. The remaining domain exists as a 130 amino-acid insertion in the OB2 fold, and takes the form of a pair of long antiparallel helices (the stem), with a three-helix bundle (3HB) at the end, which harbours a helix-turn-helix (HTH) motif. This structure, which protrudes away from OB2 by 80 Å, is termed the ‘tower domain’.

BRCA2 OB domains have structures very similar to the canonical OB fold84 (Fig. 11b), i.e. a five-stranded β-sheet (β1 – β5) coiled in on itself to form a β-barrel, and capped by an α-helix (between β1 and β5). OB2 and OB3 harbour pronounced single-stranded DNA-binding grooves characteristic of such folds. The binding site is formed by one face of the barrel, where three loops come together, i.e. L12 (between β1 & β2), L45 (between β4 & β5), and L3α (between β3 and the α-helix). The tower domain contributes an additional β-strand (Tβ1) to OB2’s groove (pairing with β1), thus increasing its surface area. OB1 has a shallower groove, due in parts to its L12 loop being disordered, and the L45 loop folding away from the groove. Furthermore, the former is longer by ~20 amino acids in OB1 than the average amongst OB folds.

OB2 and OB3 are packed in tandem, with their grooves aligned to make a continuous channel. The packing buries 1261 Å² in total surface area, and involves both van der Waals interactions and hydrogen bond networks. Residues within this interface are also mutated in
cancer, and 4% of tumour-derived missense mutations in BRCA2 DBD map onto amino acids therein, i.e. Val2829, Arg2893, Glu2921, and Asn3042. This prearrangement of the two folds is thus likely to be important for BRCA2’s function. OB1 and OB2 too pack extensively through van der Waals forces and hydrogen bonds, burying 1271 Å² in total surface area. In contrast to OB2 and OB3 however, they arrange in a pseudo two-fold symmetric (head-to-head) manner. Their grooves though remain collinear, as with OB2-OB3 packing. The OB1-OB2 packing is further stabilised by a nine amino-acid stretch of DSS1 that slots like a wedge between the two folds.

The helical domain is composed of a core of four helices (α₁, α₈, α₉, and α₁₀) and two consecutive β-hairpins (β₁ – β₄). A ~50 amino-acid fragment comprising four short helices slithers around the core structure, and constitutes an additional layer of structural folds. Helices α₉ and α₁₀ pack with OB1, by means of van der Waals interactions through hydrophobic and aromatic residues, as well as side- and main-chain hydrogen bonds. These elements are some of the best conserved in the entire BRCA2 DBD. It is also the DSS1-receptor therein alongside OB1.

DSS1 binds to BRCA2 via an N- (a.a. 7 – 25) and C-terminal (a.a. 37 – 63) fragment, with the middle segment being disordered. And while these flanking sequences are highly conserved, the intervening portion is not. Similarly, majority of the DSS1-binding BRCA2 residues are highly conserved as well. The N-terminal segment of DSS1 winds about the helical domain, spans the αD-OB1 interface, and culminates in packing with OB1. The C-terminal segment packs first with the OB1-OB2 interface, followed by OB1, and then finishes at the OB1-αD interface. Two of DSS1’s β-strands, β₁ and β₂, pack with β₆ and β₃ of OB1, respectively. All in all, the two termini of DSS1, which has a 37% acidic residue content (and 13% aromatic), follow the path of an extended, meandering groove on the αD and OB1 surface, which in turn is rich in basic, aromatic and hydrophobic residues. The DSS1- and nucleotide/sugar-binding sites lie roughly on opposite sides of the OB1 fold. Two DSS1-interacting BRCA2 residues are mutated in cancer, i.e. Ala2564 and Arg2580.

The tower domain’s stem is comprised of α-helices Tα₄ and Tα₅. The intervening three-helix bundle (helices Tα₂, Tα₃ and Tα₄) is supported by the N-terminus of Tα₅, which is incorporated into the 3HB’s hydrophobic core. Between 3HB and Tα₄ is a disordered segment of 12 amino acids. The tower harbours four of the seven most mutated residues in BRCA2 DBD, i.e. Ala2872 and Ile2865 that mediate Tα₄-Tα₅ interactions, and Glu2777 and Lys1871 which are solvent exposed on the stem. This in conjunction with the conservation of residues at the tower’s base and 3HB, is implicative of the domain’s role in BRCA2’s tumour suppressor function.
Using native gel electrophoretic mobility shift assay and end-labelled DNA probes, the BRCA2DBD-DSS1 complex was found to have affinity toward oligo(dT), oligo(dC), and single-stranded DNA of mixed sequence – but not double-stranded DNA. However, an isolated helical domain–OB1–DSS1 complex bound single-stranded DNA comparatively weakly (with an estimated affinity 1/100 that of the entire domain).68

Structure of BRCA2DBD-DSS1 in further complex with oligo(dT), single-stranded DNA, was solved separately from the former and with an internal deletion in the tower domain68 (Fig. 11c). The latter yielded better diffracting crystals than otherwise, whilst maintaining an identical core structure. Complete density could be seen for only five of the nine nucleotides. Single-stranded DNA bound in a uniform configuration in the OB2-OB3 channel, with the bases’ edges bound at the top of the channel, and the phosphodiester backbone forming a ridge along its surface. As seen in other single-stranded DNA-binding proteins, BRCA2 OB-ssDNA contacts entail stacking interactions between bases and aromatic residues, hydrogen bonds amongst bases’ edges and polar residues lining the top and sides of the channel, and phosphates interacting with polar residues or main-chain amide groups. Thereby, OB2 made the most significant contacts, and therein three residues involved in DNA binding (but without structural roles) are found mutated in cancer, i.e. Lys2754, Gln2945, and Ser2907 – underlining the importance of single-stranded DNA binding to BRCA2 function. The phosphate group of a sixth nucleotide was visible, interacting with the main-chain amide of Ala3007 from OB3’s L12 loop, but the base was presumably disordered. Remainder of the OB3 groove was unavailable for DNA binding due to being involved in crystal packing.

Tumour-derived missense mutations in the BRCA2 DBD are scattered across the five domains82, with 25.0% residing in the helical domain, 25.4% in OB1, 8.6% each in OB2 and OB3, and 32.4% in the tower domain – suggesting they are all important to BRCA2’s tumour suppressor activity. These encompass residues with structural roles, DNA- or DSS1-binding functions, as well as ones on the tower’s surface.

Role of the specific helix-turn-helix motif within the three-helix bundle at the tower’s end is not clear, though in other proteins it mostly binds double-stranded DNA. It might thence enable BRCA2 to bind at the single- and double-stranded DNA junction upon end-resection. The BRCA2 orthologue in Ustilago maydis (corn smut, a fungus), Brh2, does preferentially bind such ssDNA-dsDNA junctions85. However, BRCA2 itself has been found to exhibit no especial preference for these junctions over mere single-stranded DNA. Though the adoption of secondary structure by the long stretches of ssDNA generated during resection could mimic dsDNA.
The abovementioned structures were derived from crystals of a 736 amino-acid mouse BRCA2 DBD bound to DSS1, which diffracted to 3.1 Å; those further bound to oligo(dT)$_9$ ssDNA fragments, and then ones with a deletion of the tower therein that both diffracted till 3.5 Å. However, further crystals of an 816-residue rat BRCA2 DBD bound with DSS1 were also obtained, and diffracted to 3.4 Å.

**Figure 11.** Crystal structure of BRCA2 DNA/DSS1-binding domain$^{68}$. (a) Tertiary level view of the mouse BRCA2DBD–DSS1 structure, showing the helical domain (purple),
followed by the three successive oligonucleotide/oligosaccharide-binding folds OB1, OB2 and OB3 (orange, blue and teal respectively), and the tower (T) domain insertion in OB2 (also blue). A three-helix bundle (3HB) can be seen atop the tower. The structure is depicted from C- to N-terminus (left-right), in order for DNA and DSS1-binding surfaces to face the viewer. DSS1 (lime green) is seen bound between the helical and OB1 domains. Disordered regions are missing. (b) A generic OB-fold structure based on its smallest variant, the B subunit of verotoxin-1 (VT1B)\textsuperscript{84}. Five β-strands (ribbon arrows), fold into a closed β-sheet (i.e. β-barrel), capped by an α-helix (cylinder). Three variable loops (L\textsubscript{12}, L\textsubscript{33}, and L\textsubscript{45}), of the five in total linking the structural segments, contribute residues for the oligomer-binding site. (c) The BRCA2DBD\textsuperscript{ΔTower}–DSS1–oligo(dT)\textsubscript{9} structure\textsuperscript{68} (monochrome), with ssDNA (red sticks) and DSS1 (gold) for emphasis. Only five of the nine nucleotides are ordered and discernable. Bases bind along the side and top of the OB2–OB3 channel. For a greater discussion of this see §4.2 and Fig. 15. Created in MacPyMOL using PDB-1MIU, -1BOV, and -1MJE.

1.6 *In vitro* full-length BRCA2

Mechanistic (including structural) studies of human BRCA2 have been hindered by the challenges of purifying a single polypeptide of its size (i.e. 3,418 a.a.). Its biochemical and molecular functions were thus inferred by using fragments of the molecule, or small fungal and worm orthologues (Brh2 and BRC-2 respectively), which each comprise a single BRC motif and are known to stimulate ssDNA binding by RAD51, and/or its promotion of DNA-DNA interactions during homologous recombination.

Moreover, its insufficient solubility and propensity to degrade when expressed in the laboratory (being eight to ten times more massive than the average protein), has further hampered its isolation\textsuperscript{85}. Also, it tends to occur in complexes with other proteins, and is often unstable on its own.

Following a decade and a half endeavour, derivation of pure extracts of the full-length human BRCA2 protein was reported concurrently by the Kowalczykowski\textsuperscript{*}, Heyer\textsuperscript{†}, and

---

* Stephen C Kowalczykowski, Department of Microbiology and Molecular Genetics, University of California Davis, USA.
† Wolf-Dietrich Heyer, Department of Microbiology and Molecular Genetics, University of California Davis, USA.
West‡ groups. This is believed to constitute the largest singular protein to have been purified to homogeneity.

The Kowalczykowski laboratory employed two maltose-binding protein (MBP) tags at the N-terminus of hBRCA2, which helps its solubility and folding, and stabilises the protein that was then purified from cultured human epithelial kidney cells. The Heyer group purified the protein from yeast, having affixed a glutathione S-transferase (GST) tag to its N- and a Flag epitope and polyhistidine (His	extsubscript{10}) tag at its C-terminus, respectively. This policy allowed the group to verify the integrity of the purified product by detecting for the presence of both tags. West and team purified BRCA2 from human epithelial cancer (HeLa) cells, having inserted the gene as part of a bacterial construct – see §2.1.1 for details.

The first study showed it bound human RAD51 and DMC1, yeast Rad51, but not E. coli RecA in vitro. However, there was also no significant interaction with human RPA, E. coli SSB, or human RAD52. The amount of RAD51 molecules bound by each copy of BRCA2 scaled linearly with the former’s concentration (indicative of strong interaction in the nM domain), till about 4.5 ± 0.9 RAD51s per BRCA2 were bound. Following on, weaker binding (in the µM range) was manifest. At maximal saturation, ~6 RAD51 proteins were bound to every BRCA2. BRCA2 strongly preferred binding to either single-stranded, or double-stranded DNA with an ss-tail, over dsDNA. It also prevented and favoured RAD51’s association with ds- and ssDNA respectively. When exposed to both single- and double-stranded DNA, strand exchange activity of RAD51 is reduced to background levels; however after prior incubation with BRCA2, this inhibition was diminished in a concentration-dependent manner. Though its magnitude was less, stimulation by BRCA2 was maintained in the presence of RPA. Prior incubation of ssDNA with RPA acutely suppressed DNA strand exchange by RAD51 in an absence of BRCA2, which was then promoted by up to 20-fold upon addition of the latter. An effect was evident at substoichiometric ratios as low as 100-fold. Therefore, despite not possessing explicit RPA binding activity, it seemed that interactions of BRCA2 with RAD51 and DNA were sufficient to allow RAD51 access to RPA-coated ssDNA, for RPA to then be displaced as filament nucleation and growth ensued. Furthermore, BRCA2 stabilised the RAD51 nucleoprotein filament via inhibition of its ssDNA-dependent ATP hydrolysis activity in a concentration-dependent manner to levels observed in absence of DNA, since ATPase activity inactivates and turns over the RAD51 protein.

‡ Stephen C West, Cancer Research UK London Research Institute, Clare Halls Laboratories, South Mimms, UK.
The second study also confirmed purified human BRCA2 bound RAD51, and indicated a stoichiometry of 6 ± 1 molecules of RAD51 for each copy of BRCA2. Though dependent on accuracy of measurement of the latter’s concentration, it was nonetheless fairly consistent with the values derived by Kowalczykowski group. Single molecule studies had suggested a nucleation of as little as 2 or 3 RAD51 monomers was adequate for priming filament growth (though other estimates of 4 or 5 are higher), but in either case BRCA2 seemed to bind enough RAD51 protomers to promote filament extension. Minimal RAD51 binding to single-stranded DNA occurred in presence of saturating amounts of RPA, but this was stimulated seven-fold on addition of BRCA2. A stimulation was observed at substoichiometric concentrations of 1 BRCA2 to 33 RAD51 molecules. The effect was highest in presence of ATP and Ca\(^{2+}\), which stabilises the active ATP-bound form of RAD51 on ssDNA. BRCA2 exhibited no significant differences in binding to either an ssDNA circle, or a gapped DNA substrate comprising five 5’ and four 3’ junctions. RAD51’s binding to ssDNA, i.e. in presence of BRCA2, was enhanced five-fold by DSS1 than without.

Here it could be noted that BRCA2’s saturation at having bound 6 RAD51 molecules, is consistent with Pellegrini et al.’s observation of only six of the eight BRC repeats being capable of direct interaction with RAD51\(^{79}\) (§1.5.4).

Lastly, the study by West laboratory found BRCA2 bound single-stranded DNA in a concentration-dependent manner, but did not form stable complexes with dsDNA – demonstrating that it was an ssDNA-binding protein as suggested by structural and biochemical analyses of the BRCA2 DNA-binding domain by Pavletich group. Furthermore, its electron microscopy visualisation by rotary metal shadowcasting revealed putative rod-shaped ‘dimers’, but also ball-shaped ‘monomers’. These displayed dimensions with the tungsten coating of 15 ± 2 nm, and 270 ± 30 nm × 16 ± 2 nm, respectively. Their respective masses were inferred by comparison with ferritin to be 337 kDa and 860 kDa, i.e. consistent with BRCA2 monomers and dimers. Examination of protein molecules bound to tailed DNAs (3.5 kb linear duplexes with 54-nt 3’ ssDNA tails at one end), showed 91% were situated at an end of the DNA substrate, none were bound to both ends, and only 9% were associated with the internal dsDNA region. In assessing whether BRCA2 modulated the relative ss- and dsDNA binding affinities of RAD51, presence of low quantities of the former was found to vastly increase latter’s binding to ssDNA. Likewise, BRCA2 reduced RAD51’s binding to dsDNA, illustrative of the provision of differential targeting necessary for homologous recombination. The DNA-binding assays were performed in low-salt conditions, whereby BRCA2 binds ssDNA, while RAD51 binds ss- and dsDNA equally; and comprised ATP and Mg\(^{2+}\) to the facilitate formation of active nucleoprotein filaments. Added analysis did not show evidence of a BRCA2-RAD51-ssDNA ternary complex per se, suggesting BRCA2
acted as a molecular chaperone that facilitated RAD51’s loading onto ssDNA, and then dissociated from the two. BRCA2 was also found to promote RAD51-effected DNA strand exchange by virtue its abovementioned properties. Its dimerisation would be consistent with the observed dimeric forms of Brh288, and the crystallographic dimer of BRCA2 DBD seen its X-ray study68.

While these experiments corroborated ideas regarding BRCA2’s function, i.e. especially that it helped nucleate RAD51 filament growth, and refined understanding of the mechanism by which this is coordinated. Perhaps more importantly, the purifications provided an avenue for structural studies amongst others of full-length BRCA2. Though it is remarkable this extent of biochemical experiments could be performed with the microgram quantities derived, the still extremely low yields preclude it from high-resolution structural analyses.

1.6.1 Overview of thesis

A couple of prodigious efforts at the genetic, cellular, and latterly biochemical levels (several of which are reviewed hereby), have unravelled some valuable insights into the functions of this very important protein, whose malfunction clearly continues to blight many human lives.

However, understanding of the explicit mechanistic bases for its action remains incomplete, and structural characterisation of full-length BRCA2 is now conspicuous by its absence. Many fundamental questions can still be posed, for instance with regards to the organisation of various structured/folded domains within BRCA2; its manner of interaction with RAD51 and the number of protomers that are typically bound; whether it interacts with RAD51 monomers or polymeric assemblies such as rings and filaments; the manner of its interaction with single-stranded DNA and facilitation of RAD51’s loading thereupon; its oligomerisation state; whether it initiates filament nucleation at a single or multiple sites on ssDNA; and if filament growth proceeds from BRCA2 bidirectionally, or its polarity if otherwise. The list is by no means exhaustive, as this and more should become clarified.

Hereby, we provide structure- as well biochemically-based mechanistic insights into full-length BRCA2’s interaction with RAD51 and DNA, and its facilitation of recombinase loading and nucleoprotein filament formation; and thus delineate some of the precise functions of BRCA2 in homology-directed repair by RAD51 that have remained ill-defined in the overall scheme.

We sought to visualise BRCA2’s mechanism via single-particle negative stain electron microscopy. The low abundances of particles in the preparations precluded them from cryo-EM. We solved the structure of full-length human BRCA2 on its own and in complex with
human RAD51, and identified BRC repeats and the C-terminus in the former and RAD51 in the latter by antibody labelling. We also analysed the BRCA2-ssDNA complex and BRCA2-RAD51-ssDNA pseudo-complex by 2D electron microscopy.

Also, we quantify the affinity of BRCA2 toward various lengths of single-stranded DNA, and establish the minimum stretch requisite for effective DNA-binding. We probe BRCA2’s effect on both the length of RAD51 nucleoprotein filaments, and the frequency of nucleation events. Furthermore, we ascertain whether BRCA2 would bind at multiple sites along an ssDNA molecule; and determine the polarity that RAD51 filaments emanate from BRCA2, and subsequently grow.

Significantly, we discover that human BRCA2 might most commonly occur as dimers, and discuss the implications and mechanistic advantages that are conferred by such an arrangement in light of and with regards to the many of the above.
2 Methods

2.1 Protein expression and purification

2.1.1 BRCA2

The human BRCA2 protein used here was first purified as such in the laboratory of Prof Stephen C West, Cancer Research UK London Research Institute (Clare Hall laboratories), South Mimms, UK. Purification was later also carried out by Dr Eric H Kong of the Zhang group, Macromolecular Structure and Function Laboratory, Imperial College London, South Kensington, UK.

A bacterial artificial chromosome (BAC) was constructed in which the human BRCA2 gene complete with upstream and downstream regulatory elements was FLAP tagged at either the N- or C-terminus. The FLAP tag is composed of enhanced green fluorescent protein (EGFP), and S- and Flag-affinity tags separated by PreScission- and TEV-protease sites. Stable HeLa cell lines expressing the gene were selected and the protein purified by Flag- and GFP-affinity purification. GFP and S tags were removed by TEV cleavage, leaving only a small 1.3 kDa Flag tag. Some PALB2 was found to co-purify with N-terminally tagged BRCA2, whereas little to none was associated with the C-terminally tagged protein.66

In our preparations, HsBRCA2 was obtained by slight modifications to the above methodology. C-terminally GFP-Flag tagged BRCA2 was purified from 20 litres of HeLa cells containing a BAC-BRCA2FLAP vector. Following cell collection and lysis via a homogeniser, BRCA2 was purified with Anti-Flag M2 affinity gel (Sigma Aldrich) and GFP-Trap_A agarose beads (Chromotek). Order of the affinity steps was reversed in some derivations without apparent differences. The final protein (200 µl at 2.5 µg/ml), was saved as small aliquots at -80 °C. For initial electron microscopy analyses, the purified protein was further subject to a gel-filtration step through Superose 6 10/300 GL (GE Healthcare), to provide a more monodisperse population.

As BRCA2 was expressed from its own promoter at levels similar to the endogenous protein, the sample was very dilute (precluding it from high-resolution cryo-electron microscopy studies), but of good quality/purity and reasonably homogeneous – see Fig. 15a.

2.1.2 BRCA2-RAD51

To constitute the BRCA2–RAD51 complex, BRCA2 was partially purified with GFP-Trap agarose and incubated with excess human RAD51, i.e. 0.5 µM for 30 minutes at 4 °C. The BRCA2-RAD51 complex was purified away from the remaining RAD51 by binding to
M2-Flag agarose, washed extensively and eluted with 3x (triple-) Flag peptide (Sigma Aldrich).

HsRAD51 had been purified as per its first reported purification, again by the West group (CRUK LRI, Clare Hall laboratories, South Mimms, UK). Human RAD51 could thence simply and quickly be purified following overproduction in *E. Coli*.\textsuperscript{89} Dialysis of cell-free extracts against buffer comprising low concentrations of spermidine, results in selective precipitation of HsRAD51 as microcrystallites (observable by light microscopy). These crystals could then easily be redissolved in phosphate buffer, and the protein purified to homogeneity by hydroxylapatite, affi-gel heparin and Q-sepharose chromatography.\textsuperscript{90}

2.2 DNAs

ssDNA binding affinity of BRCA2 with increasing lengths of the former was assessed in the West laboratory. Oligos of 20, 30, 33, 37, 40, 50, 66, 74 and 100 nucleotides were employed. Those ranging between 20- and 70-nt, all constituted 5’ derivatives of the following 74-mer (written 5’–3’):

\[ \text{GCGCTACCCAGTGCACCAATGGATTGCTAGGACATCTTTGCCCACCTGCTACCC} \]
\[ \text{CGGTTCCAAATCGATAAG} \]

The 100-mer was (5’–3’):

\[ \text{GGGCCGAATTGGCGACGCCGACGTGCATGCCTCTCTAGACTCGAGAATTCGGTACC} \]
\[ \text{CCGGGGTTCGAAATCGATAAGCTTACAGTCTCCATTTAAAGGACAAG} \]

We created two gapped DNAs (gDNAs), for purposes of characterizing the BRCA2-ssDNA complex by 2D electron microscopy. The intention being for BRCA2 to bind in the intervening single-stranded DNA sequence, while the flanking duplex arms helped delineate the path of DNA-binding across its surface. This is because while the ssDNA, which besides would be occluded by BRCA2, is difficult to visualise under negative-stain, the dsDNA would not and is possible to be observed. The ssDNA gap sizes were designed for BRCA2 to bind into comfortably in light of results from the previous experiment.

gDNA1 had two 50-nucleotide oligos annealed to the 5’ and 3’ ends of a 200-mer, resulting in an ssDNA gap of 100-nt. gDNA2 was related but the complimentary oligos were 65 nucleotides in length, giving a 70-nt ssDNA interval. The 200-mer had the following sequence (5’–3’):

\[ \text{GAGTTTTATCGCTTCCATGACGCAGAAGTACACTTACCTGGATATTCTGATGAG} \]
\[ \text{TCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTTACGAAATTAAAT} \]
CGAAGTGACTGCTGGGAAAAATGAGAAAAATTGCACCTATCCTTGCGCAGCTCGAGAACGTCTTACTTTCGCCCATCAACT.

To study RAD51 filament nucleation events on single-stranded DNA mediated by BRCA2 under electron microscopy, linear viral φX174 ssDNA was used (West lab). This was prepared by annealing a 22-nucleotide oligo, 5’-TAAAAACTCTCTGCGTTGGATACGCCAAATC-3’, to φX174 ssDNA, and subsequent treatment with PstI restriction enzyme. The linearized ssDNA was then gel purified.

The gDNAs were constructed by Dr Eric Ho Kong of the Zhang laboratory, Imperial College London, UK.

2.2.1 DNA gold-labelling

To deduce the direction of RAD51 filament growth from BRCA2 on single-stranded DNA, ssDNA molecules gold-labelled at either their 5’- or 3’-terminus were constructed (West group). A 5’-biotinylated ssDNA substrate was generated via PCR amplification with a single primer (5’-biotin-TAAAAACTCTCTGCGTTGGATACGCCAAATC-3’), and φX174 RFI (replication form 1) dsDNA as template. The polymerase chain reaction product, 5’-biotinylated linear single-stranded DNA, was gel purified and labelled at the 5’-end with 5 nm gold particles coupled to streptavidin.

2.3 Protein-DNA complexes

2.3.1 BRCA2-ssDNA

Electromobility shift assays (EMSAs) were carried out to determine the minimum length of single-stranded DNA that BRCA2 could effectively/stably bind (West laboratory).

BRCA2-ssDNA binding reactions (10 µl), contained ssDNA oligonucleotides (1 nM) that were 32P-labelled at their 5’-end, together with BRCA2 (1 nM) in binding buffer (20 mM HEPES, 2 mM ATP, 2 mM MgCl2, 1 mM dithiothreitol, and a pH of 7.5). The process was analogous for 3’-gold labelled ssDNA.

After 10 minutes at 20 °C, the protein-DNA complexes were fixed with 0.25% glutaraldehyde, and incubated at 20 °C for 15 minutes. They were then analysed by electrophoresis through 1% agarose gels run for 2 hours in TAE buffer at 4 V cm⁻¹, dried on to filter paper, and visualised by autoradiography.

In order to prepare the BRCA2-gDNA complexes for electron microscopic visualisation, we incubated purified BRCA2 with 100 nM of gapped DNA for 30 minutes at 37 °C in a low salt buffer (50 mM HEPES pH 7.5, 10% glycerol, 50 mM NaCl, 0.01% Triton X-100, 1 mM
EDTA, 1 mM dithiothreitol). Binding of BRCA2 to the gapped DNAs was affirmed by EM-SA.

The latter experiments were performed by Dr Eric Kong, MSF lab, Imperial College, London, UK.

2.3.2 **BRCA2-RAD51-ssDNA**

RAD51-ssDNA complexes were prepared by incubating 150 nM RAD51 with 0.2 nM linear φX174 single-stranded DNA, in binding buffer at room temperature (as an internal control, again in the West lab).

To form BRCA2-RAD51-ssDNA complexes, 0.1 nM BRCA2 was incubated with 150 nM RAD51 for 1 minute in binding buffer at room temperature, prior to the addition of 0.2 nM ssDNA and further incubation for 3 minutes.

To probe the effect of BRCA2 on the frequency of RAD51 filament formation, 40 nM RAD51, with and without 0.2 nM BRCA2 (the latter as a control), was incubated for 1 minute in binding buffer at room temperature, before an addition of 0.2 nM ssDNA, following which the incubation was carried on for a further 1 minute.

To determine the polarity of RAD51 nucleoprotein filament formation from BRCA2 (as in §2.2.1), 0.2 nM BRCA2 and 150 nM RAD51 were premixed in binding buffer, sans dithiothreitol. After 1 minute, 5’- or 3’-gold labelled ssDNA was added, and the incubation continued for another 1 minute.

These preparations were then applied to pre-treated (glow-discharged) electron microscopy grids, and negatively-stained for analysis.

2.3.2.1 **Protein gold-labelling**

In order to ascertain the identity of BRCA2 in visualisations of filaments emanating therefrom, immunoaffinity gold-labelling of BRCA2 was performed by means of an antibody toward Flag (F3165 anti-Flag M2 mouse monoclonal, Sigma Aldrich), conjugated to 20 nm gold particles with an InnovaCoat nano-gold conjugation kit (Innova Biosciences). 0.2 nM BRCA2 was incubated with 0.1 μg/ml of antibody-conjugated gold nanoparticles for 15 minutes in binding buffer (without dithiothreitol). Afterwards, 150 nM RAD51 was added, followed in 1 minute by 0.2 nM ssDNA, and the incubation resumed for another 1 minute.

First, BRCA2 (0.2 nM) was incubated with 0.1 μg/ml antibody-conjugated gold particles for 15 min at room temperature in binding buffer without dithiothreitol. RAD51 (150 nM)
was then added for a further 1 min; this was followed by the addition of 0.2 nM ssDNA, and incubation was continued for a further 1 min.

2.4 Antibody labelling

We employed antibody labelling in order to identify the locations of various domains or regions within our electron microscopy reconstructions of BRCA2 and the BRCA2-RAD51 complex.

For BRCA2, we used an antibody against the Flag tag at the C-terminus (anti-Flag M2-peroxidase produced in mouse, Sigma Aldrich), and another specific for a region between BRC5 and BRC6 (OP95 anti-BRCA2 Ab-1 mouse monoclonal, Merck Millipore).

For the BRCA2-RAD51 complex, we utilised an anti-RAD51 antibody (14B4 ab213 produced in mouse, Abcam).

Complexes with antibodies were prepared by incubation with 10–30 nM of antibody for 30 minutes at 4 °C. In both cases, binding efficiencies of the antibodies towards BRCA2 and BRCA2-RAD51 were gauged by immunoprecipitation, prior to EM analysis. 5 µg of antibody was incubated with 2 ml of BRCA2 lysate, or 10 ng of purified BRCA2-RAD51 complex, for 2 hours in 2 ml IP buffer at 4 °C accompanied by constant mixing. 50 µl of Dynabeads Protein G (Life Technologies) was then added, and the mixture incubated for another 2 hours to capture antibody-bound protein, or complex. Beads were washed five times with 1 ml IP buffer, resuspended in SDS PAGE gel loading buffer, and analysed by SDS PAGE. Captured antibody complexes were then detected by Western blotting.

The complexes were constituted and purified by Dr Eric Kong of the Macromolecular Structure and Function Laboratory, Imperial College London, UK.

2.5 Negative-stain electron microscopy

2.5.1 Specimen preparation

For purposes of electron microscopy reconstruction of BRCA2, 2 µl of protein solution was applied to a TAAB 300-mesh continuous carbon grid, which had been glow-discharged for approximately 60 seconds and left for another ~60 s. Excess sample was blotted, and the grid stained with 2% w/v uranyl acetate solution for yet another ~60 seconds. Finally, excess stain was removed by blotting with filter paper, and the grid allowed to air dry. The grid itself was a 3 mm-wide circular copper mesh, precoated with a thin layer of continuous graphite film.
For BRCA2-RAD51, the procedure was similar in outline. Although thereby a copper Quantifoil R2/2 holey carbon grid was employed, which was coated in-house with a thin layer of continuous carbon film. This practice sometimes has the advantage of allowing easier judgement of well-stained locales on the grid, from the shading of carbon-covered holes. Also, it allows one to be more wilful with regards to the thickness or otherwise of carbon coating, which is often found to be inconsistent among commercially available ready-coated grids. The glow-discharged grid in this case, had the complex adsorbed for 2 minutes at room temperature, washed twice with water, and then stained as previously.

The coating process entailed graphite covered mica sheets whereby the carbon was stripped off the mica by floating it on water, and then lowered onto grids underneath by suctioning the water. Coated sheets were kindly supplied by Dr Fabienne Beuron, Electron Microscopy Facility, The Institute of Cancer Research, London, UK.

Similar methodology again to the two above was utilised for our gapped-DNA and antibody complexes.

When analysing RAD51-ssDNA and BRCA2-RAD51-ssDNA complexes (West group), the samples were applied 2 minutes after glow discharge and negatively stained as before. When analysing RAD51 filament nucleation on ssDNA in the presence and absence of BRCA2, the samples were administered after 1 minute (prior to staining).

The glow-discharge treatment confers an overall charge to the carbon, and increases the adsorption of the particles to the grid potentially by several orders of magnitude. The stain solution pervades the space in between the particles, possibly permeating any cavities and/or crevices present within the protein molecules themselves. Heavy metal atoms such as those of uranium in this case, are exceptionally electron-rich and as a consequence scatter any incident electrons very strongly, i.e. as opposed to focusing the electrons – that results in the background appearing dark with the particles in white (hence the name).

### 2.5.2 Data acquisition

The prepared grids were loaded into a Philips CM200 series FEG (field emission gun) transmission electron microscope (TEM), equipped with a TVIPS TemCam-F415 4,096 × 4,096 pixel charged-coupled device (CCD) camera, at the Centre for Biomolecular Electron Microscopy (CBEM), Imperial College London. This was the case for BRCA2, the BRCA2-RAD51 complex, and the antibody complexes.

For BRCA2-gDNA complexes, higher contrast was desired to enable the DNA to be seen; hence these data were collected on an FEI Tecnai F20 TEM equipped with a Falcon II 4k × 4k direct electron detector (FEI), again at CBEM Imperial.
In both cases, the microscopes were operated at their maximum operating voltages of 200 kV, and 16-bit images were acquired at 50,000× magnification. Given the pixel densities of their respective detectors, this resulted in sampling intervals of 1.76 and 2.05 Å pixel\(^{-1}\) for the CM200 and F20 TEMs respectively. Images at the former and latter were collected with electron dosages of around 40 and 20 e⁻ Å\(^{-2}\) respectively; exposure had been compensated at the F20 because of much greater sensitivity of its detector. Objective astigmatism was typically corrected at higher magnifications still (100,000 or 150,000×).

BRCA2, BRCA2-RAD51, and the antibody complexes were visualised at nominal underfocus ranging from ~ 1 or 1.5 μm to 3 μm. BRCA2-gDNA images were acquired at defoci of ~ 4 μm, again due to the requirement for higher contrast.

Collection of tilt-series and -pair data for the purposes of validating the BRCA2 and BRCA2-RAD51 3D reconstructions is set out in §2.5.4.10.

The other electron microscopic analyses were carried out by the West laboratory, using an FEI Tecnai G2 Spirit T12 twin transmission TEM comprising a Gatan Orius 3,000 × 4,000 pixel CCD camera. These observations were generally performed at a magnification of 67,000×.

2.5.3 Data preprocessing

2.5.3.1 Coarsening

For the purposes of image processing, raw micrographs or ‘boxed’-out particles from therein were ‘coarsened’ by a factor of 2, essentially coalescing adjacent pixels. This causes file dimensions, and sizes divided by 2\(^2\); and pixel sizes multiplied by 2, to 3.52 and 4.10 Å pixel\(^{-1}\) for the CM200 and F20 respectively. This allows faster subsequent processing times. It is also not sacrificial in the information sense, as given the various artefacts introduced by negative staining (addressed later), the image data is far removed from having any ‘real’ signal to such resolution scales.

It so happens as a consequence of the Nyquist sampling theorem\(^\text{92}\) (a fundamental result in digital information theory) that the theoretically achievable resolution from data sampled at 3.52 Å pixel\(^{-1}\) (a ‘spatial’ sampling frequency of \(v = \frac{1}{3.52 \, \text{Å}}\)), is limited to 2 × 3.52 Å = 7.04 Å (the Nyquist frequency of \(f_{\text{Nyquist}} = \frac{v}{2} = \frac{1}{7.04 \, \text{Å}}\)) — again much under the bounds of negative stain electron microscopy.

The coarsening process was performed in IMAGIC\(^\text{93}\) (Image Science Software), with the aid of specifically written scripts to automate the process for multiple micrographs or particles.
2.5.3.2 Particle selection

Individual particles were picked manually in EMAN2\textsuperscript{94} (e2boxer), into ‘box’ dimensions of 128 × 128 (post-coarsening). For antibody and gDNA complexes, 162 × 162 boxes were used in order to accommodate the two.

The above yielded a dataset of 8,773 particle images for BRCA2, and 6,877 for BRCA2-RAD51.

Care was exercised at avoiding obvious aggregates and potential degradation products. The latter usually exhibited significantly small sizes, as well as inconsistent dimensions and shapes amongst each other.

Neither the micrographs, nor the picked particles were CTF-corrected prior to processing.

2.5.3.3 Filtering, masking and normalisation

These raw images contain undesired data in addition to useful information about the particles. This comprises low frequency artefacts over long scales (such as background density ramps and stain gradients), and high frequency noise over short scales that could both harm the accuracy of image alignment and angular assignment procedures, which rely on cross-correlation functions. High frequency elements, although noisy, also harbour the fine details one hopes to mine from the dataset. Therefore, while it might be necessary to suppress such spatial frequencies for the purposes of determining low-resolution preliminary models and during the initial stages of refinement, they could later be reintroduced for achieving maximal resolutions.

The band-pass filter normally applied is the product of a narrow Gaussian high-pass filter to cut away the disturbing low spatial frequencies, and a wide Gaussian low-pass filter to cut away the high frequencies. The former is associated with the low frequency cut-off that gradually cuts away spatial frequencies lower than this value, and the latter with the high frequency cut-off that does the same for frequencies greater than its value. Band-pass parameters are chosen to maintain the outer envelope of the particle (a low frequency cut-off based on the size of the molecule), and to remove high frequency noise beyond the expected resolution (high frequency cut-off).

The values appointed corresponded to 270 Å for the former (a number that should be greater than the biggest dimension of the particle), and 10 Å for the latter. These were expressed in terms of their corresponding spatial frequencies as a fraction of the Nyquist frequency, which equated to \( \frac{1}{7.04 \AA} \) and \( \frac{1}{10 \AA} \) respectively. The latter’s value was changed to 1, corresponding to the Nyquist limit of 7.04 Å, during the final stages
of refinement. For the purposes of antibody and gapped DNA complexes, a low pass filter (high frequency cut-off) corresponding to 60 Å was employed, in order to enhance the contrast therein by emphasising the overall features and suppressing fine details.

Following filtering, a soft-edged circular mask of fractional radius 0.9 and a drop-off of 4 pixels, was applied to the particle images. The former is expressed as a fraction of the inner radius of the box, and the latter as the half-width at half-maximum of the roll-off that follows.

Gaussian (viz. gradual) filters are applied in Fourier space to avoid an introduction of ‘ripples’ in the real space image, which would be the effect of a sharp-edged filter. Likewise, ‘soft’ masks are applied to real images to avoid creating ripples in Fourier space. This is important, as many single particle electron microscopy algorithms involve forward or reverse Fourier transforms, and subsequently often correlation calculations.

For individual images, the range of density values (represented by grey levels) may be inconsequential, as display programs often scale these numbers for representation. But should one be performing cross-correlations for the alignment of particles images to a set of references, and some happen to contain exceptionally large numerical values, they would yield higher peaks and falsely appear to provide the best match. Moreover, the value of the Fourier transform at the origin is equal to the mean value of the image; hence a high mean value would generate a big central peak in the Fourier transform that is unrelated to the structural information.

Hence, the particle images were normalised to an average density of 0, with an arbitrarily selected standard deviation of 10.

In 16-bit greyscale images, each pixel can assume $2^{16}$ (i.e. 65,536) different grey values that indicate its brightness level.

All the above procedures were carried out simultaneously (in immediate succession) in IMAGIC-V.

### 2.5.4 Image processing (IMAGIC V)

Single molecules in solution are not held in fixed orientations and hence have six degrees of freedom, three translational ($X, Y, Z$), and three rotational corresponding to the ‘Euler’ angles ($\alpha, \beta, \gamma$) – see Fig. 12. The image collected in an electron microscope is a projection along the $Z$-direction (by matter of definition), and thus eliminates the out-of-plane translation parameter $Z$. This leaves five parameters to be determined for every particle image, the two in-plane translational degrees of freedom ($X, Y$), the in-plane rotational degree of free-
dom ($\alpha$), and the two out-of-plane rotational degrees of freedom ($\beta, \gamma$). The in-plane degrees of freedom ($X, Y, \alpha$) are removed by various forms of alignment.

It is the out-of-plane rotations of individual molecules ($\beta, \gamma$) that are the key to three-dimensional information, and must be identified in order to carry out a 3D reconstruction. The processing of single particle images is thus aimed at determining these parameters for each of the individual particles in the dataset(s).

Figure 12. The six degrees of freedom of single molecules in solution, three translational ($X, Y, Z$), and three rotational ($\alpha, \beta, \gamma$). An image collected in a microscope is a projection along the Z-axis, which leaves five parameters to be determined. Alignment of the particles removes the in-plane degrees of freedom ($X, Y, \alpha$). This leaves only the out-of-plane rotational degrees of freedom ($\beta, \gamma$) to be determined for each of the individual particles, in order to perform a 3D reconstruction.

2.5.4.1 Reference-free alignment (centring)

Alignment of low-contrast images with respect to a set of well-defined references images tends to bias the dataset towards the properties of the references, purely by selecting for similarities with the template contained within random noise variabilities in the particle images. Therefore, ‘reference-free’ alignments are to be used to obtain the first inventory of the various typical molecular views present within the dataset (most especially when no prior structural information is available with regards to the sample at hand), and in its most simple iteration can involve simply the translational ‘centring’ of the molecular images. Since only a translational alignment is performed, the rotational orientation of all the molecules remains random.

Multivariate data analysis methodologies are then used to find similar images, in similar rotations, which come to constitute ‘classes’. The resulting class averages are then compared
to find similar averages with different in-plane rotational orientations. Such systematically present views constitute good, unbiased reference images with which to commence multi-reference alignment procedures.

In our case, the pretreated particle images of BRCA2 and the BRCA2-RAD51 complex were centred by ‘mass’, which places the centre of image mass (i.e. intensities) at the centre of the image frame. Only density values within a circle of radius 0.9 as a fraction of the inner radius of the images, and above a density threshold of 0.0 were utilised for mass calculation. The former corresponds to the masking radius of the raw particles, and contributes toward some saving of processing time (especially for a few of the latter, more computationally intensive procedures).

2.5.4.2 Multivariate statistical analysis (MSA)

Multivariate statistical analysis (MSA)\(^{95-97}\), singular value decomposition (SVD), or principal component analysis (PCA), are methods employed for multivariate data analysis. Despite subtle mathematical differences, what these methods have in common regarding application to large image datasets, is that they are all based on the decomposition of the total interimage variance into mutually orthogonal components in hyperspace – eigenvectors (or eigenimages), which are ordered according to decreasing magnitude. This decomposition is specific to the dataset being analysed.

In essence, MSA looks for differences in features among the various particles, which it extracts out as ‘references’. The first eigenimage one obtains is an average of all the aligned particles, the second image represents the strongest set of differences that exist among the set of particles ignoring that average, and the third image any differences that are left after taking those two differences out etc. The object is to isolate the strongest set of varying features from amongst the particles, which can be utilised as a mechanism for eliminating noise (as not much noise is seen appearing in these images, i.e. what one mostly observes is the real variations that occur due to changes in the particles). This process hence constitutes an efficient noise reduction method, and then instead of classifying the particles directly, one classifies them after they have been ‘filtered’ as per this scheme. Having run this procedure, one can then perform k-means clustering (in MSA subspace), and derive a set of class averages for one’s particle data – see §2.5.4.3.

The biggest advantage of the MSA technique is that it enables any particle in an entire dataset comprising many thousands of individual particles to be represented as a linear combination, with different weights or eigenvalues, of the eigenimages of the set\(^{98}\). This effects a massive a compression of the information in the dataset, correspondingly reducing the computing time required for the processing of the images, and facilitating their interpreta-
An additional advantage of this methodology is that the eigenimages of the centred data allow one to have an objective indication of the symmetry properties of the constituent particles (since it would constitute one of the main features of symmetric molecules)\textsuperscript{98}.

One is then able to calculate the ‘distance’ between the images based on the eigenvalues ascribed to them\textsuperscript{98} (see Fig. 13a). In practice, an entire dataset could usually be described by the early eigenimages, with the higher eigenimages pertaining mostly to the noise components of the data.

The pre-treated and centred particles from the BRCA2 and BRCA2-RAD51 datasets were subjected to MSA, employing initially 30 eigenimages, and later the maximum allowed 69. Only pixels within a circle of radius 0.9 as a fraction of the inner radius of the input images were defined to be active during the process.

2.5.4.3 Hierarchical ascendant classification (HAC)

After MSA has culminated, one can calculate the distance between each image in a dataset and every other image. This is in order to build class averages, one needs to be able to group similar particles together into classes.

Originally, each image is a ‘class’ by itself, but then the ‘nearest’ classes can be coalesced two at a time, until at the end all the images become a member of one large class encompassing the entire dataset. This process is called ‘hierarchical ascendant classification’ (HAC)\textsuperscript{99,100} – see Fig. 13b. At each level of the classification, two classes are merged in a way that intra-class variance is minimised and inter-class variance maximised. This is done as per the Ward criterion\textsuperscript{101}, which describes the intra-class variance added upon the merger of two classes $i$ and $i'$, and is given by:

$$\text{Added variance}_{i,i'} = \frac{(w_i \cdot w_{i'})^2}{(w_i + w_{i'})} \cdot d_{i,i'}^2,$$

where $d$ is the distance between the two classes as determined by MSA, and $w_i$ and $w_{i'}$ are the ‘weights’ respectively. At any given point in the procedure, the program attempts to merge pairs of classes together, which at that level of the classification will be associated with minimal added variance.

Finally, the partition obtained with HAC is post-processed, by allowing members of all classes to ‘migrate’ to any class more conducive in terms of minimising the intra-class variance\textsuperscript{98}.

In order to obtain a set number of classes from the above process, one can ‘cut’ the ‘tree’ of this history of mergers at the appropriate level. The images contained in classes under this level could then be summed into single class averages.
For the BRCA2 dataset, the number of classes in the beginning was chosen so as to correspond to an average of ~20 particles per class, to elucidate the overall form of the molecule. Later, an increased number of classes were used, corresponding eventually to less than an average of 10 particles per class, in order to extract the smaller details by making the classes more representative of their constituent particles, though at some expense of the contrast in individual class averages. The latter however is more than compensated by the increased accuracy of alignment, classification, and angular assignment with progressive cycles of refinement, and the overall increased number of class averages representing a greater number of views with higher integrity. This remained the scenario until a final refinement step via projection matching was carried out.

For the BRCA2–RAD51 complex, the number of classes corresponded to an average of ~23 particles per class, and this continued to be the case until the final two cycles of refinement by projection matching. Due to slightly more heterogeneity in the dataset, this proved desirable in order to derive a ‘consensus’ reconstruction.

**Figure 13.** Multivariate statistical analysis (MSA) and hierarchical ascendant classification (HAC). MSA allows the calculation of ‘distance’ between individual particle images in multidimensional hyperspace (a), based on the eigenvalues describing each image. (b) HAC tree based on the similarity between individual particles. One can obtain the desired number of classes by trimming the tree at a specific level. Adapted from Dan Bose, PhD thesis, Imperial College London, 2005.

2.5.4.4 Class averaging

An electron microscopy dataset, even under negative stain, is comprised of a large number of relatively low contrast, low signal-to-noise ratio (SNR) images of individual particles, where the latter is defined as:

\[
\text{SNR} = \frac{\text{variance in signal}}{\text{variance in noise}}^{0.98}
\]
They hence prove problematic to judge or process, without prior application of some form of averaging procedure, to substantially enhance the SNR of the data. One is able to achieve this via the alignment and averaging of a number of similar molecular images together. The signal scales linearly with the number of images averaged; under the idealised assumption that all the images averaged are identical, i.e. apart from the noise that differs from image to image. Noise on the other hand scales as the square root of the number of images, and thus the SNR also scales as such. These class averages may then be used as references for a new multi-reference alignment followed by MSA-based classification round. After a number of iterations, ‘good’ class averages, with increased SNRs can usually be obtained. The high SNR values thus derived are of great importance to the accurate assignment of Euler angles to these images.

All images that had been assigned to the same class in the classification process were averaged together to generate class averages.

One has the option to exclude a certain fraction of the worst members from each class (by the measure of intra-class variance), at either the HAC or averaging stage. Neither was availed, as it yielded little practical benefit in our case.

2.5.4.5 Multi-reference alignment (MRA)

Following an initial reference-free alignment of the dataset, one begins to employ multi-reference alignments (MRA). These entail an alignment of the dataset to a number of individual reference images. Initially, they would be class averages resulting from the centring of particles.

On having obtained this set of class averages, one considers as varied a subset of these as possible (in terms of molecular views represented), as a template for a round of reference-based alignment. Such references should also as far as possible, constitute the strongest set of class averages (high SNRs), while also exhibiting the most detail.

The classifications’ accuracy was checked via extraction of particles within selected classes, followed by visual assessment of their consistency with the respective class averages. This was done first in IMAGIC, and then later using ‘classeditor’, a convenient program written by Dr Timothy Grant (Grigorieff Lab, Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA).

Upon alignment, the MSA process is run again, and a new set of class averages generated by classification and averaging, from which one makes a selection as previously, for yet another round of alignment and so forth. After having iterated this procedure a few times, class averages representative of various distinct views of the molecule should begin to emerge.
And following several iterations, one is supposed to have obtained a fairly varied set of class averages with more homogeneous components. These can then constitute the basis for Euler angle assignment and three-dimensional reconstruction.

The above process was repeated multiple times. In the case of BRCA2, starting with a preliminary selection of 5 class averages, which was broadened with every passing round, eventually 66 stable averages were carried forward for the purposes of creating an initial 3D reconstruction, after 5 such cycles of two-dimensional refinement.

Similarly, for BRCA2-RAD51, beginning with an initial 10 class averages (a number that was increased with progressive cycles), finally 80 stable averages were taken forward in order to generate a preliminary 3D reconstruction (following 4 rounds of 2D refinement).

However, in either case only a small subset of these class averages was used to construct the very first 3D, which was then widened in preparation for constructing the initial ‘reference’ 3D for alignment.

Once such an initial 3D is obtained, 2D reprojections of 3D models from successive refinements are subsequently used as alignment references, in order to resolve additional projections of the structure. Hereby, uniformly distributed projections within the asymmetric triangle, spaced at 10° and accounted for any symmetry at the given stage, were utilised throughout up until the round(s) of projection matching.

Extended 3D objects yield very varied projections in terms of intensity. Hence, it is necessary to normalise the references, as otherwise alignment shall be biased towards references with the greatest contrast.

All such references were therefore normalised to an average density of 0, and an arbitrary variance of 100, i.e. the same parameters as for the pre-treated images.

2.5.4.5.1 Brute-force alignment

All the reference-based alignments described hereby were performed using the brute-force, ‘timalign’ program optimised for noisy datasets (written by Dr Timothy Grant, Janelia Research Campus, HHMI, USA).

This algorithm performs a full translational alignment at every rotation of the particle along first a user specified ‘coarse’ sampling (set to be 10°), and then every 1°. In case of the BRCA2-RAD51 complex, a value of 4° was used for the former at the projection matching rounds.

The program searches for the highest cross-correlation peak over all the coarse rotations, and finally refines the rotational and translational parameters about the highest scoring rota-
tion. This is performed against each reference image (for every particle image), and the altogether best solution taken to be the optimum alignment. The biggest downside of such an approach is its CPU intensive nature. It is however most effective for sub-ideal samples.

By contrast, a ‘typical’ MRA scheme would involve a sequential, translational followed by rotational alignment (or vice versa). As a result, should the first of these be wrong, the latter might then also be incorrect. This can yield erroneous alignments, most especially for noisier datasets.

The program also always allows the original, pre-treated images to be used as the input, i.e. as opposed to the previously aligned images, and so has the advantage of avoiding interpolation errors between successive alignments. Furthermore, it does not mandatorily impose a restriction on the translation of particles, e.g. as a function of mask radius and box width.

In our cases, the fairly liberal default threshold of 20% the box inner radius was specified (corresponding to a maximum displacement of ~13 pixels or 45 Å along any direction), even more so because one’s best care had been taken in centring the particles during picking stage.

An unrestricted yet informed prescription of such a maximum shift for the purposes of alignment, can offer substantial saving of CPU-time whilst still maintaining maximal accuracy.

2.5.4.6 Euler angle assignment

Class averages resulting from alignment and classification procedures represent two-dimensional projections of the 3D structure along discrete directions. Having derived a decent assortment of these, a three-dimension structure can then be obtained if the orientational relationships between such images are known.

Two of the ways these may be deduced, are angular reconstitution by common lines\textsuperscript{102–104}, and projection matching\textsuperscript{105}.

2.5.4.6.1 Angular reconstitution (AR)

The angular reconstitution (AR) method allows one to \textit{a posteriori} determine the relative orientations of the 2D projections of a 3D structure. This is based on the ‘common lines projection theorem’, which states that two different 2D projections of the same 3D object, would always posses a one-dimensional (1D) line projection in common.

From angles between pairs of common line projections, the relative Euler-angle orientations of these projections can be calculation. The figuring of angular orientations for at least
three projections (class averages) is needed for solving the orientation problem for an asymmetric molecule.

The process begins with the generation of line projections at 1° intervals from the two input class averages, or a class average and reference, which generates ‘stacks’ of individual line projections over 360° called ‘sinograms’. A line-by-line correlation is then run between the two sinograms, which produces a maximum at the position corresponding to the common line projection. The angles of these lines thus provide the relative orientation of the two projections. By adding a third projection, and finding its relative orientation with respect to the first two, an ‘asymmetric triangle’ covering the unit sphere can be constructed. The angles of further projections are found via brute-force correlations of these projections’ sinograms with all other sinograms.

Correlation functions and coefficients have an intrinsic downside in that they are both multiplicative (or ‘squared’) operations, which cause the more powerful components of the data, i.e. the low spatial frequencies in any EM dataset, to become overinflated relative to the weaker components – the high frequencies harbouring smaller details that one is often interested in. By dividing the Fourier components by square roots of their amplitudes, these unfavourable effects can be modulated, resulting in better correlations based on fine features of the particle. This constitutes the basis for an ‘amplitude square root’ (ASQ) filter, which can be applied to sinograms, particularly during the later phases of refinement, when one wants to focus on the (weak) high resolution data components, and boost these to statistically significant levels. The advantage conferred by this practice in negative stain electron microscopy might however be insignificant.

The class averages were assigned angular reconstitution Euler angles as described above. In case of anchor-set based angle assignments, reference projections had been created at 30° intervals, and then at 10° during the advanced stages of refinement of the BRCA2 structure; and 10° throughout for the BRCA2–RAD51 reconstruction. 2° angular increments in ‘search’ were utilised, but reduced at the end to 1° in the case of BRCA2; though maintained at 2° for BRCA2–RAD51. ASQ-filtering was also employed during later stages of the former. Again, the two-fold increased sampling thereby causes a four-fold gain in processing time, but likely leads to no improvement.

However, it does make sense to use finer angular increments for search than those used to create the anchor-sets, as the process entails interpolation unlike projection matching.

All anchor-set references were normalised to an average density 0, and an arbitrary variance of 100.
2.5.4.6.2 Projection matching

In MSA/angular reconstitution based reconstruction methodologies, projection matching is often used as a complimentary technique to angular reconstitution, in order to refine the angular assignment and alignment of the structure. Thereby, uniformly spaced references are reprojected from the 3D reconstruction as elsewhere, and used to align the dataset – with all the particles aligning to a particular reference being averaged and assigned the Euler angles of that reference. As a consequence, the resulting structure is very biased toward the input 3D model, and there are few opportunities for alternative projections to diverge. However, this methodology can be very effective at the final refinement steps once a reliable reconstruction has been derived via other means.

Projection matching was used for the alignment and classification of particles, at the last stage of refinement of BRCA2’s structure, and for the last two stages of BRCA2–RAD51. In case of the former, references had been generated at 6° intervals in an asymmetric triangle arrangement, taking account of any symmetry. In the latter, first a 10° separation was used, and then 6°.

All projection-matching references were normalised to an average density of 0, and an arbitrary variance of 100.

2.5.4.7 Three-dimensional reconstruction

Following from angular assignment, one obtains a set of class averages with assigned Euler angles. In principle, these 2D class averages are then ‘back projected’ along the assigned angles, creating a sum of densities at their intersection that reconstructs the 3D volume of the molecule.

However, there is a limiting complication in the above’s implementation. 2D projections of a 3D object in real space, correspond to radial 2D slices in the 3D Fourier transform of the object. In the analytic case, projection operations through a 3D object extend from −ve to +ve infinity, resulting in infinitely thin central sections (i.e. mathematical planes) in its Fourier transform. Naturally, biological macromolecules posses defined maximum extents, and projection operations thus need only extend over this distance ‘D’. Due to reciprocity of Fourier space, central sections arising thereby have a discrete width (1/D). Unlike the analytic case, these ‘slabs’ will therefore overlap over more than a mathematical line. Owing to their diametric arrangement, this effect is more pronounced closer toward the origin in Fourier space (i.e. low spatial frequencies). Consequentially, should one be performing a 3D reconstruction by simply inserting stacks into Fourier space by back-projection of Euler-angle assigned class averages, the former shall overlap close to the centre of the 3D Fourier transform and
overweight the low frequency components, causing a burring of the map. The scale of this effect is then determined by the extent of the overlap. To counteract the problem, ‘exact filter’ algorithms can be applied, which at their most basic level constitute a high-pass filter in order to suppress the low frequencies; but are calculated separately for each input class average, and account for various factors, such as the dimensions, symmetry, and uneven distribution of views of the particle. The latter’s effects are ‘absorbable’ therein.98,107

In the case of BRCA2, a selection of three from the set of 66 stable class averages mentioned in §2.5.4.5 thought to represent roughly orthogonal views, were manually designated mutually perpendicular Euler angles and used to construct a tentative 3D model via exact-filtered back projection. Anchor references were created from this map and used to sequentially assign Euler angles to a number of other class averages in the set deemed valid representations of views of the particle, and compatible with the preceding subset, by comparison with reprojections (see later), and subsequently added to the subset of averages constituting the 3D volume.

For BRCA2-RAD51, beginning with a subset 10 of the 81 stable averages from 2D refinement used to generate the tentative 3D model, this was added to as before until a preliminary reconstruction composed of 26 was derived. Due to the bigger size, and clearer and more distinctive projections of these molecules, assigning angles to a larger number of these start-up averages proved more convenient.

MRA references were then created from these initial 3D reconstructions, and used to align the respective particle datasets. The newly aligned particles were run through MSA, HAC and averaging to generate a new set of class averages. These were assigned Euler angles according to anchor-set references from the initial 3D model, and used to construct a new model. This process was repeated till no further change could be observed in the reconstructions.

Thereby, once the dimeric nature of the reconstructions became apparent, the as-of-then unsymmetrised volumes were rotated for the top view to have Euler angles (0, 0, 0), i.e. facing down the Z-axis – about which they were then symmetrised. These models constituted the bases for the succeeding refinements, and C2 symmetry was retained therefrom. The refinement processes were continued until no further improvement could be deciphered within the reconstructions, following which projection matching was utilised to finalise the structures.

Quality of a reconstruction can be assessed by comparison of its constituent class averages with the reprojections from the 3D volume along the same Euler angle directions. At each stage averages with poor corresponding reprojections were thence excluded and the remain-
der used to build revised reconstructions. Class averages considered unfit for incorporation into the 3D maps, owing to for example, bad or poorly resolved features, low contrasts, or heterogeneous members, were excluded in the first instance. During the last couple of rounds of refinement, a quantitative but arbitrary criterion, i.e. a maximum of 40% error between average and reprojection, was utilised as a threshold for meriting inclusion into the reconstruction. By these stages, the reconstructions had improved greatly and so necessitated reduced user supervision and management.

The relative size of the object inside the 3D volume influences the shape of the exact filter. If the molecule is small compared with box dimensions, then it follows that the thickness of central slices in Fourier space is relatively large; and that these intersecting slabs influence each other over a wider frequency range. An object size as fraction of image size of 0.6 was specified throughout for the purposes of both reconstructions. This value was based on measurements of class averages, and earlier on observations of the length of BRCA2 ‘dimers’ in metal shadowing electron microscopy studies, which found it to be $270 \pm 30 \text{ Å}$ so that:

$$\frac{\text{maximum particle size}}{\text{cube side length}} = \frac{270 \text{ Å}}{\text{sampling} \times \text{box width (in pixels)}} = \frac{270 \text{ Å}}{3.52 \text{ Å} \times 128} = \frac{270 \text{ Å}}{451 \text{ Å}} = 0.6.$$  

Any point-group symmetry being utilised was also specified when performing reconstructions (and angular reconstitution), because if true this creates an advantageous redundancy in the data, which besides facilitating faster convergence of the refinements and enabling higher resolutions to be obtained, makes the overall procedure more resistant to deviations from the correct scenario.

Prior to generating either anchor-set or alignment references, the 3D volume in question was band-pass filtered between parameters corresponding to 10 and 270 Å for BRCA2; and the 3σ resolution and 270 Å for BRCA2–RAD51. The reconstructions were centred by mass within a spherical radius of 0.9, followed by 0.7 (i.e. as a fraction of the box size) – by the same rationale, and in the same manner as the pre-treated particles. The references were then normalised to an average density of 0, and an arbitrary variance of 100.

2.5.4.7.1 Euler angle refinement

After every round of MRA, MSA, classification and Euler angle assignment by angular reconstitution, some class averages would not get designated correct angles, which would lead to erroneous reprojections. Following intermediate phases of refinement of BRCA2’s structure, and from the beginning of BRCA2–RAD51’s, class averages with prefect reprojections were extracted and used to build reference models. Anchor set projections from these were used to refine the earlier Euler-angle assignments. From these refined angular designa-
tions to the same set of class averages, new, refined 3D maps were created. One again, averages with good correspondences, of which there would now be a greater number, were extracted and the process repeated as before, until the Euler angles had stabilised, and no further class averages could be incorporated into the reconstructions. Such iterative refinements of angular assignments and 3D reconstructions within stages, thus constitute mini-refinements in a larger scheme of iterative alignment-based refinements.

Prior to the creation of any such anchor set references, the 3Ds were filtered and centred as above; and the reference images normalised as usual to an average density of 0, and an arbitrary variance of 100.

2.5.4.7.2 3D masking

If desired, the reconstructed 3D image can be zeroed outside a sphere of certain specified radius, in order to mask any noise densities beyond this perimeter. The cut-off would not have a sharp edge, but rather a steep linear roll-off. For BRCA2, initially no mask was employed, followed by a ‘soft’ spherical mask of radius 0.9, and then 0.7 as a fraction of the box width. In the case of BRCA2–RAD51, one of fractional radius 0.7 was utilised throughout.

Later, one may impose a ‘fitted’ mask on the 3D. This generates a 3D mask delineating the 3D object in the 3D volume. One proceeds first via a low-pass filtering of the concerned 3D volume. The filtered volume is then binarised, such that all grey values above a certain threshold are set to 1 (white), and the ones below to 0 (black). To find the threshold for binarisation, a density (i.e. grey value) survey of the filtered volume may be run, and the 2σ value used as a threshold criterion. Sharp edges of the mask are then removed by averaging over the 3D environment of each input voxel. One’s original 3D volume is multiplied by this mask to create a binary masked 3D. The basic idea is that when ‘inside’ the object, one has a higher local variance level than ‘outside’ the object.

A low-pass parameter of 0.1 (corresponding to 70.4 Å), a density threshold of 2σ, and a 3D convolution volume with a linear dimension of 5 pixels, were used to create ‘soft’ binary masks for both reconstructions.

However, the masks were found to slightly cut into the protein densities along the sides of the molecules. To solve this, masks were ‘dilated’, whereby the original masks are isotropically ‘extended’ by a certain distance, to create a ‘looser’ fitted mask. In both our cases, the masks were dilated by 2 pixels, i.e. ~ 9 Å.

Binary masking was only performed for the purposes of either projection matching or creating the final reconstructions therefrom. In the case of BRCA2, it was employed for gen-
erating the finalised structure, after the terminal round of refinement by projection matching. For BRCA2–RAD51, it was utilised for generating alignment references for the penultimate and ultimate rounds of refinement via projection matching, as well as creating the concluding reconstruction.

Such masks can also be projected along the directions of input class averages of 3D reconstructions, and used to binary mask the former, after their re-binarisation and smoothening in 2D. This practice is roughly equivalent to solvent flattening in X-ray crystallography. In our cases, the projected 2D masks were binarised using the same 2σ value as used to binarise the 3D masks, and softened by convoluting over an area of linear size 5 pixels. They were then used to reconstruct the 3D volumes with no masking, a soft spherical mask of fractional radius 0.7, or binary masking yet again – for the purposes of resolution estimation.

2.5.4.7.3 3D map visualisation

Density thresholds corresponding to predicted molecular weights of the structures, which were then experimentally confirmed via scanning transmission electron microscopy, were calculated using a sub-routine in IMAGIC assuming a specific protein density of 0.844 Da Å⁻³ (see next). The maps were rendered in UCSF Chimera for voxels with density values greater than this threshold.

It should be noted however, that at the resolutions concerned, protein folds get ‘smoothed out’ such that only tertiary and quaternary features can be resolved. This causes an over- and underestimate for the values of volume and surface area respectively, for a set size and mass, than would be the case otherwise. To account for this issue, in actuality ‘corrected’ thresholds were calculated using 1.2× the supposed molecular mass as a guideline.

If available, the above procedures were carried out on binary masked maps, to be able to avoid the contribution of ‘mass’ by the surrounding noise affecting the accuracy of the above estimations.

Fitting of the Rad51 crystal structure into the BRCA2–RAD51 map was too performed in Chimera. The former was first interactively docked into the latter, using anti-BRC and anti-RAD51 labelling results of BRCA2 and BRCA2-RAD51 as the guide. The fit was then auto-refined therein by use of correlation coefficient as the criterion to be maximised.

2.5.4.8 Scanning transmission electron microscopy (STEM)

Scanning transmission electron microscopy (STEM) of our preparations of the BRCA2 protein and the BRCA2-RAD51 complex, were carried out at the STEM facility, Brookhaven National Laboratory, Upton, New York, USA; and the data analysed by Dr Eric
Kong, Zhang group, Macromolecular Structure and Function Laboratory, Imperial College, London, UK – in order perform molecular mass measurements of the samples as a substitute for mass spectrometry.

The principle of mass measurement by STEM is based on the incoherent nature of large-angle elastic scattering off atomic nuclei. The signal observed thereby is a sum of the scattering from all the atoms irradiated by the beam, regardless of their mutual bearings or bonding states. Scattering off atoms with different atomic numbers is very nearly proportional their respective atomic weights, and so the signal from the detector is highly proportional to the local mass traversed by the beam.

Areas of the grid with clean background, and adequate numbers of specimen, as well as tobacco mosaic virus (TMV) particles being used as an internal control and for calibration, are selected for mass analysis. Background is computed in clear areas of the image, and the signal minus background integrated over each particle and multiplied by a calibration factor, to yield a mass value.

We employed this technique mainly to ascertain the oligomeric state of our molecules. We fitted two independent Gaussians corresponding to BRCA2 and BRCA2–RAD51, and their cumulative curve to the distribution of masses, and determined the statistics therefrom.

2.5.4.9 Resolution

In electron microscopy, signal cannot directly be distinguished from noise, i.e. the extension of Fourier rings to a certain value does not imply that information is interpretable at that resolution; hence the indirect measure of Fourier shell correlation (FSC) is usually utilised in its stead, the 3D equivalents of rings in 2D Fourier space being ‘shells’.107

FSC is calculated by splitting the set of class averages constituting a 3D volume into two random halves, and calculating the correlation across the full range of spatial frequencies between the two consequent reconstructions. A typical curve shall exhibit high correlations over the low frequency components, which then decay away steeply toward the high frequencies.

Correlations between reconstructions derived from two halves of the dataset, are known to be spuriously affected by having the same mask imposed on both, either spherical or binary – especially if the two have step function drop-offs. However, equally the surrounding ‘solvent’ region does not comprise one’s particle, but rather extra-molecular noise. The latter is significant because its masking therefore does not impact upon inside noise variations, and influence information contained in the object itself. Since our reconstructions did not employ ‘hard’ masks, the latter effect most probably compensates for the former.
Nevertheless, as both the half-sets of class averages were still ultimately derived from data aligned to the same set of references, the resulting FSC is thus regarded as ‘semi-independent’. Truly independent FSC calculations necessitate generating two separate reconstructions from data divided into two halves at the outset, the so-called ‘gold-standard'\textsuperscript{109}. The latter is then justifiably accompanied by the use of more lax resolution criteria such as 0.143. However, this practice is not usually considered worthwhile for negatively stained electron microscopic reconstructions.

The global resolution of a reconstruction is estimated by the intersection of its FSC curve, with curves defining set resolution criteria, such as 0.5, 0.143, 3σ and \( \frac{1}{2} \)-bit\textsuperscript{110}. The former has a constant value of 0.5 over all spatial frequencies, corresponding to the point where the frequency-dependent signal-to-noise ratio of the map, \( SNR_{MAP}(v) \) equals \textsuperscript{111}, as in:

\[
SNR_{MAP}(v) = \frac{FSC(v)}{1 - FSC(v)},
\]

so that when \( FSC(v) = 0.5 \):

\[
SNR_{MAP}(v) = \frac{0.5}{1 - 0.5} = \frac{0.5}{0.5} = 1;
\]

where \( v \) denotes spatial frequency (resolution), and \( FSC(v) \) is the resolution dependent FSC. This criterion is the strictest and most commonly employed, but is known to be overly conservative, e.g. for gold-standard reconstructions but also for less symmetric particles, where useable information is present beyond this threshold\textsuperscript{110}. The 0.143 criterion is based on the estimated correlation between a map calculated from the full dataset (by averaging the two half-maps), and a ‘perfect’ reference containing no noise; and aims to make FSC comparable with the crystallographic figure-of-merit (FOM)\textsuperscript{111}. When the two halves of the data are combined to produce the best estimate of the structure, the SNR of the full average is increased by a factor of \( \sqrt{2} \) relative to the half-reconstructions. This results in a scaled FSC (\( FSC' \)) as follows:

\[
FSC' = \frac{2FSC}{1 + FSC};
\]

\( FSC \) being the ‘old’ FSC\textsuperscript{111}. Therefore, if \( FSC = 0.5, FSC' = 0.67 \); but if \( FSC' = 0.5, FSC = 0.33 \). Comparing this merged reconstruction with noise reduced by \( \sqrt{2} \), against a perfect reference where noise equal zero, results in a correlation (\( C_{\text{ref}} \)) given by:

\[
C_{\text{ref}} = \sqrt{\frac{2FSC}{1 + FSC}};
\]
where $FSC$ is again the old FSC$^{111}$. It can also be shown that $C_{\text{ref}}$ is equal to the cosine of the average phase error, and equivalent to FOM (a common measure of map interpretability in x-ray crystallography). Thereby, if $C_{\text{ref}} = 0.5$, $FSC = 0.143$. In analogy to crystallography, a FOM of 0.5 is commonly used to mean that the map is interpretable, as in that a molecular structure can be built. Other criteria measure the point at which the FSC crosses the level that would be expected of pure Gaussian (i.e. random) noise$^{110,111}$. The latter is given by $1/\sqrt{N}$, where $N$ is the number of voxels in a given resolution shell. A value of two or three times this figure is then commonly utilised, making up the $2\sigma$ and $3\sigma$ criterions respectively. These criteria are therefore true ‘curves’, and attempt to denote the point at which the signal in a reconstruction lies significantly above the noise floor, but are not in common use at present.

Naturally, various parts of a reconstruction do not all conform to the same resolution. However, assigning a single resolution value to the map where the FSC descends under a certain threshold, provides a convenient means of tracking one’s progress over the course of refinement.

Following from intermediate stages of the BRCA2 reconstruction, and the beginning of BRCA2–RAD51’s, resolution was charted using FSC and the various threshold criteria. Any point-group symmetry utilised for the reconstructions was specified, which in order to account for its influence, multiplies the $N\sigma$ and $\frac{1}{2}$-bit threshold curves by the square root of the number of asymmetric units for this cyclic symmetry; as then the number of independent voxels in Fourier space 3D shells is thus lower by the degree of symmetry, necessitating its division with this number$^{110}$. A threshold of 3 was specified for the sigma curve throughout. The size of the molecule within the 3D reconstruction volume also affects the expected FSC levels. Thereby, the smaller the former is inside the latter, the larger is the corresponding ‘convolution’ sphere in Fourier space whereby the complex structure factors are highly correlated. A ‘filling degree’ of 0.6 was hence specified always.

2.5.4.10 Structural validation

The overall correctness of the two structures was established by a variation of the tilt-pair and tilt-series validation methods, as described by Scheres$^{112}$ (Nat Struct Mol Biol, 2011).
The particle sets used in this analysis each constituted a tilt series ranging between nominal tilts of 0° and 40°, in steps of 5°. Each particle was independently assigned Euler angles using the respective model’s projections. The particle with a nominal tilt of 0° was then used as a reference to calculate all the other relative tilt transformations in pairs, as described below.

The ‘standard’ spherical polar coordinate system (Fig. 14, left) is defined by the ‘azimuthal’ angle \( \theta \) in the \( x-y \) plane from the \( x \)-axis, the ‘polar’ or ‘zenith’ angle \( \phi \) from the positive \( z \)-axis (so that \( \phi = 90° - \delta \), where \( \delta \) is the ‘latitude’), and the distance or radius \( r \) from the origin to a point, i.e.:

\[
(r, \theta, \phi), \text{where} \quad 0 \leq \theta \leq 360°, \text{and} \quad 0 \leq \phi \leq 180°.
\]

In the IMAGIC convention (Fig. 14, right), the azimuth \( \theta \) is defined as \( \gamma \), and the zenith \( \phi \) defined to be \( \beta \), such that by the above order:

\[
(r, \gamma, \beta), \text{where} \quad -90 \leq \gamma \leq 90°, \text{and} \quad -180 \leq \beta \leq 180°.
\]

But for the special case of C2 symmetry:

\[
0 \leq \beta \leq 180°.
\]

As per the above equivalencies between standard and IMAGIC angles, and the known relationships of Cartesian to spherical coordinates, one can then write:

\[
x = r \cos \theta \sin \phi = r \cos \gamma \sin \beta,
\]

\[
y = r \sin \theta \sin \phi = r \sin \gamma \sin \beta, \text{and}
\]

\[
z = r \cos \phi = r \cos \beta.
\]

Considering vectors \( \vec{v}_1 \) and \( \vec{v}_2 \) for two equidistant points on the sphere \( P_1 \) and \( P_2 \) respectively, one has:

\[
\vec{v}_1 = (r, \theta_1, \phi_1) = (r, \gamma_1, \beta_1), \text{and}
\]

\[
\vec{v}_2 = (r, \theta_2, \phi_2) = (r, \gamma_2, \beta_2).
\]

Writing these in terms of Cartesian coordinates:

\[
\vec{v}_1 = (r \cos \gamma_1 \sin \beta_1)\hat{x} + (r \sin \gamma_1 \sin \beta_1)\hat{y} + (r \cos \beta_1)\hat{z}, \quad \text{and}
\]

\[
\vec{v}_2 = (r \cos \gamma_2 \sin \beta_2)\hat{x} + (r \sin \gamma_2 \sin \beta_2)\hat{y} + (r \cos \beta_2)\hat{z}.
\]

- 70 -
One can then reckon the scalar (dot) product of the two vectors, as a product of their magnitudes $|\mathbf{v}_1|$ and $|\mathbf{v}_2|$, and the cosine of the angle $\tau$ between them. This would be the angle they subtend at the centre of the sphere, and hence the ‘tilt angle’ between a pair of Euler angles. Explicitly,

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = |\mathbf{v}_1||\mathbf{v}_2| \cos \tau = r^2 \cos \tau,$$

$$\therefore |\mathbf{v}_1| = |\mathbf{v}_2| = r.$$

By expansion, one therefore has:

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = r^2 (\cos \gamma_1 \sin \beta_1 \cos \gamma_2 \sin \beta_2 + \sin \gamma_1 \sin \beta_1 \sin \gamma_2 \sin \beta_2 + \cos \beta_1 \cos \beta_2).$$

Factorising,

$$= r^2 [(\sin \beta_1 \sin \beta_2 (\cos \gamma_1 \cos \gamma_2 + \sin \gamma_1 \sin \gamma_2)) + \cos \beta_1 \cos \beta_2].$$

Using the trigonometric identity whereby $\cos (A) \cos (B) + \sin (A) \sin (B) \equiv \cos (A - B)$, one gets:

$$= r^2 [\sin \beta_1 \sin \beta_2 \cos (\gamma_1 - \gamma_2) + \cos \beta_1 \cos \beta_2].$$

Equating,

$$r^2 (\sin \beta_1 \sin \beta_2 \cos (\gamma_1 - \gamma_2) + \cos \beta_1 \cos \beta_2) = r^2 \cos \tau,$$

$$\{\sin \beta_1 \sin \beta_2 \cos (\gamma_1 - \gamma_2) + \cos \beta_1 \cos \beta_2\} = \cos \tau.$$

Therefore, the tilt angle $\tau$ between two views with Euler angles $(\beta_1, \gamma_1)$ and $(\beta_2, \gamma_2)$ is expressed as:

$$\tau = \cos^{-1} [\sin \beta_1 \sin \beta_2 \cos (\gamma_1 - \gamma_2) + \cos \beta_1 \cos \beta_2].$$

The reconstructions were validated by comparing the expected versus the measured tilt of each of the (nine) pairs in the series, calculated as per the formula derived above. Angles were converted from the IMAGIC $(\alpha, \beta, \gamma)$, to the SPIDER/FREALIGN convention $(\psi, \theta, \phi)^{113,114}$, for the purposes of determining out-of-plane errors in the tilts in TILTDIFF as follows:

$$\psi = \alpha + 90^\circ,$$

$$\theta = \beta, \text{ and}$$

$$\phi = 270^\circ - \gamma.$$

A value of $360^\circ$ was added to and subtracted from, negative and positive angles respectively, so as to all conform to the range $0^\circ \leq \psi, \theta, \phi \leq 360^\circ$.

Whilst it is possible to obtain perfectly valid electron microscopy reconstructions from unilted negatively stained particles, tilted views often tend not to appear as expected for a particular angle of tilt, due to the effects of stain, especially that surrounding the molecules.
The phenomenon of ‘flattening’ of the particles perpendicular to the carbon support film is also often quoted\(^\text{115}\), and believed to be a consequence of the heavy-metal staining. Some sources espouse the severity of this effect to be as much as 50\%, or even up to 70\%.

Furthermore, in the specific case of BRCA2, the reconstruction does not possess especially distinct projections over a relatively wide range of angles, further compounding the procedure’s difficulty.

We also analysed \(-20^\circ/20^\circ\) and \(0^\circ/20^\circ\) tilt-pairs for both datasets. Euler angles were assigned for either tilt in the tilt-pairs by projection matching in SPIDER (and the correct assignments visually adjudged by comparison with model projections). Tilt transformations were computed via their substitution into the aforementioned formula. As the more familiar IMAGIC alignment references were employed, conversion of the angles from SPIDER format was not necessitated. References had been projected from binary masked final reconstructions, accounting for C2 symmetry, and normalised. Since the nearest view can be across the other half of the Euler sphere, the angles were translated using the symmetry relationship for an object with two-fold rotational symmetry, so as to always correspond to the shortest angle between them. For such an object in IMAGIC, every projection at \((\beta, \gamma)\), has an equivalent projection at \((\beta, \gamma - 180^\circ)\) for positive \(\gamma\); and \((\beta, 180^\circ + \gamma)\) for negative \(\gamma\).

While the results were confirmatory, i.e. the mean tilt values lay close to the applied tilts; the associated percentage errors were large (more so for the smaller-angle \(0^\circ/20^\circ\) tilt-pairs) – see §3.2.6. However, amongst the distribution of tilt-angle values for each, there were somewhat tight clusters with lesser errors.

The tilt-pair series methodology as in the negative stain electron microscopy study of human telomerase (also a dimer)\(^\text{112}\), proved to be much more successful in our case, than the usual approach of analysing a set of tilt-pairs with the same angle. This is conceivably because while certain individual particles can be free of staining artefacts, enabling their analyses as a series of tilt pairs; in a collection of tilt-pairs, several such pairs might be affected by the skewing effects of negative stain.

Even though the grids were adjusted for eucentric height in the electron microscope at high magnifications, owing to an inherent ‘wobble’ in the stage/grid-holder, the overall defoci of images would alter upon application of tilts, i.e. move closer to focus with increasing angles. Therefore, the defoci were manually compensated in the TEM to be the same as the untilted images to a good approximation.

All these data were collected at the FEI Tecnai F20 TEM, at the Centre for Biomolecular Electron Microscopy, Imperial College London, UK. And since its detector yields a lesser pixel size for the same magnification, consequently causing the particles to appear smaller;
the 3D models (derived from CM200 data) were scaled by the requisite factor (i.e. 3.52 Å pixel\(^{-1}\) / 4.10 Å pixel\(^{-1}\) = 0.86 or 86%), and the volumes padded to have the same box side as the particles (hereby 162 pixels) by addition of empty voxels. References were then created from these models as per usual.

\[
\begin{align*}
(r, \theta, \phi); & \text{ where} \\
0^\circ & \leq \theta \leq 360^\circ \\
0^\circ & \leq \phi \leq 180^\circ
\end{align*}
\]

\[
\begin{align*}
(r, \gamma, \beta); & \text{ where} \\
-90^\circ & \leq \gamma \leq 90^\circ \\
-180^\circ & \leq \beta \leq 180^\circ
\end{align*}
\]

**Figure 14.** Comparison of the standard and IMAGIC spherical coordinate systems. The zenith \(\phi\) in the former corresponds to \(\beta\) in the latter, and azimuth \(\theta\) is equivalent to \(\gamma\). However, whilst in the standard convention the azimuth covers a full rotation (\(0^\circ \leq \theta \leq 360^\circ\)), and zenith covers half (\(0^\circ \leq \phi \leq 180^\circ\)); in IMAGIC the zenith encompasses \(360^\circ\) (\(-180^\circ \leq \beta \leq 180^\circ\)), and azimuth \(180^\circ\) (\(-90^\circ \leq \gamma \leq 90^\circ\)). Although in the case of C2 molecules, the former too is then obviously restricted to \(180^\circ\) (\(0^\circ \leq \beta \leq 180^\circ\)). IMAGIC incorporates another angle \(\alpha\) representing the inplane orientation of the particles for a fixed \(\beta\) and \(\gamma\), and ranging the full rotation. Radius \(r\) denotes distance from the centre, but is made redundant for most angular calculations. In SPIDER/FREALIGN formats, the angles then relate to IMAGIC’s as follows: \(\psi = \alpha\), \(\theta \equiv \beta\), and \(\phi \equiv \gamma\), i.e. the same as standard\(^{1,3,11}\) (see text for details, §2.5.4.10).

2.5.4.10.1 Handness

Beyond accurately determining orientations for structural validation, elucidating the handedness of a reconstruction in negative stain can be more intractable.
We proceeded by first determining the approximate tilt axis of our images. This could be deduced by comparing untilted and tilted micrographs, and observing the direction in which the field of view would get ‘compressed’ in the latter. In our case this was horizontal, implying a tilt axis that was roughly vertical. We then compared the defocus on the left and right halves of the micrographs; for tilted micrographs the former would be less underfocused than the latter, and hence up and down respectively. This was confirmed by displacing an underfocused, untilted image up- and downwards in the TEM and observing the change in defocus. There was no difference between the top and bottom halves.

Having determined the tilt direction, we compared the surface views from both unmirrored and mirrored models that gave rise to the (equivalent) projections as seen in the tilt-series (and -pair) validations. The object being to identify the mirror whose tilt direction was commensurate with that of the micrograph.

The results thereby were statistically inconclusive, with slightly better than even splits for the hands depicted in Fig. 19 from the pairs/series analysed (see §3.2.6 for numbers on the latter). The task was made more difficult by the fact that the BRCA2 reconstruction harbours pseudo-mirror symmetry within its dimeric structure. Further ambiguity is added at the resolutions concerned, as the mirrors can tend to look similar.

Hence, we predicated as to the handedness of the structures, by selecting the mirror of the BRCA2-RAD51 reconstruction that was most compatible with the fitting of Rad51’s crystal structure. Following therefrom, we then based the handedness of the BRCA2 reconstruction on the mirror that corresponded best with the aforementioned handedness of the BRCA2-RAD51 structure, again on the premise of fitting (see §3.2.8).

2.5.4.11 Antibody complexes

Owing to the large mass of HsBRCA2, and challenges in deriving the requisite amounts of protein, we were not able to employ internal labelling and nanogold particles, to be able to identify specific regions within the reconstructions.

We therefore utilised monoclonal antibodies, their distinctive overall Y-shapes being visible in negative-stain electron microscopy; and often also the constitutive globular lobes, when viewed uncomplexed on their own (see Fig. 22a).

Guided by this shape, we selected individual antibody-bound particles. The antibodies were then masked-out from the particles, by making soft hand-drawn binary masks around the particles. This was to avoid angular misassignment of the particles, due to the attached antibody densities.
The particles had been low-pass filtered to 60 Å (and a high pass of 270 Å), to facilitate their Euler angle assignment, i.e. through enhancement in SNR afforded by suppression of noisy high frequencies, thus assisting angular designations based on overall shape.

A series of closely spaced dots were placed around the perimeter of the particles, and used to create initial sharp-edged masks. These were then smoothed by a low-pass parameter of 0.1 (~70 Å), and multiplied with the original particle images, to yield antibody-bound particles with soft binary masks.

The particles were assigned Euler angles by matching against projections from the respective 3D models. Reconstructions were rotated in UCSF Chimera via the command line, to exactly correspond to each of these views in turn. In every instance, a cylinder was placed perpendicularly into the screen, touching the surface where the antibody appeared to make contact with the particle in its projection image. This helped localise the antibody’s position in two-dimensions at each step (as should be the case, since each particle is a 2D projection image). After repetition with multiply oriented antibody-bound particles, one obtained a substantial overlap of these cylinders, indicating the antibody’s location in three-dimensions. Cylinders of diameter 20 pixels had been employed to account for error in individual placings. For a visual illustration, see §3.2.7 and Figs. 22b-d.

We deployed this technique, in our case because of an inability to derive class averages, owing to the sparsity of antibody-bound particles. Thence, while each separate assignment has an associated uncertainty (encapsulated in the cylindrical width), and only defines the position of the antibody in two-dimensions; a series of such assignments reduces this error, and pinpoints its location in three-dimensions.

The methodology was utilised for each of the three antibody complexes, i.e. anti-BRC and (C-terminal) -Flag antibodies against BRCA2, and the anti-RAD51 antibody against BRCA2–RAD51, in order to identify these domains within the respective reconstructions. In the latter, as there were multiple targets for the antibody, particles with several antibodies bound were frequently observed.

For BRCA2–Ab1 (anti-BRC) complex, an initial selection of 86 particles was made. Of these, the eight very best in terms of clarity were incorporated into the analysis in Fig. 22c. For the BRCA2–anti-CFlag complex, a total of 65 antibody bound particles were selected. These were first separated into nominally ‘good’ (22), ‘average’ (23) and ‘bad’ (20) particles. Similarly to BRCA2-Ab1, eight of the most clear were then utilised for analysis in Fig. 22b. For BRCA2–RAD51–anti-RAD51, three good and 43 average particles were preliminarily identified. From amongst these, two of the former and eight of the latter were finally employed in the analysis (Fig. 22d).
The reconstructions had been low-pass filtered to 60 Å for creating reference projections, and 10 Å for generating surface views.

2.5.4.12 gDNA complexes

To outline the trajectory of DNA across BRCA2, i.e. in its complex with single-stranded DNA, we exploited its binding to gapped DNA substrates, as and for the reasons described in §2.2.

Particles of BRCA2-gDNA complexes, showing clear DNA density flanking a central protein density, were selected for analysis. Since binary masks were not necessitated hereby, a soft circular mask of inner radius fraction 0.9, and an 8-pixel drop off was used. The images were band-pass filtered between parameters of 60 and 300 Å, former to aid visibility of the DNA strands, and the latter to account for added length due to DNA.

Euler angles were designated using matching model projections of the BRCA2 reconstruction. The corresponding surface views were oriented in Chimera as previously, and DNA traced thereon as per its track in the particle projections (see §3.2.10 and Fig. 24b).

A total of 77 eligible BRCA2-gDNA particles were boxed. These comprised 48 from the 100 nt, and 29 from the 70 nt gapped DNA complexes; of which the 29 most clear-cut cases were used in the analysis, including 15 from the former, and 14 from the latter dataset respectively.

As BRCA2-gDNA data were collected at the F20 TEM, the BRCA2 3D reconstruction was appropriately scaled and padded like before prior to generation of angular assignment references and surface views; and low-pass filtered to 60 Å for purposes of the former, and 10 Å of the latter.

2.5.4.13 BRCA2–RAD51+ssDNA

We sought also to delineate the trajectory of nucleated RAD51-ssDNA filaments from BRCA2 and analyse where on its surface they emerged.

Hereby, the complexes were boxed into size 180 × 180 squares to better accommodate the filament, though this needed not be the full filament.

Similarly to gDNA complexes, the central particles were matched to the best-fit BRCA2 model projections, albeit the projections were then rotated to match the particles (and not vice versa), so as to correspond to their original orientations in the micrograph. Surface views corresponding to the rotated projections were rendered in IMAGIC at the threshold discussed in §2.5.4.7.3.
Initially, approximate points were placed on the projections and surface views, where it was thought the filaments were most likely originating. Thick lines denoting RAD51-ssDNA filaments were then traced onto the surface views, based on matching projections with the dot markers, and original particles with issuing filaments (for subsequent interpretation) – see §3.2.11.2.1 & Fig. 28.

This (tentative) analysis was performed using seven micrographs of the type in Fig. 27c derived from the West laboratory, and a sum total of a mere four particles that had the ‘clearest’ appearance.
3 Results

3.1 Purification

Human BRCA2 protein was purified, and the human BRCA2–RAD51 complex constituted and purified, as described in §2.1.1 and §2.1.2 respectively. The samples were soluble, and their analysis by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and silver staining, revealed the preparations to be of good quality (see Fig. 15a).

Since PALB2 had previously been found to be associated with affinity purified N-terminally tagged BRCA2, but not with the C-terminally tagged protein, the latter was employed throughout in our single-particle electron microscopy and other analyses. Thereby, the residual compositional heterogeneity visible in the SDS gel is due mainly to non-specific degradation of the large BRCA2 protein, as confirmed through mass spectrometry by the West laboratory. This is compatible too with the presence of smaller particles with self-inconsistent sizes and shapes observed under negative-stain EM (Fig. 16, left).

Above affirmations are valuable as affinity steps were performed in low salt buffers that would favour BRCA2’s association with its partner proteins, and potentially skew various subsequent results such as those of DNA binding via contribution of their own respective activities.

As before, BRCA2 was found to stimulate RAD51–ssDNA nucleoprotein filament formation, as determined by gel-shift assays (Fig. 15b). In its sole presence, only ssDNA was detectable. The BRCA2–ssDNA complex was moderately apparent in BRCA2’s presence (sans RAD51), as opposed to RAD51–ssDNA that was relatively weakly visible in RAD51’s (sans BRCA2). However, BRCA2–ssDNA and RAD51–ssDNA complexes were strongly obvious in the sum presence of BRCA2 and RAD51, and unbound ssDNA notably absent. It should be noted that a ternary BRCA2–RAD51–ssDNA is thought not to exist per se, with BRCA2 dissociating from RAD51 and ssDNA on loading former upon the latter and facilitating initial filament growth, hence the absence of a third ‘super-shifted’ band.

The binding of BRCA2 to gapped DNA substrates, i.e. as opposed to single-stranded DNA, was also established through electromobility shift assays (EMSAs) – see Fig. 15c.
**Figure 15.** Purification of BRCA2 and BRCA2–RAD51, and binding to single-stranded DNA. (a) SDS-PAGE showing purified proteins used in electron microscopy analyses. (b) BRCA2 promotes RAD51–ssDNA complex formation as seen by gel shift assays. (c) BRCA2 binds gapped DNA, as determined by gel-shift assays.

### 3.2 Negative stain electron microscopy

#### 3.2.1 Micrographs and particles

As BRCA2 was expressed from its endogenous promoter at *in vivo* levels (500 ng from 20 L of cells), negative stain electron micrographs of our preparations were generally quite scant in terms of particles (see Fig. 16 - left), which precluded them from analysis by cryo-electron microscopy that necessitates significantly higher concentrations of samples to be practical.

The micrographs also showed visible aggregation, and wisps of smaller particles that exhibited no mutually consistent size or shape, facilitating their ready exclusion as degradation products (i.e. as opposed to valid and intact subpopulations).

The 8,873 and 6,877 particle projections respectively of BRCA2 and the BRCA2–RAD51 complex so-picked (Fig. 16 - right), exhibited a respective ‘long’ and ‘short’ dimension of ~ 120 Å and 250 Å in the former case, which were then larger in the latter case especially as to the short dimension (by ~ 40 Å).

Interestingly, we observed a propensity for ‘dimeric’ shapes, broadly similar to previous rotary shadowing electron microscopy visualisations of BRCA2. The observation then that the sizes of BRCA2–RAD51 particles was consistent with these, suggested that at the very least RAD51 binding did not dissociate the BRCA2 ‘dimers’.

Either of the particles was rather elongated, with those of BRCA2–RAD51 being ‘wider’ in the other two dimensions. This lent a ‘dumbbell’ or ‘kidney-bean’ appearance to BRCA2
(Fig. 16a - right), whereas a more ‘heart-’ or ‘V-shaped’ appearance to BRCA2–RAD51 (Fig. 16b - right).

With regard to antibody labelling, complexes were constituted using low concentrations of antibodies (i.e. 10–30 nM, with ~5 nM BRCA2), in order not to saturate the micrographs’ background with antibody particles – see §2.4. As a result, only a fraction of the observed particles had visibly bound antibodies (~5%), and in turn the number of particles where both ‘monomers’ had antibody attached was exceedingly low. The situation was slightly different in case of BRCA2–RAD51 particles as each particle presumably had multiple sites to which (anti-RAD51) antibodies could bind.

In any case, having a singular antibody bound per particle, enabled easier adjudgement of the antibody’s point of binding on the particle’s projection, by eliminating confusion arising from several attached antibodies.

Likewise for BRCA2-gDNA complexes, we detected approximately one particle with DNA bound per scanning every two frames, further compounded hereby by the inevitable difficulty in identifying and unambiguously tracing naked DNA under negative stain conditions, especially in complex with proteins.

Incidentally, DNA is better observed (with higher contrast), in cryo-rather than negative stain EM, due to lack of interference by heavy metal ions or ‘masking’ by uranyl acetate nano crystallites.
Figure 16. Example transmission electron microscopy captures of negatively stained BRCA2 (a) and BRCA2–RAD51 (b) samples, and selections of particles therefrom. Whilst some heterogeneity in terms of aggregates and degradation products can be seen, the presence of a homogenous species therein is evident. With BRCA2 a preponderance of dimeric shapes is observed (a, right), which then remains consistent with BRCA2–RAD51 (b, right), except that the latter appear ‘fatter’.

3.2.2 MSA and classification

As can be seen in the predominant five eigenvectors of the centred dataset of 8,773 BRCA2, and 6,877 BRCA2–RAD51 particle images (Fig. 17a, b); a stark 2-fold component is obvious in the second and third eigenimages of the former, and the second, third, and fourth such images of the latter that is suggestive of a dimeric species.
It is worth reiterating here that the first eigenvector represents the sum of a collection of particle images, which in the above cases had only been subject to ‘centring’ (i.e. reference-free alignment), and thus appears as a ‘disc’ since the particles’ rotational orientations had remained arbitrary. Its diameter however can be used as measure of the particle’s size, and is visibly larger for BRCA2–RAD51 than for BRCA2. The observation that a two-fold rotational symmetry was apparent in the very next few eigenimages, was indicative of the specimens’ oligomerisation state being that of a dimer. Nonetheless, symmetrical eigenimages can be incidental and/or spurious, and hence their implication was treated with caution and not presumed.

Following initial reference-based alignment, the first eigenvector appears increasingly characteristic of the particle’s shape (including in this case exhibiting two-fold symmetry) as opposed to merely its size, as the particles get rotationally as well as translationally aligned.

A two-fold symmetry in the molecules, was also apparent in the MSA and HAC derived class averages, i.e. prior to either commencing 3D-based MRA, or imposition of C2 symmetry on the structures.

It should noted that two-dimensional projections of a three-dimensional object with 2-fold cyclic symmetry, do not possess two-fold symmetry in themselves, unless either along or perpendicular to the symmetry axis of the 3D object, in which case they will display 2-fold rotational symmetry and mirror symmetry respectively. Though one could detect clear two-fold symmetry in class averages along such angles; it is usually also intuitively apparent from ‘intermediate’ views if the projections are derived from a three-dimensional molecule with C2 symmetry.

As rounds of MRA, MSA, and HAC based classification progressed, class averages depicted visible improvement in terms of SNR, the variety of views depicted, intra-class variance, (i.e. becoming gradually more faithful), and observable detail – that are indicative as to an improvement along the refinement process.

Viewing individual particles constituting the class averages, one could see that by the final rounds of MRA followed by MSA and HAC (or projection matching) based classification; particles comprising most individual classes were quite homogeneous and resembling of the averages (see Fig. 17c, d), reflective of more accurate alignments and convergence of the refinement procedures – especially given the low SNRs of most particles images and artefacts of negative stain.
Figure 17. Multivariate statistical analysis and classification. The five predominant eigenvectors of the centred dataset of 8,773 BRCA2 (a), and 6,877 BRCA2–RAD51 particles (b). Images 2 and 3 in the former, and 2, 3 and 4 in the latter show clear 2-fold symmetry, suggestive of such a component within the particles themselves (i.e. dimers). These in fact constitute the first non-trivial vectors of the data, as the very first eigenimage denotes the total average of the particle images (and hence looks circular). Image 5 in either case might be indicative of the sub-optimal centring of a subset of particles. The latter is then larger for BRCA2–RAD51 than BRCA2 reflective of its bigger size. From toward the ending stages of BRCA2 (c) and BRCA2–RAD51 (d) refinements, a selection of 10 class averages (top) and 5 individual particles from each (below), reveals the former to be representative of the latter (i.e. not a spurious summation of noise and/or inaccurately aligned/classified particles).

3.2.3 Angular assignment and particle alignment

As rounds of MRA and MSA/HAC-based classification followed by averaging progressed, anchor-set based Euler angle designations of the class averages improved alongside, i.e. the percentage error in individual angular assignments decreased, whilst the spread of ‘views’ widened, and the averages’ α angles settled close to 0° – a good signifier as to the in-plane parameters being eliminated through repeated alignments. All such developments constituted indicators as to an improvement along the refinement process.
Toward the end, angular reconstitution Euler angle designations of the class averages were accurate and precise enough so as not to necessitate significant adjustment and fine-tuning via miniature refinements (as in §2.5.4.7.1), yielding increasingly little change or improvement. Their extent was therefore reduced during these stages.

At the end of this phase for BRCA2, the average angular reconstitution error of the set of 1000 class averages that were created was 21.31%, and 18.55% for the 80% best. The angles were very well distributed over the Euler sphere by the expectations of negative stain electron microscopy (see Fig. 18a), with only a handful of devoid views – principally a cluster situated close to (90°, 0°), and to lesser extent (135°, 0°) and (45°, 0°), i.e. the ‘head-on’ side view, and 45° tilts from the equator toward the north and south ‘poles’. These represented 387 of the above averages that were used to generate the reconstruction, which was then carried over to the final projection matching step.

The spread was even more uniform for BRCA2–RAD51 (Fig. 18b), with no notably absent views. Hereby, they represented 211 of the 294 class averages used to create the reconstruction that was taken forward to the first of the projection matching steps.

Adoption of ‘preferential orientations’ by fixated molecules can be a concern in negative stain EM, especially for elongated molecules such as the ones here. The presumption being that these particles would ‘roll’ along their long axes, and for instance prefer not to ‘stand’ on their ends. This situation is clearly avoided in cryo-EM, whereby particles are flash frozen in random orientations. And indeed a vastly uneven distribution of orientations can cause artifacting in the structure, as a result of poor sampling. However, it should be considered that gravity has negligible influence at molecular scales, which would facilitate such a state of affairs. Rather, the distribution of charge over the molecule and the grid’s carbon support should wield by far the greatest effect. Therefore, it would not seem reasonable to expect particular views on purely geometric bases.

At the projection matching stages, there were no references to which any particle failed to align; i.e. for all 578 angular orientations at 6° separations over (half) the Euler sphere in case of BRCA2, and 214 followed by 578 at 10° and 6° separations respectively in the case of BRCA2–RAD51. However, naturally some of the resulting classes were more ‘populated’ than others.

Because of this, projection matching derived class averages shall evidently be evenly spaced in angle, as the unequal distribution of particles within the classes shant be reflected. As a result, angular maps so obtained are perhaps less meaningful than those using angular reconstitution, as the latter works by interpolation and consequently does not ‘force’ particles into predefined Euler angles.
Nonetheless, the failure of no one particle to not get assigned to a particular reference out of up to 578, was indicative of a good spread of views amongst the particle datasets.
Figure 18. Distribution of Euler angles of the 387 and 211 class averages of BRCA2 (a) and BRCA2–RAD51 (b) respectively, which constituted the last 3D reconstructions derived via MSA-based classification, averaging and anchor-set based angular assignment. Due to C2 symmetry only half the Euler sphere is relevant and is shown. A fairly even spread of views is apparent in either case, with only comparatively scarce coverage in a few areas for BRCA2; whilst no regions like such are obvious in the case of BRCA2–RAD51. It should be minded that as a price paid for 2D projection of the Euler (hemi)sphere, the scaling is non-linear and much more compressed toward the periphery than the middle, which exaggerates the angular spacings closer thereto.

3.2.4 Structures of BRCA2 and BRCA2–RAD51

As 3D refinements advanced, the reconstructions exhibited improvement in a number of respects. The agreement between class averages and their corresponding reprojections became more exact, and hence the correlation error between the two decreased; whilst the observable detail amongst them increased. A greater percentage of averages as well as number of particles could be incorporated in the structures, drawing on a bigger proportion of the image datasets, i.e. either by visual adjudication, or a criterion such as an upper bound on the 3D error between class average and reprojection for meriting inclusion.

Despite the additional inputs, there was an accompanying reduction in the average correlation error in the 3D reconstructions, all of which could be considered signifiers of a globally convergent refinement in entropy space.

For BRCA2, 387 of the 1000 class averages were included in the last 3D reconstruction derived by MSA/HAC-based classification and averaging, and assignment of Euler angles by angular reconstitution. This model possessed an overall 3D error of 35.69%, i.e. the mean correlation error between the individual averages and their reprojections, and was carried over to the projection matching stage.

The concluding structure obtained after projection matching (with binary masked class averages), incorporated 568 out of the 578 averages, i.e. the vast majority of the particle dataset, and possessed an average correlation error in the map of 18.22%. All of these class averages displayed excellent, i.e. close to perfect, correspondence with their respective reprojections (see Fig. 19a, ii).

The reconstruction itself shows an elongated molecule, with maximum dimensions of ~250 Å (length) × 135 Å (width) × 120 Å (height) – see Fig. 19a, i. This comprises main densities that form smooth arcs at the far sides of the molecule, and antiparallel elongated densities along the side rims that connect the two halves and cage a central space, as seen
from the concave face (i.e. top view). This lends the molecule a distinctive overall ‘jelly
bean’ appearance, which is apparent in the side view and tilts thereof, i.e. with a groove in
the middle, but also a relatively flat base. When observed from the bottom view, one sees
densities from the adjacent halves enclosing a central rectangularly shaped ‘channel’. The
two halves corresponding to monomers themselves resemble ‘scoops’ and pack together in a
‘yin-and-yang’ arrangement to constitute the dimer.

The binary masked structure, at the density rendering threshold that yields 1.2× the cor-
correct mass (for the reasons in §2.5.4.7.3), provides a volume of 1,090 nm³, and a surface area
of 1,430 nm². Whilst this probably more or less compensates for the former, the latter likely
remains a marked underestimate due to the absence of secondary structure elements at the
resolutions concerned.

In the case of BRCA2–RAD51, ultimately 211 out of 294 class averages were included in
the reconstruction derived via MSA, HAC and averaging, and anchor-set based angular as-
signment. This 3D had a mean correlation error between the averages and their reprojections
of 24.22%, and was carried forward to the next (projection matching) step.

The ending structure, obtained after two rounds of projection matching, incorporated 553
of the 578 class averages; and contained a 21.28% average error in the map, again using bi-
nary masked class averages, and comprising the vast majority of the dataset. And as before,
the averages showed outstanding, close-to-exact agreement with their corresponding repro-
jections (Fig. b, ii).

Once again, the structure revealed a prolonged shape (Fig. 19b, i), i.e. with two halves
connected by extended densities along the outer rims of the central cavity; albeit being
slightly longer by ~15 Å (i.e. 265 Å), noticeably wider by 30 Å (165 Å), and substantially
taller by 45 Å (again 165 Å) than BRCA2. Apparent in the side view, was the appearance of
extra density along the sides of the central cage, which contributed to the added width, and
the disappearance of a clear middle groove. Also evident was a steeply sloping underside,
which culminated in a pointed ‘tip’ at its base, i.e. as opposed to being flat as previously.
These gave the reconstruction a more inverted triangular appearance. And in the top view,
this now resembled a parallelogram. Furthermore, the central ‘channel’ could no longer be
found in the bottom view.

The binary masked map, rendered at a contour level that as before provided 1.2× the ex-
pected molecular mass (as also derived by STEM measurements), yielded a volume and
surface area of 1,560 nm³ and 1,800 nm² respectively. And likewise, whilst the former is
probably approximately accurate, the latter is most likely a considerable underestimate.
Due to the overall change in shape in comparing the two reconstructions, and thus an apparent tertiary reorganisation in going from BRCA2 protein to the BRCA2–RAD51 complex; it is not manifestly obvious which density in the latter corresponds to RAD51, as one could not then simply perform a difference map for this purpose. To assist with this, and the fitting of BRCA2 into the BRCA2–RAD51 structure, we employed our anti-BRC and anti-RAD51 labelling results (as detailed in §3.2.7 and §3.2.8).

It must also be stated that the BRCA2 and BRCA–RAD51 reconstructions were carried out entirely independently, without using projections of the former as references for either alignment or angular assignment of the latter’s particles, in order to prevent the well documented phenomenon of model bias\textsuperscript{118–120} (as mentioned in §2.5.4.1 and §2.5.4.6.2).
Figure 19. 3D reconstructions of BRCA2 (a) and the BRCA2–RAD51 complex (b). (a i) The BRCA2 structure shows a dimer with significant densities on the far ends of an elongated molecule. The region of less mass in the middle comprises two prolonged antiparallel densities along the sides of the molecule that connect the two presumably monomeric halves of the structure. This constitutes a central groove in the structure that gives the molecule its roughly characteristic kidney-bean shape, albeit with a flat base at the bottom comprising a cavity, formed by two distinct regions of density packed in a rectangular arrangement. The monomers themselves pack in a head-to-tail conformation to constitute the dimer. (ii) A representative selection of 20 of the
568 (binary masked) class averages (C) contributing to the reconstruction, displaying excellent agreement with their corresponding reprojections (R) from the (spherically masked) structure. This theme was maintained throughout the full set of averages and reprojections. The respective surface views (S) from the binary masked structure are also shown. (b, i) The BRCA2–RAD51 complex has a more distinctive overall triangular appearance, without a channel at the bottom or a clear V-shaped groove in the middle (c.f. BRCA2). Whilst the former is indicative of conformational change, the latter indicates RAD51 binding, as its place is now assumed by extra density along the side rims of the structure. (ii) 20 illustrative class averages out of the 553 constituting this reconstruction and their respective reprojections, once again displaying excellent correspondence – as throughout the entire set. As before, surfaces are also shown. In either case, the class averages (C) are hardly distinguishable from the reprojections (R).

3.2.5 Resolution

FSC calculations differ based on the kind of mask imposed on the half reconstructions, and the filling degree of the structure inside the volume. For instance, binary masked reconstructions shall produce the highest correlation values, whilst unmasked maps the least. Various programs will also differ in the way they take into account of the structure’s filling degree, while some might not do so at all. Furthermore, all else being the same, FSC values still differ – presumably owing to the slightly different correlation algorithms amongst the programs.

Values of resolution will then of course vary depending on the threshold criterion employed. For 3σ and ½-bit curves, their calculation in turn depends on the specified imposed symmetry of the reconstructions. Symmetry averaging creates correlations that these criteria compensate for, which fixed value criterions obviously do not.

Scaled FSC (i.e. FSC*, see §2.5.4.9) may also be used, which accounts for the higher SNR in the combined reconstructions. Its use is most often found in ‘gold-standard’ FSC calculations, whereby its rationale is very well established.

For the sake of consistency, in our case all Fourier shell correlations were performed using spherically soft masked half-reconstructions (of radius 0.7× the box width) in IMAGIC, intended as a compromise between the potentially optimistic and pessimistic FSC values provided by binary and unmasked models respectively. The molecules’ filling degree of 0.6, and C2 symmetry (in applicable stages) were also specified. For assessment of resolution, a blanket criterion of the fixed 0.5 threshold was applied (though other curves were retained for reference purposes). This was intended to yield conservative values, in light of the semi-
independent FSCs utilised, whereby the half-reconstructions are already biased toward the previous 3D model (having been obtained from class averages derived from particles aligned to a common template). These data were then rendered using an Excel macro created by Dr Timothy Grant§.

The resolution as determined by Fourier shell correlation using the 0.5 criterion, then gave rise to a value of 18.4 Å for BRCA2 and 19.5 Å for BRCA2–RAD51 (see Fig. 20). At ~19 Å, these are close to the physical limits of negative stain electron microscopy, whereby the size of uranyl acetate crystallites tends to impose a lower bound of ~18 Å on these samples115, and thus far less limited by either the size of the datasets or quality of the preparations.

![Figure 20](image)

**Figure 20.** FSC coefficient and resolution criteria plots for the finalised BRCA2 and BRCA2–RAD51 reconstructions. FSCs are calculated from spherically masked half reconstructions, albeit from binary masked class averages. The 0.5, ½ bit, 0.143 and 3σ threshold curves are overlaid. From their intersections with the FSC curves one obtains resolution values of 18.4, 15.2, 13.5 and 12 Å for BRCA2, and 19.5, 16.0, 14.2 and 13 Å for BRCA2–RAD51 respectively. Higher resolutions (lower values) toward the +ve x-axis.

---

§ Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA.
3.2.6 Structural validation

Due to the difficulties of tilt-pair analysis in negative stain electron microscopy, we adopted the modified ‘tilt-pair series’ approach as our primary means of validation of the reconstructions, as explained in §2.5.4.10.

Thereby, the observation that the measured tilts lay as close as an average ~2° of the expected tilts in the best cases for both, verified the overall correctness of the structures (Fig. 21a, i). The values were then (2.4 ± 1.4)° for BRCA2, and (1.6 ± 1.4)° for BRCA2–RAD51 – see Fig. 21a, ii.

Tilt-angle determinations were more accurate in the case of BRCA2–RAD51 complex in general, presumably facilitated by its larger size and more distinctive shape.

Three and two such series for BRCA2 and the BRCA2-RAD51 complex respectively were analysed.

We had also carried out conventional tilt-pair analysis on both samples (Figs. 21b & c). We selected particles from each tilt set (untilted and tilted) that were in agreement with projections from the corresponding model along the same angles as they (to check for correct angular assignment), and independently calculated the tilt angle for each of the pairs of projections using the formula derived in §2.5.4.10.

For 0°/+20° tilt pairs, in the case of BRCA2 (Fig. 21b, i); of the 45 pairs selected a cluster of 32 exhibited a mean tilt of (25.6 ± 19.8)°, from which a further subset of 26 then displayed a tilt of (18.6 ± 14.6)°. On the other hand, the full set of 45 pairs gave an average tilt of (40.9 ± 30)°, skewed greatly by a number of large outliers. For BRCA2–RAD51 (Fig. 21b, ii); of the 30 selected pairs, 26 produced a mean angular difference of (25.7 ± 16.7)°, with a smaller subgroup of 22 yielding (21.2 ± 13.9)°, and a yet smaller one of 20 giving (18.9 ± 12.4)°. The complete set of 30 provided an average tilt angle of (30.6 ± 20.0)°, again influenced highly by a few big outliers. A consistent theme of more accurate (but not more precise) measurements, from a larger proportion of the selection, was observed in the case of BRCA2–RAD51 complex compared with BRCA2 alone.

For −20°/+20° datasets, in the case of BRCA2 (Fig. c, i); 43 tilt pairs were selected, of which a 35 strong subset that excluded outliers produced a measurement of (34.5 ± 22.1)°. A still tighter cluster of 22 yielded (37.3 ± 15.3)°. The set of 43, i.e. incorporating anomalous tilt values, gave an overall value of (43.4 ± 31.6)°. In the case of BRCA2–RAD51 complex (Fig. c, ii); again 43 pairs were selected, which altogether produced a measurement of (32.7 ± 26.3)°, while a subset therein of 23 gave a value (36.9 ± 16.5)°. Improved determinations for BRCA2–RAD51 were thus again obvious hereby.
The outliers in each case consisted mostly of upper extreme outliers, i.e. tilt values >90°, but present also were lower far outliers, such as values of 0°.

\[ \delta = 18° \]

\[ \delta = 20° \]

\[ \delta = 24° \]
b, ii

\[ (119°, -10°) \quad (95°, 3°) \quad \delta = 27° \]

\[ (125°, 10°) \quad (143°, -13°) \quad \delta = 24° \]

\[ (95°, 21°) \quad (113°, 29°) \quad \delta = 24° \]

c, i

\[ (83°, 33°) \quad (47°, 36°) \quad \delta = 36° \]

\[ (131°, 11°) \quad (125°, 62°) \quad \delta = 40° \]

\[ (119°, -24°) \quad (149°, 3°) \quad \delta = 35° \]

c, ii

\[ (95°, -57°) \quad (83°, -88°) \quad \delta = 33° \]

\[ (95°, -39°) \quad (119°, -58°) \quad \delta = 30° \]

\[ (71°, 66°) \quad (65°, 29°) \quad \delta = 30° \]
**Figure 21.** Tilt series (a), and tilt pair (b, c) analysis. (a, i) Particles from 0–40° tilt series with 5° steps, of BRCA2 (top) and BRCA2–RAD51 (bottom), and matching reprojections with measured angles indicated. (a, ii) Table indicating the applied versus measured tilts for each of the tilted particles, of BRCA2 (left) and BRCA2–RAD51 (right). The difference for each pair is indicated, as well as their mean/standard deviation calculations. Measurements are stated to one decimal place. (b, i) Selected 0–20° tilt pairs of BRCA2, and matching reprojections with Euler angles shown. Calculated tilt angles are denoted. These examples have measured tilts close to 20°. As can be seen (c.f. matching reprojections), a 20° tilt of the BRCA2 structure can entail very little change in its projection (§2.5.4.10). (b, ii) Same for BRCA2–RAD51. These examples too have measured tilts of ~20°, and as seen here, a 20° tilt of the BRCA2–RAD51 structure tends to yield slightly more distinct projections. In either case references are rotated to match the particles. (c, i) Selection of 0–40° BRCA2 tilt-pairs, matching reprojections with angles, and calculated tilts. (c, ii) Same for BRCA2–RAD51. In these cases particles are rotated to match the references. Various filterings and maskings are depicted.

### 3.2.7 Structural domains

In order to identify various domains within the structures, we had utilised antibody labelling (as in §2.4). Specifically, we used an antibody against the C-terminal Flag tag of BRCA2, and another between BRC5 and BRC6 (Fig. 22a). Furthermore, we employed an antibody against RAD51 in the BRCA2–RAD51 complex. We did not find other good (monoclonal) antibody candidates against BRCA2 as tested through immunoprecipitation experiments.

Thereby, we then focused on particles with a single bound antibody, to reduce ambiguity in positional assignment of the antibody’s binding site in the particle’s projection. However, this was helpful as well for angular assignment of the particles themselves, as a crowd of antibodies would often render a particle’s boundary somewhat ambiguous and hence unfavourable to accurate masking. Nonetheless, as they were all too many, we did analyse several cases of particles bound to multiple antibodies for BRCA2–RAD51.

Euler angles were assigned to particles with attached antibodies, as per the respective 3D map’s projections. Using these angular designations and the locations of the antibodies, we then derived their (three-dimensional) positions on the respective structures via (a manner of) triangulations – as discussed in §2.5.4.11 (see Figs. 22b-d).

From these we saw that BRCA2’s C-terminus was situated on the top surface towards the far ends of the dimer (Fig. 22b). Similarly, we found the anti-BRC antibody was located on
the side rim of the structure, but toward the main monomer density (Fig. 22c), i.e. indicating that BRC repeats lay along these elongated densities. The latter would also suggest the extra density responsible for RAD51 was that observed on the outer rim of the BRCA2–RAD51 reconstruction (between the two globular monomeric halves). To test this, we labelled the BRCA2–RAD51 complex with an anti-RAD51 antibody.

We did indeed then find these to be positioned along the side rims of the BRCA2–RAD51 structure (Fig. 22d). However, they proved harder to localise to specific points, presumably due to the multiple antibody targets within this complex, which meant they could label the several cognate RAD51 monomers in any permutation. Regardless, the results indicated that RAD51 molecules were bound about the central cage region (on either side), toward the top of the structure.
Figure 22. Antibody complexes and identification of structural regions. (a) Field of antibodies (left), and selected antibody particles (right). The globular oval domains are most often times visible, but appear markedly flexible relative to each other. (b, i) A selection of three BRCA2-CFlag particles bound to anti-Flag antibody (left, circled blue), matching projections from the reconstruction (middle), and their corresponding surface views (right). Where the antibody meets the particle’s projection, a cylinder (also blue) is inserted at right angles into the surface. (b, ii) The result upon many repetitions creates an overlap of cylinders where the antibody is located (top), which is then highlighted (below). (c, i, ii) The similar process for BRCA2 labelled with anti-BRC5/6 antibody (magenta), and BRCA2–RAD51 bound to anti-RAD51 antibody (d, i, ii, gold). Coloured areas then indicate the regions as identified for BRCA2 and BRCA2–RAD51, i.e. the BRCA2 C-terminus (blue), BRC5/6 (magenta), and RAD51 (yellow/gold).
3.2.8  Fitting, conformational change in BRCA2 on RAD51 binding, and handedness

The localisation of both BRC repeats and RAD51 along the outer rims of the BRCA2 and BRCA2–RAD51 reconstructions respectively, provided inter-consistent evidence as to each other’s locations within the two structures (§3.2.7 and Figs. 22c & d).

Since RAD51 bound at the top of the BRCA2 dimer, the altered shape at the bottom of the BRCA2–RAD51 reconstruction is thus presumably owing to BRCA2, suggesting the occurrence of an induced reconfiguration of the BRCA2 molecule upon binding to RAD51.

The underside slope of the BRCA2–RAD51 complex, indicates there having been a tilt of the BRCA2 monomers toward the vertical axis (i.e. in relation to the dimer plane) – see Fig. 23a. In BRCA2, the bottom channel is constituted by two L-shaped densities, which we posit are each contributed by a BRCA2 monomer. The closure of this channel in the complex then seems to be brought about by a reorganisation of the L-densities, so as to become more closely associated in an intercalated arrangement (Fig. 23b).

We implemented the above changes in Chimera, via segmentation of the BRCA2 reconstruction into two presumably monomeric halves, and their independent fitting into the BRCA2–RAD51 density. This resulted in an increase in correlation coefficient between BRCA2–RAD51 and the BRCA2 models from 0.64 to 0.78. Although some uncertainties remained as to the exact boundary between the two BRCA2 monomers, it was largely apparent the conformational change had predominantly entailed a rotational and tilting motion on part of the monomers in order to accommodate the steep slope along the sides of the complex and its pointed tip.

Modelling this rearrangement spared the antiparallel side densities along the top of BRCA2–RAD51’s structure, i.e. those connecting its main polar densities, which had been identified previously as belonging to RAD51 via antibody labelling (cf. §3.2.7 and Fig. 22d). The two extra density regions could then each accommodate approximately four RAD51 monomers arranged in a filament-like fashion (Fig. 23c). For its potential significance and implications, see discussion in §4.2.

Hereby, we also adjudicated as to the likely handedness of the BRCA2 reconstruction, by splitting the two mirrors into their monomeric halves and gauging how well they could be modelled into the BRCA2–RAD51 structure, whose handedness we had in turn adjudged on the basis of fitting of the Rad51 filament crystal structure. We found the opposite BRCA2 mirror to the one hereby described failed to provide a sensible fit.

The above steps were performed by Prof Xiaodong Zhang, MSF Laboratory, Faculty of Medicine, Imperial College London, South Kensington, UK.
Figure 23. BRCA2’s binding to RAD51. (a) Interlay of BRCA2 dimer (cyan and yellow) and BRCA2–RAD51 (pink mesh) structures, highlights occurrence of conformational change. (b) Rearrangement of BRCA2 dimer inside the BRCA2–RAD51 complex involving a rotation and tilt of BRCA2 monomers. (c) This leaves extra density in the BRCA2–RAD51 structure that accommodates two oppositely directed sets of ~4–5 RAD51 monomers in a filament-like manner (orange ribbon). Handedness was ascribed to BRCA2 on the basis of its fitting into BRCA2–RAD51, whose determination was in turn based on the fit of RAD51 therein.

3.2.9 STEM mass measurements

To affirm BRCA2 remained a dimer upon RAD51 binding, we performed molecular mass measurements of the BRCA2–RAD51 complex by means of scanning transmission electron microscopy – as in §2.5.4.8.

A histogram of individual mass values (see Fig. 24), binned at 50 kDa intervals, and ranging from 0 to 1600 kDa; revealed two subpopulations at ~800 kDa and ~1200 kDa respectively, i.e. consistent with a BRCA2 dimer, as well as a dimeric complex. With regards to the former, we believe the process of repeated washings during STEM specimen preparation partly dissociated the complex.
By fitting two separate Gaussians functions to the histogram (i.e. one for each peak), as well as a double Gaussian distribution (the sum of the two); we derived molecular mass measurements of $760 \pm 220$ kDa for BRCA2, and $1240 \pm 110$ kDa for BRCA2–RAD51.

These data are therefore congruent with a BRCA2–RAD51 complex in which a set of $\sim 4$–5 RAD51 copies is bound per BRCA2 (§1.6), i.e. in a 2-fold arrangement as discussed above.

**Figure 24.** Histogram of mass determinations and statistical analysis. An independent Gaussian curve is fitted corresponding to either subpopulation, such that their cumulative provides the optimum fit for the entire bimodal distribution (viz. double Gaussian dissociation). The two peaks were measured to be centred at $760 \pm 220$ kDa (BRCA2) and $1240 \pm 110$ kDa (BRCA2–RAD51). This is commensurate with a BRCA2 dimer, and the binding of each monomer to $\sim 4$–5 RAD51s in the BRCA2–RAD51 complex. The number of particles able to be analysed was rather low, which perhaps contributed to the relatively large measurement errors (but that could also have been owing to heterogeneity). This implementation was performed in OriginPro 9.1 (OriginLab, Northampton, Massachusetts, USA).
3.2.10 BRCA2–ssDNA

Alongside recombinase binding, the most important property of BRCA2 as a recombination mediator, is binding to single-stranded DNA, i.e. as opposed to double-stranded DNA. This preference is critical, as RAD51 itself binds ss- and duplex DNA with similar affinities; and is therefore dependent on BRCA2 for loading onto its rightful substrate.66

To characterise BRCA2’s ssDNA binding, the minimum required length of single-stranded DNA for binding to BRCA2, was probed in the West laboratory. Using 32P-labelled oligonucleotides of defined lengths (i.e. 20–100 nt) and electromobility shift assays – see Fig. 25a, it was discovered that at least ~50 nucleotides were requisite for the formation of a stable BRCA2–DNA binding, although longer oligos of >66 nt were needed for maximal binding. Minimal to no binding was seen with ssDNAs shorter than 40 nucleotides.

In comparison, previous analysis of the BRCA2 DNA-binding domain alone68 (§1.5.6 and see later in §4.1); found it bound 12–20 nt of ssDNA weakly, with stronger binding occurring at 36 nucleotides. Our assays are therefore consistent with the notion whereby stable association of single-stranded DNA to BRCA2 necessitates the former’s synergistic binding to both monomers of the latter.

In our group, we visualised BRCA2–gDNA complexes under negative stain electron microscopy and analysed these through 2D image processing (Fig. 25b), in order to elucidate the course of single-stranded DNA bound across the BRCA2 dimer, i.e. as detailed in §2.2 and §2.5.4.11.

We discover the relative positions of flanking dsDNA arms about BRCA2 molecules bound to the intervening single-stranded DNA, were consistent with a scenario whereby ssDNA bound along the long-axis of the dimer at its top surface; which in turn is compatible with the positioning of the C-terminus on BRCA2 as determined by antibody labelling (§3.2.7), i.e. since it is close to the DBD (being <7% of the sequence length away).
Figure 25. ssDNA binding of BRCA2. (a) BRCA2’s binding to 5'-32P-labelled ssDNAs of lengths 20–100 nt as assessed by EMSA. BRCA2 exhibits nominal binding to ≤ 40 nt, stable binding with 50 and 66 nt, but strongest binding only at ≥ 74 nt, suggesting their simultaneous binding to either BRCA2 monomer. (b) A selection of five BRCA2 particles bound to gapped DNA substrates (left column), matching projections from the structure (middle), and corresponding surface views (right). dsDNA arms are highlighted in the particle images (orange dots), and in accordance with matching projections, traced onto the surface views (orange lines) for interpretation. These then demonstrate that ssDNA binds along the long axis of the BRCA2 dimer at its top surface.
3.2.11 BRCA2–RAD51–ssDNA dynamics (West laboratory**)

3.2.11.1 Filament frequency and length

We sought to visualise the precise effect of BRCA2 on RAD51–ssDNA filaments via negative stain electron microscopy (Fig. 26a, b).

On investigating initial nucleation events, there were found to be ~4× nucleoprotein filaments in the presence of BRCA2 than without, having quantified 149 and 204 randomly collected frames of the former and latter respectively (giving \( P < 0.0001 \) by a two-tailed \( t \)-test) – see Fig. 26d.

However, relatively short filaments (40–100 nm) were observed either with or in the absence of BRCA2, and measurement of 332 of the former and 314 of the latter, revealed no appreciable difference in length between the two, i.e. either as an average, or using 20 nm bins from 0–120 nm followed by a singular > 120 nm bin (Fig. 26c).

![Figure 26](image)

**Figure 26.** Nucleation of RAD51–ssDNA filaments by BRCA2. Difference in the number of nucleation events in absence (a) versus the presence of BRCA2 (b) as seen by negative stain electron microscopy. Right hand columns show enlargements of nucleoprotein filaments. Quantifying lengths, and difference in the number of such fil-

**Except §3.2.11.2.1.**
aments per frame, revealed no appreciable change in the former (c), but a 4-fold significant increase in the latter (d). Blue and red bars denote RAD51–ssDNA filaments with and without BRCA2 respectively. Scale bar (black) indicates 100 nm.

3.2.11.2 Filamentisation polarity

As discussed earlier (§1.6 and §3.1), the BRCA2–RAD51–ssDNA ‘supercomplex’ is transient, and does not occur stably.

Nonetheless, remarkable findings were made in visualising BRCA2–RAD51–ssDNA samples under negative stain EM, i.e. the observance of RAD51–ssDNA filaments with BRCA2 clearly situated at one end (Fig. 27c). The latter were confirmed to indeed be BRCA2 particles using a monoclonal anti-Flag antibody coupled to 20 nm gold particles (Fig. 27d). However, at longer reaction time points such terminal BRCA2 particles were detected only infrequently, indicating its dissociation once nucleoprotein filament formation had been established, as per the theorised model. This was a striking demonstration of a hitherto unknown mono-directional growth of these filaments from BRCA2. To determine the polarity of filament growth itself, ssDNA molecules were gold-labelled at either their 5’ or 3’ termini. Electron microscopy analysis revealed that RAD51 filaments invariably extended toward the 5’ gold-labelled end – see Fig. 27e. By contrast, using ssDNA with 3’ gold label, similar complexes oriented in the other direction (i.e. 5’–3’), failed to be observed.

Our obvious interpretation of these results was hence that BRCA2 promoted unipolar RAD51–ssDNA filament growth out of nucleation foci in a 3’–5’ direction, consistently with the likewise polarity of DNA strand exchange (relative to the bound ssDNA) that takes place later.

For a description of how the above is in agreement with our structural data, especially that concerning BRCA2 and BRCA2–RAD51’s seemingly dimeric nature, see discussion in §4.2 and §4.3.
Figure 27. Polarity of RAD51–ssDNA filament growth. EM visualisation of RAD51–ssDNA filaments (a) and BRCA2–ssDNA complexes (b), followed by a BRCA2–RAD51 complex + ssDNA mixture (c) – all after short incubation times. In the latter, one clearly sees RAD51 filament emanating from easily distinguishable BRCA2 particles. To establish the reality of this situation, BRCA2 (d) and the 5’ ssDNA end (e) were in turn immunogold labelled, to enable their localisation within these structures. In the former case, gold particles (dark spots) are found on the globular entities capping the end of the filaments, whilst in the latter case they are situated at the far end of the filament from these bodies. This indicates the latter to be BRCA2, and that RAD51 filaments issue therefrom along ssDNA toward its 5’ terminus, while BRCA2 caps the 3’ end.
3.2.11.2.1 Image analysis of nucleation events (Zhang group)

By finding matching model projections and corresponding surface views for the BRCA2 particles hereby observed capping recently nucleated RAD51–ssDNA filaments, we had sought to visualise the points of exit and outline paths of such filaments out of BRCA2 (§2.5.4.13).

Thereby, we invariably found these filaments to emerge from one side, and one end of BRCA2 dimers, suggesting that productive nucleation ensued from one (and only one) RAD51-laden monomer, i.e. that in the correct polarity for RAD51 filament growth on bound ssDNA – see Fig. 28.

The sampled particles were too few for analysis to be rigorous (§2.5.4.13), but it nonetheless provided an interesting qualitative assessment. One should note that neither the BRCA2 nor the BRCA2–RAD51 structures are necessarily accurate representations of the capping BRCA2 particles. Presumably, the BRCA2–RAD51 dimer would have offloaded half its cognate copies of RAD51, with the concerned BRCA2 monomer then also dissociating from the ssDNA – but crucially not the other monomer, which also remains bound to ssDNA.

Therefore, these might adopt a conformation rather different to that seen in either structure. Naturally, this makes their analysis difficult, until such a time that BRCA2 can be purified at the gross quantities necessary to permit analysis of its and its complexes’ various states.
3.2.11.3 Multinucleation

Intriguingly, multiple observations of linearly aligned BRCA2 particles and short RAD51 filaments were made (see Fig. 29), presumably along the same ssDNA molecule. These suggested that BRCA2 molecules, and by inference BRCA2–RAD51 complexes, bind along multiple sites on single-stranded DNA and seed multiple nucleation events.
Altogether, the data indicate that BRCA2 facilitates the interaction of RAD51 with ssDNA, by creating their critical mass thereon at repair foci, from which filaments then self-polymerise in a 3’–5’ direction. In accordance with earlier work, nucleoprotein filament assembly itself seems to entail a formation of multiple nucleation clusters followed by subsequent ‘gap-filling, as opposed to continuing elongation from a single or small number of such sites. The growth of these short nascent filaments into one extended filament capable of efficient homologous pairing and strand exchange, may require the action of other recombination factors, perhaps in particular RAD52 but also RAD51 paralogues, RAD54, and PALB2.

![Image of multiple RAD51 filament nucleations along ssDNA.](image)

**Figure 29.** Observation of multiple RAD51 filament nucleations along ssDNA. Micrograph shows various BRCA2–RAD51–ssDNA nuclei in a linear array, presumed to be along a single ssDNA molecule. Though this cannot be said definitively, i.e. as the ssDNA itself does not get visualised, it is nonetheless likely as DNA was at an extremely low concentration, and such images were observed throughout the sample grid.
4 Discussion

Our research offers new insights as to the role of BRCA2 in driving RAD51 nucleoprotein filament assembly on ssDNA during homology-directed repair, which constitutes the core enzymatic function of this protein\textsuperscript{123}. The dysfunction of this very process in individuals with BRCA2 mutations, leads it to be one of a small group of the most notorious cancer-related genes\textsuperscript{3}.

Hereby, we put forward the first structure of full-length BRCA2 protein, on its own and bound to its cognate recombinase RAD51. Crucially, we find that BRCA2 occurs, and is active as a dimer. Amongst others (see §1.6), we show it binds to two sets of ~4–5 RAD51 molecules in opposing polarities; that single-stranded DNA binds at the top surface of the dimer along its long axis; BRCA2 increases the frequency (4-fold), but not the length, of nascent RAD51 nucleoprotein filaments; and that it instigates monodirectional 3′–5′ filament growth from each nucleation site.

To encapsulate our results in prior context, detailed below are the salient aspects of our proposed model of the mechanism of action of the BRCA2 breast cancer tumour suppressor, in view of the structural information.

4.1 BRCA2 DNA-binding domain and RPA

Presynaptic filament formation, strand exchange and pairing are all highly dependent upon RPA. It is believed this is because of RPA minimising secondary structure in single-stranded DNA. This would facilitate recruitment/binding of recombination mediators and auxiliary factors, loading of RAD51 onto single-stranded DNA, and assembly of more continuous RAD51-ssDNA nucleoprotein filaments. Though RPA is needed for RAD51-affected recombination \textit{in vitro}, its stimulatory effect is minimal if granted access to single-stranded DNA prior to RAD51. Presumably, this is owing to RPA competing with RAD51 for the ssDNA substrate, given the former has \textasciitilde100 times the affinity for single-stranded DNA than the latter. It is hence thought accessory factors including recombination mediators like BRCA2 help remove RPA from ssDNA.\textsuperscript{68}

In the study by Pavletich and colleagues of the structure of BRCA2DBD-DSS1, its ability to facilitate RPA-dependent recombination activity of RAD51 was too tested.\textsuperscript{68} As might be expected, reaction efficiency was lower when RAD51 and RPA were added simultaneously to ssDNA, than if RPA was added after RAD51. However, when BRCA2-DSS1 was added concurrently with RAD51 and RPA, catalysis of the reaction was observed as a function of the former’s amount. A sub-stoichiometric concentration 1/15\textsuperscript{th} that of RAD51’s sufficed to
stimulate the reaction 2.4-fold. The stimulation plateaued at 4.5-fold upon reaching 16/15th the RAD51 concentration. BRCA2DBD-DSS1 also facilitated the reaction when RPA was added after RAD51/ssDNA, though the amount of stimulation achieved was less by 1.9-fold.

There are manifold parallels between the BRCA2 DNA-binding domain and RPA (see Fig. 30). Amongst all proteins carrying OB folds, BRCA2’s are most similar to those in RPA, as measured by 3D homology search and visual inspection of several such proteins. RPA, a heterotrimeric protein, possesses two OB folds with high affinity for single-stranded DNA (DBD-A and -B), and two others with weaker affinity ~1/100th that of the former (DBD-C and -D), but which contribute to its overall binding to ssDNA. This is highly reminiscent of BRCA2, whereby an isolated αD-OB1 domain also has circa 100 times less avidity for single-stranded relative to the full DBD.68

Structural homology is closest between the BRCA2 OB2 and RPA DBD-A domains, the two being superimposable within 1.4 Å r.m.s.d. for 89 out of the 109 alpha-carbons. The second closest homology is that between BRCA2’s OB3 and RPA’s DBD-B domains, with an r.m.s.d. of 1.5 Å for 84 Cα atoms. These deviations are actually lower than even those amongst the RPA OB folds. Although OB1 of BRCA2 is more structurally divergent, its closest homologue is also an OB fold of RPA’s, i.e. DBD-D, of 1.9 Å r.m.s.d. for 65 Cα’s.

The arrangement of OB folds is also similar to those seen in the RPA-ssDNA crystal structure, which shows the DBD-A and -B domains bound to oligo(dC)₆. And as with BRCA2, RPA displays minimal affinity for oligo(dA) and oligo(dG), supposedly due to their high propensity for forming secondary structures. Contacts between OB folds and single-stranded DNA also follow the same paradigm in RPA and BRCA2 DBD.

Furthermore, Pavletich and colleagues’ data showed that the BRCA2DBD-DSS1 complex can produce two different complexes with single-stranded DNA, i.e. a strongly avid complex requiring the tower and involving 32- to 36-nucleotide oligo(dT), and a weaker complex only formed at higher concentrations entailing as few as 8 to 12 nt, and apparently unaffected by deletion of the tower. Remarkably, such two binding modes have also been demonstrated for RPA. One employs only the DBD-A and -B OB folds (as seen in the crystal structure), has low affinity but high cooperativity, and has a DNA-binding footprint of 8 nucleotides, i.e. 8-nt mode. The other mode engages all DBD domains (A/B/C/D), has high affinity but low cooperativity, and involves a binding site of ~30 nucleotides, i.e. 30-nt mode (as shown by deletion, mutagenesis, and crosslinking experiments). By analogy, it thus appears that the binding of OB2-OB3 domain to single-stranded DNA observed in the crystal structure of BRCA2DBD₄Tower-DSS1-oligo(dT)₉, represents the weak complex and mimics the 8-nucleotide binding mode of RPA. The strong complex could therefore form by utilising
the remaining DNA-binding domains of BRCA2, especially since the αD-OB1 region does possess some additional DNA-binding activity.68

Provided the linker between αD-OB1 and OB2-OB3 folds, and in light of our reconstructions and the above explained identicalities between BRCA2 DBD and RPA, we believe the former could adopt a more compact, ‘horseshoe’-shaped overall quaternary structure like the latter in the context of full-length BRCA2 (Fig. 31).

In support of the above hypothesis, heterologous fusion polypeptides comprising a single (BRC3 or BRC4), or multiple BRC repeats of human BRCA2, fused to the large hRPA subunit, which harbours the DBD-A and -B domains – were found to significantly complement Brca2-deficient mammalian (hamster) cells. They suppressed spontaneous chromosomal aberrations, and promoted Rad51 focus formation and cell survival upon exposure to DNA damaging agents. It remains remarkable that as little as 2% of BRCA2 by sequence, combined with RPA70 could at least partly recapitulate its function. But perhaps not surprisingly, some of these constructs could be considered analogues of BRCA2 in certain organisms. For instance, the BRC3-RPA peptide (the smallest produced) is equivalent to the single BRC repeat orthologues found in U. maydis (Brh2) and C. elegans (BRC-2), or those containing truncated DNA-binding domains (such as the latter, which lacks the tower).125

In view of the complexity of the human genome, the RAD51 recombinase may have evolved to rely on BRCA2 for its localisation to sites of processed DNA double-strand breaks, or specific structures therein. Its lack of preference for single- over double-stranded DNA might hence reflect this fact.68

Figure 30. Comparison of BRCA2DBDΔTower-DSS1-ssDNA68 (a) and RPA-ssDNA124 (b) crystal structures. The two are highly similar with regards to both structure and function, with differences only in gross quaternary organisation. Given this, and the
flexible linker between the αD-OB1 and OB2-OB3 domains of BRCA2DBD, leads us to propose that it might adopt the compact fold of RPA within the 3,418 amino-acid full-length BRCA2. Homologous domains are like-coloured, i.e. OB1 and DBD-D (orange), OB2 and DBD-A (blue), and OB3 and DBD-B (teal). DSSI (lime green) and ssDNA in stick-mode (red) are depicted.

4.2 Dimerisation and its significance

Our data show that BRCA2 exists principally in dimeric form, and these do not dissociate on forming complexes with either RAD51 or ssDNA. BRCA2 then appears to bind four to five RAD51 monomers per each subunit, which matches or exceeds the minimal nucleation length needed to set off filament extension via self-polymerisation (§1.6).

Each BRC repeat can mimic, with varying efficiencies, RAD51 interprotomer contacts requisite for filament and ring assemblies (§1.5.4); which suggests the notion that recombinase molecules are aligned by BRCA2 (§1.5.5), in a filamentous or pseudo-filamentous manner, in preparation for loading unto ssDNA as such.

Due to the intrinsic two-fold symmetry imposed by the dimer, such that the monomers are organised in a head-to-tail conformation, the two sets of RAD51 molecules bound to the BRCA2 dimer would thus be oriented in opposite directions. Upon imposition of a defined polarity through ssDNA-binding, one set of these molecules will then be poised correctly for interaction with DNA (Fig. 31ii) – as RAD51 filaments grow preferentially in the 3’–5’ direction (see §4.3). Despite invoking only half the bound substrate, this arrangement offers a mechanistic advantage, since regardless of the orientation of ssDNA binding, a dimer shall always present one set of recombinase molecules in the right direction for productive nascent filament formation.

The potentially decreased diffusion rate of a dimeric versus monomeric species in the dense nuclear interior should also be of little consequence, as in the light of even a purely passive scenario, an added two times bulk would only slow the diffusivity by a factor of \( \sqrt[3]{2} \approx 1.26 \). And given that nucleoplasmic BRCA2’s diffusion rate was demonstrated to be in the millisecond timeframe\(^{26}\) whilst DNA damage signalling and repair processes take place over minutes or even hours, a doubled molecular weight is unlikely to hamper the repair’s efficiency – especially provided the substantial reaction benefits afforded by a dimer.

Even though our data clearly show that human BRCA2 forms dimers in vitro (alongside reported evidence for dimerisation of \( U. \) \textit{Maydis} orthologue Brh2\(^ {88}\)), one cannot exclude the possibility that it might also assume a different form, e.g. \textit{in vivo}, upon interaction with partners such as PALB2 and/or (indirectly) BRCA1, or DNA damage or cell cycle mediated
post-translational modifications that may trigger a dimer to monomer transition. A similar mechanism has been proposed for Brh2, whereby a switch to its monomeric state occurs upon DSS1 binding. In avian DT40 cells, a transition from slower-diffusing species of BRCA2 (i.e. large macromolecular assemblies), to ones capable of faster diffusion (smaller complexes), has been observed upon induction of DNA breakage. Such transitions have also been proposed as regulatory mechanisms for a couple of DNA damage signallers, e.g. ATM and ATR (ataxia telangiectasia and Rad3-related) proteins.

Nonetheless, in contrast to ATM or ATR, BRCA2 dimers are active, i.e. in their case at RAD51 and ssDNA binding and seeding presynaptic filament formation. Whether BRCA2 monomers too are similarly or at all active as such, remains to be seen and cannot unambiguously be prognosticated toward (though monomeric BRCA2 could conceivably always bind ssDNA in the proper orientation that allows RAD51 loading, and thus not be futile during half the instances – a problem that is automatically solved by the dimer).

---

**Figure 31.** Mechanism of action of BRCA2. (i) Crystal structure of RPA in its 30 nt ssDNA binding mode (as in Fig. 21), showing a compact U-shaped quaternary arrangement of its DNA binding domains/OB folds. As discussed in §4.1, the 704 a.a. BRCA2 DBD could adopt a similar configuration inside the 3,418 a.a. full-length BRCA2. Two symmetry related RPA molecules with separately coloured domains (1, 2, 3, 4), are then colour-coded onto the BRCA2–RAD51 structure in place of the BRCA2 DBDs (ii). When binding to ssDNA (depicted above in a left to right 5'–3' po-
larity), it will be able to do so through two of the DNA binding domains on either monomer, which will be oriented correctly for ssDNA binding notwithstanding its relative polarity (i.e. folds 3 and 4 / OB2 and OB3 on the left monomer, and folds 1 and 2 / αD and OB1 on the right monomer in the above example). Since RAD51 filaments grow with a 3′–5′ bias, one of the two sets of RAD51 molecules bound by the BRCA2 dimer in opposing directions (orange with yellow outline), will then always be able to productively engage the ssDNA. This benefit would not be afforded to monomeric BRCA2, which might not be effective in 50% of its DNA binding occasions. The dimer also seems to help by increasing the interaction surface with ssDNA, and promoting maximally stable complexes therewith (see §3.2.10).

4.3 Polarity of filament growth and strand transfer

RPA, and presumably also the BRCA2 DBD, preferentially bind DNA with a defined polarity. How might then an ssDNA strand, with its defined polarity, bind across a BRCA2 dimer (i.e. if the DBD atop one monomer will be oriented incorrectly)?

As reasoned in §4.1, the four DNA binding folds in full-length BRCA2, might instead be arranged in a similar compact U-shaped configuration as those of RPA, such that two of the four are then facing the ‘right’ direction.

Therefore, notwithstanding the relative polarity of its binding, the ssDNA would be able to engage these subdomains on one monomer (e.g. OB2 and OB3), and the other two in the second monomer (αD and OB1) in the same polarity – thus enabling the correct polarity for ssDNA binding on either monomer.

Our determination that BRCA2 exhibits weak binding toward ssDNAs of up to 40 oligonucleotides (§3.2.10), and necessitates >70 nt for strong binding, supports the concept that ssDNA contacts the DBD on both monomers and the two act in concert.

Recombination is initiated at extensive 3′ ssDNA tails created by exonuclease action at DNA double-stranded breaks. They serve as a substrate for multiple BRCA2 effectuated RAD51 nucleation and initial filamentisation events. These ‘buds’ then extend and merge to form a single nucleoprotein filament that shall instigate homologous pairing and exchange.

Our analyses of the BRCA2–RAD51–ssDNA quasi-complex, reveal BRCA2 to be situated at the 3′ end of each emergent RAD51–ssDNA filament, thereby corroborating the model in which such filaments grow in the 3′–5′ direction – congruous with the polarity of RAD51-mediated strand exchange\textsuperscript{128,129}.

One prospect is that BRCA2 ‘caps’ one terminus of the nucleoprotein filament (i.e. the 3′ end), and blocks its dissociation, whilst allowing it to self-polymerise and extend at the other
end. Since neither BRCA2 nor RAD51 display any preference for binding at either the ssDNA/dsDNA junction or the 3’ terminus of the resected ssDNA tail\textsuperscript{86,87} and would thus likely bind internally, RAD51 polymerisation in the 3’–5’ direction might beg a situation whereby several nucleotides at the far 3’ ssDNA end itself fail to be covered by RAD51 (see Fig. 2 for reference). However, this is not a problem, as a wholly coated ssDNA tail without a 3’ protruding end, is not requisite for strand pairing\textsuperscript{130,131}. In fact, it has been demonstrated that RAD51 can promote incipient ssDNA-dsDNA pairing interactions at either 3’ or 5’ extended ssDNA tails\textsuperscript{132,133}. Though once achieved, only the invading 3’ end can serve as a primer for DNA repair synthesis.

Incidentally, the \textit{U. Maydis} Brh2 protein exhibits a distinct bias toward binding ss-/ds-DNA junctions, and promotes RAD51 filamentisation at the 3’ ssDNA overhang, i.e. whilst BRCA2 evidently lacks such a preference (§1.5)\textsuperscript{85–87}. Given that BRCA2 appears to cap the 3’ terminus of the RAD51–ssDNA filament, which then extends in a 3’–5’ direction relative to the bound DNA – implies that some differences exist between these orthologues. Though whether they reflect genuine mechanistic diversity, or are a product of different experimental conditions is not entirely clear.

Whilst there might not necessarily be a direct relationship between the polarities of filament growth and strand transfer (that occurs a while later in the process), it is worth noting that in RecA filaments, which extend in the 5’–3’ direction, the polarity of strand transfer is too 5’–3’\textsuperscript{134–136}. Therefore, not least there is a certain consistency amongst the two.

We reason that the differential directions of filament growth and strand transfer for RecA and Rad51, relate to the positioning of an additional domain on opposite sides of the central core in each case (§1.4.2), which gets stabilised only upon filament formation.

Explicitly, RAD51 harbours a supplemental N-terminal domain, whilst RecA contains a C-terminal extension. These are situated on opposite sides of the ATPase core relative to the single-stranded DNA. With RAD51, filamentisation entails a β-strand connecting the ATPase domain and the N-terminus that is ordinarily disordered, but is stabilised upon packing against the core domain of an adjacent monomer (§1.4.3). This interaction is thereby enhanced toward the growing N-terminal end of the filament, i.e. the 5’ direction.

By comparison, intersubunit interactions in RecA filaments further involve a loop, i.e. one that is also only ordered in the filament. This is located to the C-terminal flank of the central domain, such that subunit addition in the 3’ direction would stabilise the loop, and subsequent interactions.

Hence, owing to disparate domain configurations of RecA and Rad51, these filaments extend with opposite polarities.
All in all, our results imply a mechanism for BRCA2 tumour suppressor mediated repair of DNA double-stranded breaks by homologous recombination (see Fig. 32). A dimer thereof binds two sets of RAD51 molecules, each arranged optimally for its loading onto ssDNA (derived via end-resection) and initial filament formation – given the right polarity of bound ssDNA, which it binds across its top surface. Upon ssDNA binding, one set always binds thereon, because of the directionality imposed thereby (an advantage only available to dimers). A RAD51 nucleoprotein filament then extends in the 3′–5′ direction, dictated by its molecular architecture, leaving BRCA2 to cap its 3′ terminus. The highly substoichiometric ratios of BRCA2 necessitated for RAD51 filament nucleation, suggests that BRCA2 dissociates upon the nascent nucleoprotein filament undergoing extension, and is subsequently recycled in order to initiate the formation of further such filaments – thereby serving as a molecular chaperone for RAD51 filament assembly.

**Figure 32.** Further mechanism. Structure data, observations of multiple nucleation sites along single tracts of DNA, polarity of nucleoprotein filament growth out of BRCA2, and the extremely low amounts of BRCA2 relative to RAD51 required to promote these, among others indicate a mechanism whereby the BRCA2–RAD51 complex binds at several points along end-resected 3′ ssDNA tails, and stimulates RAD15 filament formation in a 3′–5′ direction therefrom. These emanate from one monomer and cause the BRCA2 dimer to partially dissociate from the RAD51–ssDNA filament, and then self-extend to link up with similarly nucleated filaments downstream in the 5′ direction, displacing BRCA2 altogether and creating a single long nucleoprotein filament.
The dissociated BRCA2 can be utilised to further catalyse the reaction, contributing to its efficiency. This scheme presumably facilitates the formation of longer RAD51–ssDNA filaments that are desirable for productive strand pairing, homology search, and D-loop formation.
References


100. VAN HEEL, M. Classification of very large electron microscopical image data sets. *Optik (Stuttg).* **82**, 114–126


