Improved RAD51 binders through motif shuffling based on the modularity of BRC repeats

Laurens H. Lindenburg[1], Teodors Pantelejevs[1], Fabrice Gielen[3], Pedro Zuazua-Villar[2], Maren Butz[1], Eric Rees[4], Clemens F. Kaminski[4], Jessica A. Downs[2], Marko Hyvönen[1] & Florian Hollfelder[1*]

[1] Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, UK

[2] The Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK

[3] Living Systems Institute, University of Exeter, Exeter, EX4 4QD, UK

[4] Department of Chemical Engineering and Biotechnology, New Museums Site, Pembroke Street, Cambridge, CB2 3RA, UK

ABSTRACT

Exchanges of protein sequence blocks support leaps in function unavailable through point mutations during evolution. Here we study the shuffling of short modules within peptide repeats, focusing on the binding between the eight BRC repeats of BRCA2 and RAD51. The RAD51:BRCA2 interaction is a key factor in double-strand break repair by homologous recombination. Each BRC repeat consists of two modules interacting with two binding sites on RAD51. We shuffled the two modules from all eight repeats, using a microfluidic-based system to rapidly screen the resulting 64 chimeric variants. Seven BRC repeat peptide chimeras were found to display higher affinity than any of the natural repeats. Surprisingly, we found that certain modules, that were weak binders as part of their parental combinations, turned out to be much stronger once shuffled: for example, the chimera BRC8-2 exhibited binding improved by -2.44 kCal/mol. A crystal structure of the high-affinity complex of monomeric RAD51 and the BRC8-2 chimeric repeat revealed the basis for rationalizing this binding enhancement: shuffling enabled an improved interface fit and provided an extended beta-hairpin. The chimeric high affinity BRC repeat was shown to inhibit the function of RAD51 in human cells and prevent the formation of nuclear foci after ionizing radiation.

Keywords: BRCA2 / RAD51 / peptide repeat modularity / protein shuffling

- Investigating modularity within peptide repeats by shuffling
- Insight into BRC repeat peptides binding RAD51
- Shuffled repeat with improved RAD51 binding, inhibits RAD51 function in human cells

INTRODUCTION

The daunting combinatorial diversity arising from simultaneous mutation of all amino acid positions even in small proteins (leading e.g. to 10^{39} variants of a 30-mer), renders exploration of such sequence space

futile. It is therefore attractive to view proteins not only as combinations of variant amino acids, but as combinations of units of stretches of amino acids, because the combinatorial diversity is dramatically reduced if such 'modules' are instead recombined (1–6). Understanding protein modularity at different length scales is key to elucidating natural protein evolution and to achieving full control in protein design and engineering. Shuffling is an empirically proven approach used in directed protein evolution (7–9) and has been made more efficient by the automated determination of minimally folded domains for shuffling, for example using the SCHEMA algorithm (10, 11). The exon-intron architecture of eukaryotic genes may in fact serve the modular evolution of proteins through *exon shuffling* – homologous recombination at introns to bring exons into novel combinations – and help bring about new protein functions (12–15).

Here we address the relationship of modularity and function in a protein-protein interaction pair, RAD51:BRCA2, involved in in DNA double-strand break repair. BRCA2 exerts a multitude of functions on RAD51 in the cell, such as localization, nucleofilament assembly and its depolymerization and has been aptly termed the 'custodian' of chromosomal numerical and structural integrity (16). BRCA2 is a 3418 amino acid protein whose central part consists of eight conserved repeats (referred to as 'BRC repeats' followed by the number 1 to 8, each consisting of around 35 residues (17); Figure 1A). Recognizing this modularity, we probe the idea that functional sophistication is brought about by combination of these relatively simple peptide building blocks by testing whether rearranged BRC repeats can lead to functional chimerae that interact with RAD51.

The crystal structure of BRC4 in complex with RAD51 (18), together with structural modeling and biochemical experiments, revealed the existence of two distinct parts in the BRC repeats that interact with RAD51 (19). The first of these, known as the 'FxxA module', forms a β-hairpin structure and binds RAD51 with a Phe and Ala in two small binding pockets of the ATPase domain. The C-terminal part of the BRC repeat, with a conserved LFDE motif, interacts with the distal part of the ATPase domain in an α-helical conformation. In doing so, the BRC repeats directly compete with another FxxA module located in RAD51 itself (with the sequence FTTA), on an oligomerization epitope between RAD51's C-terminal ATPase and N-terminal DNA-binding domains. The BRC4 repeat peptide has been shown to cause dissociation of RAD51 oligomers and conditional expression of the repeat in breast cancer cells disrupts the RAD51:BRCA2 interaction and sensitizes them to radiation treatment (20). In isolation, the FxxA module makes a relatively weak contribution to the binding: a 4-residue FHTA peptide, representing the FxxA hotspot from the corresponding module in BRC4, bound RAD51-surrogate HumRadA2 with a K_d of 290 μM (21). Even the entire FxxA module – that is the FxxA hotspot and the surrounding residues – is not a strong binder of RAD51; about 500 μM of a 17-residue FHTA-containing peptide (the N-terminal half of the BRC4 repeat) was required to effect full disruption of the RAD51:BRC4 interaction in a competitive ELISA assay (19). It is the C-terminal LFDE module that ensures significantly enhanced affinities are achieved. This second module binds to a groove on another surface of the RAD51 ATPase domain (18). Although the phylogeny of the BRC repeats remains to be fully elucidated, it is thought that the emergence of the BRCA2 repeats predates the radiation of the mammalian class (22) and perhaps even the divergence of birds and mammals 230-300 million years ago (23). As the eight

2

repeats found in the BRCA2 protein all occur on the same exon (22), throughout their evolutionary history, these repeats would never have been subject to natural exon shuffling.

This led us to pose the questions: i) would artificial shuffling of repeat modules engender affinity maturation by bringing together the most binding-proficient FxxA and LFDE modules; and ii) would these chimeric peptides give us improved insight into module-specific contributions to RAD51 binding? Our objective was to explicitly explore the recombination of entire natural repeats in the creation of new functional proteins, in order to demonstrate the role of modularity in functional adaptation. We discovered that the natural, 'parental', combination of modules often turned out to be suboptimal for RAD51 binding, and that upon decoupling of natural BRC module combinations, more potent RAD51 binders could be obtained.

MATERIALS AND METHODS

Reagents

Preparation of BRC4 repeat peptide, N-terminally labeled with fluorescein (BRC4[fl], sequence CKEPTLLGFHTASGKKVKIAKESLDKVKNLFDEKEQ) was described previously (24). CHES was from Sigma, Pico-Surf 1 was from Dolomite, HFE-7500 was from 3M.

Plasmid constructs & cloning

For the construction of 8 parental and 56 shuffled BRC peptides, as well as several more mutant peptides, please see SI (Supplementary Figure S1 & S2 & Table S1). The *E. coli* expression construct for monomeric RAD51 (pBAT4-HumRadA22), has been described previously (25). Cloning of plasmids GFP and GFP-BRC8-2 for mammalian cell transfection is described in SI (Supplementary Figure S3 & Table S2).

Protein expression & purification

The 64 different GB1-BRC peptide fusion constructs were separately transformed to chemically competent *E. coli* BL21(DE3). Overnight LB pre-cultures were used to inoculate 20 mL LB, which were grown up to mid-log phase (OD_{600} of 0.5). Expression was induced using 1 mM IPTG and cultures were incubated for a further 3 hours at 37 °C. Cells were then harvested through centrifugation and lysed by the addition of BugBuster/Benzonase lysis reagent (Novagen, with 5 mM imidazole, 20 mM Tris-HCl, 100 mM NaCl, pH 8). The resuspended mixture was incubated for 20 minutes at room temperature, then loaded directly onto a Ni-NTA protein miniprep column (His Spin Protein Miniprep, Zymo Research). Protein was washed following the manufacturer's instructions. Proteins were eluted in 500 mM imidazole, 20 mM Tris-HCl, 100 mM NaCl, pH 8, 150 µL. Protein concentrations were quantified by UV absorption at 280 nm (using a Nanodrop spectrophotometer) and corrected for the presence of truncated side-products by SDS PAGE (Supplementary Figure S4). Monomeric RAD51 expression and purification for fluorescence anisotropy measurements was carried out as described previously (25), where monomeric RAD51 was called 'HumRadA22'.

3

Monomeric RAD51:BRC8-2 complex purification for crystallography

E. coli BL21(DE3) cells carrying pUBS520 plasmid for rare AGA/AGG encoding tRNA were transformed with GB1-BRC8-2 or monomeric RAD51 constructs and grown at 37 °C in 1 L of 2x YT medium in shaker flasks in the presence of 100 µg/mL of ampicillin and 25 µg/mL kanamycin until OD_{600} of 0.8. Expression was induced with 0.4 mM IPTG for three hours. Cells were resuspended in 25 mL of 50 mM Tris-HCl (pH=8.0), 100 mM NaCl, 20 mM imidazole and lysed on an Emulsiflex C5 homogenizer (Avestin). Cell lysate was centrifuged at 15 000 *g* for 30 min and supernatant collected. GB1-BRC8-2 lysate was loaded on a 3 mL Ni-NTA agarose matrix (Cube Biotech), followed by the application of monomeric RAD51 lysate. Column matrix was washed with 5 column volumes 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 20 mM imidazole. Complex was eluted with 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 200 mM imidazole into 2 ml fractions. Fractions containing the proteins of interest were pooled and incubated with 100 µL of 2 mg/ml TEV protease overnight at 4°C. Cleaved GB1 fusion partner was removed from the solution by a second Ni-NTA affinity step, collecting the flow-through that contains the monomeric RAD51:BRC8-2 complex. Flow-through was concentrated on a centrifugal filter (Amicon, MWCO 3000 Da) to 2 ml volume and loaded into a Superdex 75 16/60 prep grade size exclusion column (GE Lifesciences), previously equilibrated with 20 mM CHES pH 9.5, 100 mM NaCl, 1 mM EDTA. The complex eluted at 75 ml, the fractions containing the complex were pooled and the complex concentrated to 0.45 mM.

Fluorescence polarization competition assay and microfluidic based measurements

The method was essentially as described in our previous paper (24). Briefly, nanoliter droplets were generated from a well in which an increasing amount of peptide is allowed to compete with BRC4^{fl}. The 4-channel parallelized device was used for all measurements. A 10x objective was used and the power of the 488nm diode laser was 50 mW. BRC peptides were pre-loaded into PTE tubing (internal Ø, 0.38mm) to avoid cleaning syringes between runs. To this end, 40 µL of each peptide was aspirated in tubing, followed a plug of 10 µL of HFE-7500 & 0.5% Pico-Surf 1. Typically, 4 peptides were pre-loaded in each tubing, so that 16 samples could be screened at a go. All measurements were performed in CHES buffer pH 9.5, 1% BSA with HFE-7500 oil & 0.5% Pico-Surf 1 surfactant as carrier phase. Flow rates were 3 µL/min for withdrawal and 40 µL/min for 30 seconds for peptide injection. Data was fit to a competitive binding model using 12 nM for the K_d of the monomeric RAD51:BRC4^{fl} interaction (24).

Crystallography of monomeric RAD51:BRC8-2 complex

Monomeric RAD51:BRC8-2 complex was crystallized using sitting-drop vapor diffusion in a 96-well MRC plate format. 40 mM ADP/Mg^{2+} water solution was added to 0.45 mM complex in a 1:9 ratio. 200 nL of the complex was then mixed with 200 nL of the crystallization condition using a Mosquito liquid handling robot (TTP Labtech). Crystals were observed in 0.2 M NH$_4$Cl, 20% w/v PEG 3350 and used directly for data collection without the need for further optimization of the crystallization conditions.

A crystal was cryo-cooled in liquid nitrogen without the application of a cryo-protectant and diffraction data were collected at Diamond Light Source (Harwell, UK) synchrotron radiation source. Images were

4

processed with autoPROC (26). Molecular replacement phasing method was used with apo monomeric RAD51 structure ('HumRadA22', PDB: 5KDD) as a search model. The structure was refined without peptide first and the peptide was built into the clearly visible electron density manually (see Supplementary Figure S10). Manual real-space refinement was done in Coot (27) and automated refinement with phenix.refine (28) and autoBUSTER (29). Crystallographic data and refinement statistics are shown in Supplementary Table S5. The final model contains 7 complexes of monomeric RAD51 complexed with BRC8-2 peptide and one monomeric RAD51 chain with no peptide (chain H). Chain H has poorly defined electron density, which is likely caused by the lower number of crystal contacts it makes compared to other monomeric RAD51 molecules in the asymmetric unit. Individual atomic B-factors were not refined for chain H. The protein structure is fully defined in all of the complexes, but the peptide density had more variable quality. Chains B and J represent the best-defined monomeric RAD51:BRC8-2 complex and were used in the analysis. The coordinates and corresponding structure factors have been deposited to the PDB under accession number 6HQU. Contact area within the complex of both monomeric RAD51:BRC8-2 (6HQU) and RAD51:BRC4 (1n0w) was calculated (for atoms within 3.9 Å distance of atoms of the other binding partner) using a Pymol script written by Martin Christen (contact_surface v.3.0, available at https://pymolwiki.org/index.php/Contact_Surface). The script was adapted for Python3 using the 2to3 program (https://docs.python.org/2/library/2to3.html).

Cell line

U2OS cell line (ATCC, HTB-96) was grown in DMEM media supplement with 10% FBS (Gibco™ Fetal Bovine Serum, 11573397) and 100 U/mL penicillin/streptomycin (15140122, Gibco) at 37 ºC and 5% CO_2.

Transfection and cell treatment

Cells were transfected using Lipofectamine 3000 Transfection Reagent (Invitrogen) following manufacture's protocol. Plasmid DNA and transfection reagent amounts were scaled for a 10-cm dish; 4 μg DNA, 7.75 μL Lipofectamine 3000 Reagent and 7.75 μL Lipofectamine 3000.

Approximately 18 hours after transfection, cells were either exposed to 3 Gy caesium-137 g-irradiation (GammaCell 1000, Atomic Energy of Canada Ltd) or unirradiated and allowed to recover for 3 hours before being collected for analysis.

Immunostaining

Coverslips were washed twice in phosphate buffered saline (PBS) before fixation with 4% paraformaldehyde/PBS for 15 minutes, then washed three times in PBS and permeabilized in 0.5% Triton-X/PBS for 7 minutes. Coverslips were washed three times in PBS, blocked for at least 30 min in 1% BSA-Fraction V (A3059-50G, Sigma-Aldrich)/PBS, and followed by 1 hr incubation at room temperature with RAD51 (RAD51 H-92, sc-8349, Santa Cruz) primary antibody diluted 1:100 in 1% BSA-Fraction V/PBS. The coverslips were washed three times with PBS, then incubated with anti-rabbit Alexa Fluor 647 secondary antibody (A21244, Invitrogen) diluted 1:500 in 1% BSA-Fraction V/PBS for

45 min in the dark at room temperature. Coverslips were wash three times in PBS, mounted on to slides using ProLong Gold Antifade Mountant with DAPI (P36941, Invitrogen) and stored at 4 ºC for further analysis.

Cells were visualized using a Nikon Eclipse e-400 microscope with 60X objective. Images were processed and analyzed for GFP signal and RAD51 foci using CellProfiler 3.1.8.

Flow cytometry

Cells were trypsinized, washed twice in PBS and fixed by gently vortexing while adding 1 mL ice-cold 70% ethanol drop-wise. Samples were stored for a minimum of 12 h at -20 ºC. Prior to flow cytometry analysis, cells were spun down, washed twice in PBS and resuspended in around 0.5 mL Staining Solution (5 μg/mL propidium iodide (P3566, ThermoFisher Scientific), 100 μg/mL RNase A (R5503-100MG, Sigma-Aldrich) in PBS) and incubated for at least 30 min in the dark at room temperature. Cells were analyzed on a BD LSR II Flow Cytometer (BD Biosciences) and cell cycle profiles were generated after gating GFP positive cells using FlowJo v10.1 software. Around 10,000 GFP positive cells were analysed per condition and experiment.

RESULTS

Systematic shuffling of the two binding modules of the eight BRC repeats and evaluation of the 64 resulting chimerae in a microfluidic fluorescence anisotropy assay

To be able to shuffle the two modules of the eight BRC repeats found in BRCA2 that bind RAD51 (Figure 1A) a crossover point was defined immediately C-terminal to the hairpin structure found in the FxxA module (Figure 1B), as suggested by the RAD51:BRC4 crystal structure (18). The resulting 64 variant BRC peptides (Figure 1C) were cloned through an oligonucleotide cassette method (Supplementary Figure S1 & Table S1), as C-terminal fusions to the GB1 domain from protein G (Supplementary Figure S2; we will refer to shuffled variants by two digits denoting the identity of the N-terminal FxxA and the C-terminal LFDE module, respectively, e.g. BRC2-4 is a peptide with the FxxA module from BRC2 and the LFDE module from BRC4). To confirm that the GB1 domain does not interfere with the BRC-repeat interaction, an isothermal titration calorimetry (ITC) affinity measurement was carried out to confirm that the BRC4 peptide, upon fusion to the GB1 C-terminus, maintained its ability to bind HumRadA22 (a faithful yet monomeric model of RAD51 and for simplicity we will refer to this as monomeric RAD51) (25)) (Supplementary Figure S5 & Figure S6 and accompanying Supplementary Text 3) and matched affinities previously measured for BRC4 peptide (25, 30, 31).

To obtain binding data from small sample quantities, a microfluidic setup to measure dose-response curves by coupling nanoliter 'droplet-on-demand' formation with fluorescence anisotropy measurements was employed (24). The 64 different variant peptide repeats were assayed in a competition assay (Figure 1D) with a fluorescence anisotropy imaging platform using a fluorescently labelled BRC4 repeat (BRC4[fl]-peptide) as the tracer (Figure 1E) (24). Although the droplet-on-demand method supported procurement of measurements over at least two orders of magnitude in concentration

of titrant, we expected even larger differences in affinity between the different chimeras. Consequently, the BRC repeats were diluted to an appropriate concentration (as established by an initial, single concentration-point screen, Table S3) for acquisition of a dose-response profile in microdroplets (24), representing individual titrations of the chimeras (Supplementary Figure S7). High yields of peptides obtained during protein expression enabled screening close to saturating conditions (with the exception of the poorest binders), resulting in good quality data which could be fit to a competitive binding model (Figure 2A) to derive dissociation constants for all 64 variants (Figure 2B).

Shuffling leads to BRC peptide binders with enhanced affinity over wildtype

The 64 measured K_d values for BRC peptide binding to monomeric RAD51 spanned a range of three orders of magnitude from 11 µM (BRC7-8) to 6 nM (BRC8-2). To the best of our knowledge, most of the shuffled variants had never been measured before (except BRC4-5 and BRC5-4 (19)), so it is useful to compare the values we measured here for the natural repeats (which can be read as a diagonal from the top left corner to bottom right corner in Figure 2B) to previously reported values. First, using this competitive fluorescence anisotropy assay, BRC4 was found to have a K_d of 38 nM for monomeric RAD51, a value that falls within the range (6.2-64 nM) of previously measured affinities for this protein-protein interaction (25). Also, BRC4 was found to be the tightest binder of all the natural repeats, in agreement with previous studies (32, 33). The repeats 1, 2, 3 and 4 displayed higher affinity (median K_d 245 nM) than repeats 5, 6, 7 and 8 (median K_d 1636 nM). This is consistent with previous observations that BRC repeats 1 to 4 have higher affinity than repeats 5 to 8 for uncomplexed RAD51 (33). Beyond this broad analysis, more detailed comparisons (including the absolute K_d values) to previous studies are of limited value as there are important differences in the assays employed and the binders titrated against, e.g. full length RAD51 (33, 34) or truncated RAD51 consisting of the catalytic domain only (32).

Having established that the affinity ranking for the natural repeats were consistent with previous reports, we next analyzed the affinities we found for the novel recombinant peptides. The combinations with the FxxA module of BRC5 were found to be the weakest binders, easily explained by the fact that the conserved alanine in the FxxA module of BRC5 is replaced by a serine (FYTS), which has a hydrophilic side-chain that would not form favorable steric and electrostatic contacts with the compact Ala pocket. Interestingly, the recombinant peptide BRC4-5 was also found to be a relatively weak binder (K_d 2 µM), despite previous findings that this chimera displayed a stronger affinity (19). This paradox is addressed below. Remarkably, a few chimeras containing the FxxA module from BRC8, a natural repeat from the 'weak' group of repeats 5-8, were found to be the strongest binders from the entire set of 64 variants (Figure 2B) with BRC8-2 being the peptide with highest affinity with a K_d of 6 nM. This highlights the that shuffling can lead to the bringing together of elements that are in a non-optimal combination in nature.

7

Discerning module-specific contributions to binding affinity and the effect of the cross-over point placement

To allow comparisons across repeats and modules, the dissociation constants were expressed in units of Gibbs free energy (ΔG, Table S4). The effect of shuffling was quantified by expressing each of the 56 novel, unexplored combinations in terms of relative change vis-à-vis their two parental repeats, e.g. the two parents of BRC1-2 are BRC1 and BRC2 (Figure 2C & Supplementary Information Text 5.1). A positive value for $\Delta\Delta G_{parental}$ – indicating that the product of the shuffling was detrimental to the binding function – was observed in 33 out 56 peptides. Also, the average $\Delta\Delta G_{parental}$ for all 56 shuffled repeats was 0.16 kCal/mol, indicating that shuffling had a net-detrimental effect on binding function. Nevertheless, 23 out of these 56 repeats had negative $\Delta\Delta G_{parental}$ values and thus represented variants that were improved over the average of their parents. We asked whether the identity of the FxxA module affected binding to monomeric RAD51 more strongly than the identity of the LFDE module. At a first approximation, the FxxA module might appear to have dominated the interaction as proven by the fact that any combination with imperfect FxxA module of BRC5 (containing the stretch of residues Ser1662^{BRC5} to Arg1677^{BRC5}) resulted in exceedingly weak interactions ($\Delta\Delta G_{FxxA5}$ = 0.91 kCal/mol) (Figure 2C). The LFDE modules from BRC3 and BRC8 were found to cause the most significant reduction to binding in each of their respective seven recombinant peptides ($\Delta\Delta G_{LFDE3}$= 1.07 kCal/mol; $\Delta\Delta G_{LFDE8}$ = 1.05 kCal/mol). As both the FxxA and LFDE modules in each repeat could thus make a significant contribution to binding, we considered whether the net contribution to binding was equally distributed within each repeat. In BRC repeat 5, both the FxxA and LFDE modules conspired to make a weak binder ($\Delta\Delta G_{FxxA5}$ = 0.91 kCal/mol; $\Delta\Delta G_{LFDE5}$= 0.33 kCal/mol). By contrast, BRC repeat 8 could be considered 'Janus-faced', as it is composed of a net contributor ($\Delta\Delta G_{FxxA8}$= -0.96 kCal/mol) and a net disruptor ($\Delta\Delta G_{LFDE8}$= 1.05 kCal/mol) to binding. To a slightly lesser degree, BRC repeat 2 displayed the same contrast in intra-repeat properties, although in this natural repeat, the FxxA module was a net disruptor overall ($\Delta\Delta G_{FxxA2}$ = 0.82 kCal/mol), while the LFDE module was a net contributor ($\Delta\Delta G_{LFDE2}$ = -1.15 kCal/mol). This analysis is validated by the observation that BRC8-2, which combines the overall best FxxA module with overall best LFDE module, is the BRC peptide with the highest affinity of all, which we were able to cross validate by ITC measurements (Supplementary Figure S5) and is also the most improved over its parental sequences ($\Delta\Delta G_{parental}$ = -2.44 kCal/mol). The rank affinity ordering of individual modules based on their $\Delta\Delta G_{FxxA1-8}$ or $\Delta\Delta G_{LFDE1-8}$ values, next to the rank order of the natural repeats' affinities (Figure 2D), highlighted that within the eight natural BRC repeats, binding function was not always equally distributed between modules.

The observation of a $\Delta\Delta G_{parental}$ of 1.24 kCal/mol (Figure 2C) for BRC4-5 represented a notable discrepancy to previous data, obtained by competitive ELISA with synthetic chimeric peptides BRC4-5 & 5-4 (19). As expected, in agreement with our findings, BRC5-4 turned out to be a weak binder, due to the lack of conservation in repeat 5's FxxA module. However, Rajendra and Venkitaraman found BRC4-5 to be a *stronger* binder than the natural BRC4, whereas we found BRC4-5 to bind monomeric RAD51 with 55-fold *lower* affinity than BRC4. What could explain this? Apart from the obvious difference in the assays (heterogeneous ELISA-based assay vs homogeneous polarization-based assay), the main remaining difference is the cutoff between the end of the FxxA module and the start of the LFDE

module in the shuffled peptide. While Rajendra & Venkitaraman defined Lys1530[BRC4] as the last residue of the FxxA module and Ile1534[BRC4] as the first of the LFDE module, our peptides were based on the cut-off point occurring between Lys1530[BRC4] and Lys1531[BRC4]. The cutoff is arbitrary but resulted in our chimeric BRC4-5 repeat bearing two mutations compared to the other study, Val1532[BRC4] → Thr1679[BRC5] and Lys1533[BRC4]→ Ser1680[BRC5]. When we applied the SCHEMA computational algorithm to identify the optimal cross-over points in a protein sequence (10), the exact cross-over employed by Rajendra and Venkitaraman was found to be optimal (Supplementary Figure S8 & Supplementary Text 5.2). Furthermore, we found that the deletion of Lys1530[BRC4], located at the cross-over between BRC4's two modules, resulted in a significant loss of affinity (the K_d shifted from 21 nM to 4.1 µM, Supplementary Figure S9 & Supplementary text 5.3). Thus, subtle differences in the linker region can lead to dramatic differences in affinity, explaining the critical effect of the exact placement of shuffle cut-off points.

BRC8-2 forms a more extensive β-hairpin and has improved helicity compared to BRC4

To gain structural insight into the increased affinity of BRC8-2 for RAD51, we determined a 1.95 Å crystal structure of the monomeric RAD51:BRC8-2 complex (PDB ID: 6HQU). There are eight complexes in the asymmetric unit of these crystals, all of which are very similar to each other with average RMSD of 0.664 Å for 198 Cα atoms in monomeric RAD51. The BRC8-2 peptide was visible in seven out of eight complexes. In all cases, the overall conformation of the peptide was the same and we used the best-defined complex (chains B and J) in the subsequent analyses. Representative electron density difference maps of the BRC8-2 peptide prior fitting and after final refinement are provided in Supplementary Figure S10. Comparison of the refined structure for monomeric RAD51:BRC8-2 with that of the RAD51:BRC4 complex (18) (PDB: 1n0w) reveals a similar overall topology (Figure 3A). The FxxA module of BRC8 interacts with monomeric RAD51, with Phe2058[BRC8] and Ala2061[BRC8] forming identical contacts to those seen between BRC4 and RAD51. C-terminal to Ala2061[BRC8], the peptide forms a β-turn and bends back in the opposite direction, resulting in a β-hairpin structure that extends the central β-sheet of monomeric RAD51 in an inter-molecular fashion, reminiscent of the RAD51:BRC4 complex (18, 35). Five residues at the C-terminal ends of the BRC8 and BRC4 FxxA modules are identical in sequence (TASGK) and both hairpins are stabilized by the Thr2060[BRC8]/Thr1526[BRC4] hydroxyl groups, which hydrogen-bond with the nearby backbone NH of Lys2064[BRC8]/Lys1530[BRC4] and the hydroxyl of Ser2062[BRC8]/Ser1528[BRC4] (Figure 3B). In the RAD51:BRC4 complex, the C-terminal LFDE module forms a ten-residue α-helix that interacts with RAD51 through a shallow interface, comprised of a mixture of hydrophobic and polar contacts. The LFDE motif of BRC4 is replaced by LFSD in the BRC2 module. Leu1240[BRC2] and Phe1241[BRC2] bind the same hydrophobic interface that BRC4 interacts with and Asp1243[BRC2] interacts with a nearby Arg270[monomeric RAD51] as seen in BRC4 (Figure 3C). There was little difference between the contact area calculated for the RAD51:BRC4 and monomeric RAD51:BRC8-2 complexes (1042 and 942 Å^2, respectively), consistent with the previously noted weakness of the correlation between buried surface area and binding affinity (36).

The most obvious difference between the binding modes of the BRC4 and BRC8-2 peptides is in the extent of the intra-molecular hydrogen-bonding network that forms the β-hairpin in the FxxA module

9

(Figure 3B). In BRC8-2, the β-hairpin is significantly extended, with its N-terminal end, before Phe2058[BRC8], folded back towards the rest of the peptide. As a result, the hairpin feature extends from Ser2053[BRC8] to Thr1231[BRC2], a total of 19 amino acids. The hairpin in BRC4 is formed by only nine residues. A likely structural determinant for the formation of the extended hairpin is Ser2056[BRC8], whose side chain fits tightly between the two anti-parallel strands of the peptide and the surface of monomeric RAD51, is hydrogen bonding with the carbonyl of Leu1227[BRC2] and the NH of Phe2058[BRC8]. This allows the peptide to fold back on itself and to form the extended intra-molecular H-bonding network (Figure 3B). BRC4 has a bulky and hydrophobic Leu1522 in the equivalent position of Ser2056[BRC8], which is not able to satisfy the steric and electrostatic requirements of the topology we observe in BRC8-2, forcing the N-terminus of the peptide to point away from the β-hairpin and the rest of the peptide.

The observation that peptides in *all* complexes in the asymmetric unit maintain the same conformation, despite the different packing interactions that surround them, argues against the possibility that the extended β-hairpin of BRC8-2 within the RAD51:BRC8-2 complex is a crystal packing artefact. Interestingly, the FxxA module from BRC3 has the second highest $\Delta\Delta G_{FxxA1-8}$ ($\Delta\Delta G_{FxxA3}$ = - 0.35 kCal/mol) and contains a threonine at the position equivalent to Ser2056, which may similarly enable an extended hairpin conformation by hydrogen-bonding to the backbone amides of residues equivalent to Leu1227[BRC8] and Phe2058[BRC8]. To examine the contribution of Ser2056 to binding, we designed a (untagged) mutant repeat BRC8-2[S2056A]. Its affinity for monomeric RAD51 was measured as a K_d of 5 nM (Supplementary Figure S5F), i.e. an order of magnitude lower affinity than that measured for untagged BRC8-2 (Supplementary Figure S5E), confirming the significance of Ser2056 for binding.

There are also differences in the binding modes of the LFDE module of BRC8-2 compared to BRC4 (Figure 3C). In BRC4, the interface-forming residue Val1542[BRC4] has a bulky hydrophobic side chain projected into the ATPase domain that disrupts the optimal helical geometry of the peptide, forming an outward-facing bulge. This residue is changed to Ala1237[BRC2] in BRC8-2, causing a shift of the α-helix backbone towards the surface of the protein and resulting in a more optimal geometry of the helix and closer interaction with monomeric RAD51. Thus, the increased binding affinity of the BRC8-2 repeat for monomeric RAD51 appears to result from the extended hydrogen-bonding network of the FxxA module and the improved helical geometry of the LFDE module in BRC8-2.

BRC8-2 can disrupt RAD51 foci formation in cell-based experiments

To validate the utility of improved binding of BRC8-2 binding to RAD51 for biological intervention, we investigated the ability of this peptide to disrupt RAD51 function in human cells. Following treatment with ionizing radiation (IR), RAD51 translocates to the sites of DNA damage and forms foci that are visible by immunofluorescence. The formation of such foci is dependent on BRCA2 (37), and it has previously been shown that foci formation can be disrupted by expression of native BRC repeats (38, 39).

To determine whether BRC8-2 impaired RAD51 foci formation, the peptide was fused to a green fluorescent protein (GFP) containing a nuclear localization signal (NLS) and transfected into U2OS osteosarcoma cells. Cells were transfected at the same time with the parental GFP-NLS construct and

RAD51 foci formation was monitored in GFP positive cells (Figure 4A). As expected, the control GFP cells showed an increase in the median number of RAD51 foci after exposure to ionizing radiation (IR, 3 Gy) (Figure 4B and 4C). In addition, a small number of foci were present in the absence of IR, most likely reflecting HR events associated with replicative stress. In contrast, cells expressing GFP-BRC8-2 had fewer RAD51 foci prior to irradiation and following IR exposure (Figure 4B and 4C). Because RAD51 foci formation is limited to the S and G2 phases of the cell cycle, we wanted to determine whether the reduction in foci formation was due to an increase in G1 phase cells in the GFP-BRC8-2 expressing cells. We therefore monitored the cell cycle profile of GFP expressing cells by FACS and found that the GFP-BRC8-2 expressing cells have a decrease in the proportion of G1 phase cells, indicating that the effect on RAD51 foci formation is not due to cell cycle alterations. These data suggest that BRC8-2 interferes with RAD51 foci formation in human cells through binding and sequestering RAD51 away from sites of DNA damage. In support of this, we also noted that the pan-nuclear signal of RAD51 in GFP-BRC8-2 expressing cells is greater than in the GFP control cells (Supplementary Figure S11).

DISCUSSION

Functional implications of module-specific contributions to binding across the BRCA2 repeats

Affinity measurements of all eight naturally occurring BRCA2 repeats brings about a rank order of affinity as 4>1>3>2>8>7>5>6, largely matching previously established literature values (4>2>1>8>7>3>5=6 from (32) and 2>4>1>3>5>6=7=8 from (33)). Broadly, repeats BRC1, 2, 3 and 4 bind the monomeric RAD51 more strongly than repeats BRC5, 6, 7 and 8. These two different affinity levels would seem to suggest alternative functions. Indeed, reports are emerging of the specialist functioning of two different subgroups of the BRCA2 repeats, where affinity levels are also sometimes inverted depending on the RAD51 complex (33, 40, 41). Although repeats 1-4 were found to bind *free* RAD51 more strongly than repeats 6-8, repeats 5-8 were reported to have higher binding affinity for the RAD51:ssDNA complex than repeats 1-4 (33). Repeats 5-8 may also bind in concert to stimulate certain RAD51 functions (41). The affinity levels of the two groups of repeats were also found to be inverted in the context of interactions with DMC1, a meiosis-specific DNA strand exchange protein (40), highlighting the adaptability of the BRCA repeats. BRCA2 repeats 6, 7 and 8 were shown to bind to DMC1 more strongly than the subgroup 1-4. In fact module 4, the strongest RAD51 binder, was found to be the weakest DMC1 binder (40). Quite how the different BRC repeat groupings achieve differential specificity for the various targets (e.g. RAD51 vs ssDNA-bound RAD51 vs DMC1) remains poorly understood.

It is interesting to speculate whether the unequal distribution of binding function across repeats and across modules has any functional implications. The two elements in the BRC repeats, the FxxA and the LFDE module, could be readily recombined resulting in functional chimeras with altered properties, including surprising increases in binding affinity for some module combinations. The two modules interact with different sites on RAD51, with the FxxA module disrupting the RAD51 dimer interface, whereas the LFDE module could interact with the oligomeric form of RAD51 and perhaps function as a docking site for the BRC repeats, facilitating the disruptive function of the FxxA modules. The function of individual repeats might depend on the relative affinity of the two modules for RAD51, and the so-

called 'weak' RAD51 binding repeats 5-8 have functions that are different from the 'strong' binders, suggesting that affinity ranking of the full repeats is misleading.

Shuffling of modules: taking full advantage of the sum of anciently diverged parts

If combinations of modules achieve a wide range of affinities, their assembly context may matter (in addition to intrinsic effects of each module), pointing to cooperative effects of the different modules. We found that both the FxxA and LFDE modules of the eight BRC repeats make relevant contributions to the binding of monomeric RAD51, as evidenced by both FxxA and LFDE being associated with net disruption of binding function upon shuffling with other modules. This is also consistent with the finding by previous studies that the FxxA module in isolation (i.e. lacking the LFDE module) has only a modest affinity for RAD51 (19, 31, 39). Interestingly, we found that both modules often failed to act in concert to give high affinity binders within the natural repeats.

Our shuffling approach helps to pick apart the role played by both modules in their various repeats, allowing us to discern module-specific effects that are otherwise obscured when measured in their parental combinations. The BRC repeats provide a fascinating example of how Nature can exploit modularity to fine-tune function. By recycling variations on the FxxA β-hairpin and LFDE α-helix 'themes', a range of affinities (spanning two orders of magnitude for the natural BRC repeats) was achieved, without the need to resort to entirely novel sequences.

It is interesting to compare our approach based on shuffling of natural diversity with previous efforts aimed at understanding and/or improving BRCA2 repeat affinity for RAD51. These efforts are partially motivated by the therapeutic potential of blocking the BRCA2-RAD51 interaction. Nomme *et al* succeeded, through careful rational analysis and sophisticated molecular modeling techniques, in generating a BRC4 repeat peptide mutant that was 10-times more efficient in inhibiting the RAD51:ssDNA complex than the original BRC4 repeat peptide itself (31). Similarly, Venkitaraman, Skylaris and colleagues calculated molecular mechanics energies combined with the Poisson–Boltzmann surface area continuum solvation (MM-PBSA) to successfully identify the BRC repeat binding hotspots as well as enabling an accurate prediction of relative binding free energies of the natural BRC repeats (32). Scott *et al* probed the contribution of individual residues in the FxxA epitope identifying changes that resulted in increased affinity towards RAD51 (42). Unlike typical computational approaches, the affinity enhancement achieved by our shuffling approach requires no *a priori* knowledge of the binding mechanism. Guided by the fast evaluation of K_d values in microfluidic droplets, empirical models can be developed that yield novel insights into binding mechanism, and practically chimeras with improved affinity for use in various diagnostic and therapeutic applications can be obtained.

Structural insight into beneficial effect of shuffling and scope for future work

The crystal structure of the RAD51:BRC8-2 complex has provided novel insight into the BRC repeat binding to RAD51, in particular in identification of the extended β-hairpin formed by the FxxA module and the critical role Ser2056[BRC8] in facilitating the formation of the structure. The enhanced affinity of BRC8-2 may render it an attractive tool in studies that seek to investigate the effect of disrupting the

RAD51:BRCA2 interaction using cell penetrating peptide derivatives of BRC repeats (39, 43). As a first step, we were able to confirm the utility of BRC8-2 in a functional cellular assay for disruption of radiation-induced RAD51 foci formation. It is tempting to speculate that further stabilization of this β-hairpin, for example through the use of tryptophan-tryptophan cross-strand pairs (44) or even artificial cross-links such as triazole (45) might result in an additional enhancement of binding through a reduction in conformational heterogeneity prior to complex formation. Our work also revealed that the linker between the FxxA module and LFDE module is also likely to play a role in determining the affinity of the peptide. Our data on the importance of the linker region is further corroborated by a study showing that mutating the wildtype Val1532^{BRC4} to Ile or Phe results in respectively enhanced and diminished binding of BRC4 peptide to RAD51 (46).

Based on the structural insights, the crossover points that define functional units could be revised for future engineering: by splitting the BRC repeats into *three* parts, namely an N-terminal FxxA end, middle 'linker region' and C-terminal LFDE end, further libraries of chimerae (now with 512 members) could be constructed and screened over the course of a few days using our droplet approach. The increasing amount of sequence data can be harvested to create larger repertoires of building blocks: BRC4 repeats from some mammalian species can bind *human* RAD51 with higher affinity than human BRC4 (46), indicating that libraries with potential for functional improvement can be constructed through inclusion of repeat module sequences from the diversity of mammalian homologues of BRCA2.

Conclusion

Features that may help to combine rational and combinatorial engineering productively emerge from this work. Knowledge of functional units of proteins that are autonomously folded and functional, bypasses the need to design proteins from scratch, while their shuffling reduces the library complexity vastly, compared to total sequence randomization. Defining such functional modules better will provide the basis for more sophisticated libraries as well as design approaches to reach the goal of eliciting functional proteins more quickly.

ACCESSION NUMBER

The coordinates and corresponding structure factors for the monomeric RAD51:BRC8-2 complex have been deposited to the PDB under accession number 6HQU.

SUPPLEMENTARY DATA

Supplementary data contains detailed descriptions of the cloning of bacterial expression constructs for the 64 shuffled BRC peptide variants, cloning of mammalian expression constructs and notes on the soluble expression of the shuffled BRC peptide variants. Also included is a description of ITC used to cross-validate the microfluidic measurements, single concentration point measurements carried out with microfluidics and exemplary titrations carried out by microfluidics. The supplementary data also contains an analysis on the effect of shuffling of BRC peptides and in particular on the effect of the exact shuffle cut-off point placement. X-ray crystallography electron density maps, data collection and

refinement statistics are also to be found in the supplementary data. Additional cell images highlighting the pan-nuclear signal of RAD51 are also included in the supplementary data.

AUTHOR CONTRIBUTIONS

LL, FG, MB, MH & FH conceived the study. LL, MB & TP produced recombinant proteins. FG, ER & CFK set up the fluorescence anisotropy equipment. FG carried out the high throughput fluorescence anisotropy measurements. FG & LL analyzed the fluorescence anisotropy data. TP purified the monomeric RAD51:BRC8-2 protein complex, determined its crystal structure, conducted ITC measurements and analyzed the data. MH & TP interpreted the structural data. PZ-V and JAD designed and performed the cell-based assays. LL, TP, FG, MH & FH wrote the paper.

REFERENCES

1. Söding,J. and Lupas,A.N. (2003) More than the sum of their parts: On the evolution of proteins from peptides. *BioEssays*, **25**, 837–846.

2. Moore,A.D., Björklund,Å.K., Ekman,D., Bornberg-Bauer,E. and Elofsson,A. (2008) Arrangements in the modular evolution of proteins. *Trends Biochem. Sci.*, **33**, 444–451.

3. Bornberg-Bauer,E. and Albà,M.M. (2013) Dynamics and adaptive benefits of modular protein evolution. *Curr. Opin. Struct. Biol.*, **23**, 459–466.

4. Maervoet,V.E.T. and Briers,Y. (2017) Synthetic biology of modular proteins. *Bioengineered*, **8**, 196–202.

5. Scaiewicz,A. and Levitt,M. (2018) Unique function words characterize genomic proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, 6703–6708.

6. Alva,V. and Lupas,A.N. (2018) From ancestral peptides to designed proteins. *Curr. Opin. Struct. Biol.*, **48**, 103–109.

14

7. Stemmer,W.P.C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, **370**, 389–391.

8. Stemmer,W.P.C., Crameri,A., Raillard,S.-A. and Bermudez,E. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, **391**, 288–291.

9. Stemmer,W.P.C., Christians,F.C., Scapozza,L., Crameri,A. and Folkers,G. (1999) Directed evolution of thymidine kinase for AZT phosphorylation usingDNA family shuffling. *Nat. Biotechnol.*, **17**, 259–264.

10. Voigt,C.A., Martinez,C., Wang,Z.-G., Mayo,S.L. and Arnold,F.H. (2002) Protein building blocks preserved by recombination. *Nat. Struct. Biol.*, **9**, 553–8.

11. Bedbrook,C.N., Rice,A.J., Yang,K.K., Ding,X., Chen,S., LeProust,E.M., Gradinaru,V. and Arnold,F.H. (2017) Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E2624–E2633.

12. Kolkman,J.A. and Stemmer,W.P.C. (2001) Directed evolution of proteins by exon shuffling. *Nat. Biotechnol.*, **19**, 423–428.

13. Zhang,J., Yang,H., Long,M., Li,L. and Dean,A.M. (2010) Evolution of Enzymatic Activities of Testis-Specific Short-Chain Dehydrogenase/Reductase in Drosophila. *J. Mol. Evol.*, **71**, 241–249.

14. Wang,X., Gao,B. and Zhu,S. (2016) Exon Shuffling and Origin of Scorpion Venom Biodiversity. *Toxins (Basel).*, **9**.

15. Smithers,B., Oates,M. and Gough,J. (2019) 'Why genes in pieces?'-revisited. *Nucleic Acids Res.*, **47**, 4970–4973.

16. Venkitaraman,A.R. (2014) Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. *Science*, **343**, 1470–5.

17. Bork,P., Blomberg,N. and Nilges,M. (1996) Internal repeats in the BRCA2 protein sequence. *Nat. Genet.*, **13**, 22–3.

18. Pellegrini,L., Yu,D.S., Lo,T., Anand,S., Lee,M., Blundell,T.L. and Venkitaraman,A.R. (2002) Insights into DNA recombination from the structure of a RAD51–BRCA2 complex. *Nature*, **420**, 287–293.

19. Rajendra,E. and Venkitaraman,A.R. (2010) Two modules in the BRC repeats of BRCA2 mediate structural and functional interactions with the RAD51 recombinase. *Nucleic Acids Res.*, **38**, 82–96.

20. Nomme,J., Takizawa,Y., Martinez,S.F., Renodon-Cornière,A., Fleury,F., Weigel,P., Yamamoto,K., Kurumizaka,H. and Takahashi,M. (2008) Inhibition of filament formation of human RAD51 protein by a small peptide derived from the BRC-motif of the BRCA2 protein. *Genes Cells*, **13**, 471–81.

21. Scott,D.E., Coyne,A.G., Venkitaraman,A., Blundell,T.L., Abell,C. and Hyvönen,M. (2015) Small-molecule inhibitors that target protein-protein interactions in the RAD51 family of recombinases. *ChemMedChem*, **10**, 296–303.

22. Bignell,G., Micklem,G., Stratton,M.R., Ashworth,A. and Wooster,R. (1997) The BRC repeats are conserved in mammalian BRCA2 proteins. *Hum. Mol. Genet.*, **6**, 53–8.

23. Warren,M., Smith,A., Partridge,N., Masabanda,J., Griffin,D. and Ashworth,A. (2002) Structural analysis of the chicken BRCA2 gene facilitates identification of functional domains and disease causing mutations. *Hum. Mol. Genet.*, **11**, 841–851.

24. Gielen,F., Butz,M., Rees,E.J., Erdelyi,M., Moschetti,T., Hyvönen,M., Edel,J.B., Kaminski,C.F. and Hollfelder,F. (2017) Quantitative Affinity Determination by Fluorescence Anisotropy Measurements of Individual Nanoliter Droplets. *Anal. Chem.*, **89**, 1092–1101.

25. Moschetti,T., Sharpe,T., Fischer,G., Marsh,M.E., Ng,H.K., Morgan,M., Scott,D.E., Blundell,T.L., R. Venkitaraman,A., Skidmore,J., *et al.* (2016) Engineering Archeal Surrogate Systems for the

Development of Protein–Protein Interaction Inhibitors against Human RAD51. *J. Mol. Biol.*, **428**, 4589–4607.

26. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T., Bricogne,G. and IUCr (2011) Data processing and analysis with the *autoPROC* toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 293–302.

27. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.

28. Afonine,P. V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**, 352–367.

29. Bricogne,G., Blanc,E., Brandl,M., Flensburg,C., Keller,P., Paciorek,W., Roversi,P., Sharff,A., Smart,O.S., Vonrhein,C., *et al.* (2017) BUSTER version 2.10.3.

30. Gielen,F., Hours,R., Emond,S., Fischlechner,M., Schell,U. and Hollfelder,F. (2016) Ultrahigh-throughput-directed enzyme evolution by absorbance-activated droplet sorting (AADS). *Proc. Natl. Acad. Sci. U. S. A.*, **113**, E7383–E7389.

31. Nomme,J., Renodon-Cornière,A., Asanomi,Y., Sakaguchi,K., Stasiak,A.Z., Stasiak,A., Norden,B., Tran,V. and Takahashi,M. (2010) Design of Potent Inhibitors of Human RAD51 Recombinase Based on BRC Motifs of BRCA2 Protein: Modeling and Experimental Validation of a Chimera Peptide. *J. Med. Chem.*, **53**, 5782–5791.

32. Cole,D.J., Rajendra,E., Roberts-Thomson,M., Hardwick,B., McKenzie,G.J., Payne,M.C., Venkitaraman,A.R. and Skylaris,C.-K. (2011) Interrogation of the protein-protein interactions between human BRCA2 BRC repeats and RAD51 reveals atomistic determinants of affinity. *PLoS Comput. Biol.*, **7**, e1002096.

33. Carreira,A. and Kowalczykowski,S.C. (2011) Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms. *Proc. Natl. Acad. Sci.*, **108**, 10448–10453.

34. Subramanyam,S., Jones,W.T., Spies,M. and Spies,M.A. (2013) Contributions of the RAD51 N-terminal domain to BRCA2-RAD51 interaction. *Nucleic Acids Res.*, **41**, 9020–9032.

35. Kowalczykowski,S.C. (2002) Molecular mimicry connects BRCA2 to Rad51 and recombinational DNA repair. *Nat. Struct. Biol.*, **9**, 897–899.

36. Erijman,A., Rosenthal,E. and Shifman,J.M. (2014) How structure defines affinity in protein-protein interactions. *PLoS One*, **9**, e110085.

37. Yuan,S.S.F., Lee,S.Y., Chen,G., Song,M., Tomlinson,G.E. and Lee,E.Y.H.P. (1999) BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo. *Cancer Res.*, **59**, 3547–3551.

38. Chen,C.F., Chen,P.L., Zhong,Q., Sharp,Z.D. and Lee,W.H. (1999) Expression of BRC repeats in breast cancer cells disrupts the BRCA2- Rad51 complex and leads to radiation hypersensitivity and loss of G2/M checkpoint control. *J. Biol. Chem.*, **274**, 32931–32935.

39. Trenner,A., Godau,J. and Sartori,A.A. (2018) A Short BRCA2-Derived Cell-Penetrating Peptide Targets RAD51 Function and Confers Hypersensitivity toward PARP Inhibition. *Mol. Cancer Ther.*, **17**, 1392–1404.

40. Martinez,J.S., von Nicolai,C., Kim,T., Ehlén,Å., Mazin,A. V., Kowalczykowski,S.C. and Carreira,A. (2016) BRCA2 regulates DMC1-mediated recombination through the BRC repeats. *Proc. Natl. Acad. Sci.*, **113**, 3515–3520.

41. Chatterjee,G., Jimenez-Sainz,J., Presti,T., Nguyen,T. and Jensen,R.B. (2016) Distinct binding of BRCA2 BRC repeats to RAD51 generates differential DNA damage sensitivity. *Nucleic Acids Res.*, **44**, 5256–70.

42. Scott,D.E., Marsh,M., Blundell,T.L., Abell,C. and Hyvönen,M. (2016) Structure-activity relationship of the peptide binding-motif mediating the BRCA2:RAD51 protein-protein interaction. *FEBS Lett.*, **590**, 1094–1102.

43. Przetocka,S., Porro,A., Bolck,H.A., Walker,C., Lezaja,A., Trenner,A., von Aesch,C., Himmels,S.-F., D'Andrea,A.D., Ceccaldi,R., *et al.* (2018) CtIP-Mediated Fork Protection Synergizes with BRCA1 to Suppress Genomic Instability upon DNA Replication Stress. *Mol. Cell*, **72**, 568-582.e6.

44. Cochran,A.G., Skelton,N.J. and Starovasnik,M.A. (2001) Tryptophan zippers: stable, monomeric beta -hairpins. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 5578–83.

45. Merritt,H.I., Sawyer,N. and Arora,P.S. (2020) Bent into shape: Folded peptides to mimic protein structure and modulate protein function. *Pept. Sci.*, **112**.

46. Ochiai,K., Yoshikawa,Y., Yoshimatsu,K., Oonuma,T., Tomioka,Y., Takeda,E., Arikawa,J., Mominoki,K., Omi,T., Hashizume,K., *et al.* (2011) Valine 1532 of human BRC repeat 4 plays an important role in the interaction between BRCA2 and RAD51. *FEBS Lett.*, **585**, 1771–1777.

47. Roehrl,M.H.A., Wang,J.Y. and Wagner,G. (2004) A General Framework for Development and Data Analysis of Competitive High-Throughput Screens for Small-Molecule Inhibitors of Protein−Protein Interactions by Fluorescence Polarization [†]. *Biochemistry*, **43**, 16056–16066.

FIGURES

FIGURE 1
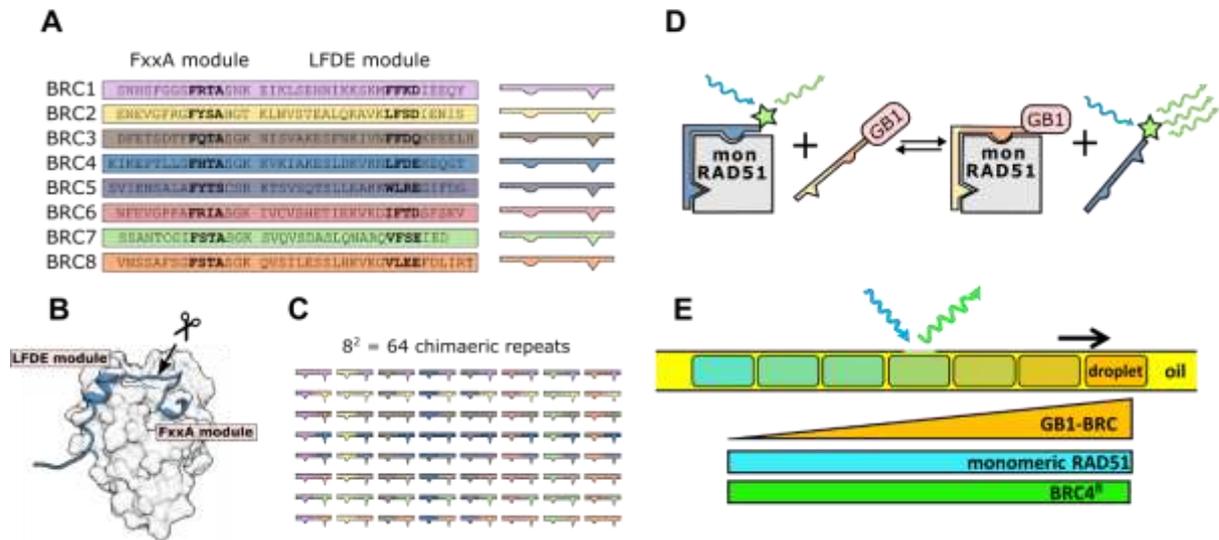
Figure 1. Systematic shuffling of the two binding modules comprising each of the eight RAD51-binding repeats in BRCA2. (**A**) The eight natural human BRC repeats as first identified in 1997 (22). The FxxA and LFDE motifs are highlighted in bold. (**B**) Cartoon representation of RAD51:BRCA2 repeat 4 peptide crystal structure (PDB ID: 1n0w). The arrow indicates the cut site chosen as the crossover point between the FxxA and LFDE modules for this study. This point corresponds to the space left between the sequences for FxxA and LFDE modules in (A). (**C**) Schematic representation of the 56 chimeric and eight natural repeats resulting from the shuffling around the crossover point described in (B). (**D**) Schematic representation of the fluorescence polarization competition assay used for the affinity determination of the BRC repeat shuffle set for monomeric RAD51 (mon RAD51) by a microfluidic droplet-on-demand system. GB1-BRC recombinant peptide fusions were titrated into a complex of monomeric RAD51 and BRC4^{fl} (a fluorescein-labeled BRC4 synthetic peptide). (**E**) Schematic depiction of a droplet train passing through one of the four simultaneously imaged channels, where a concentration gradient of the GB1-BRC variant, together with a constant concentration of monomeric RAD51 and BRC4^{fl}, was moving (in the direction indicated by the arrow) across a fluorescence polarization interrogation point. The quantification of bound complex by fluorescence polarization gives rise to the titration curves shown in Figure 2A that yield the binding affinity.

18
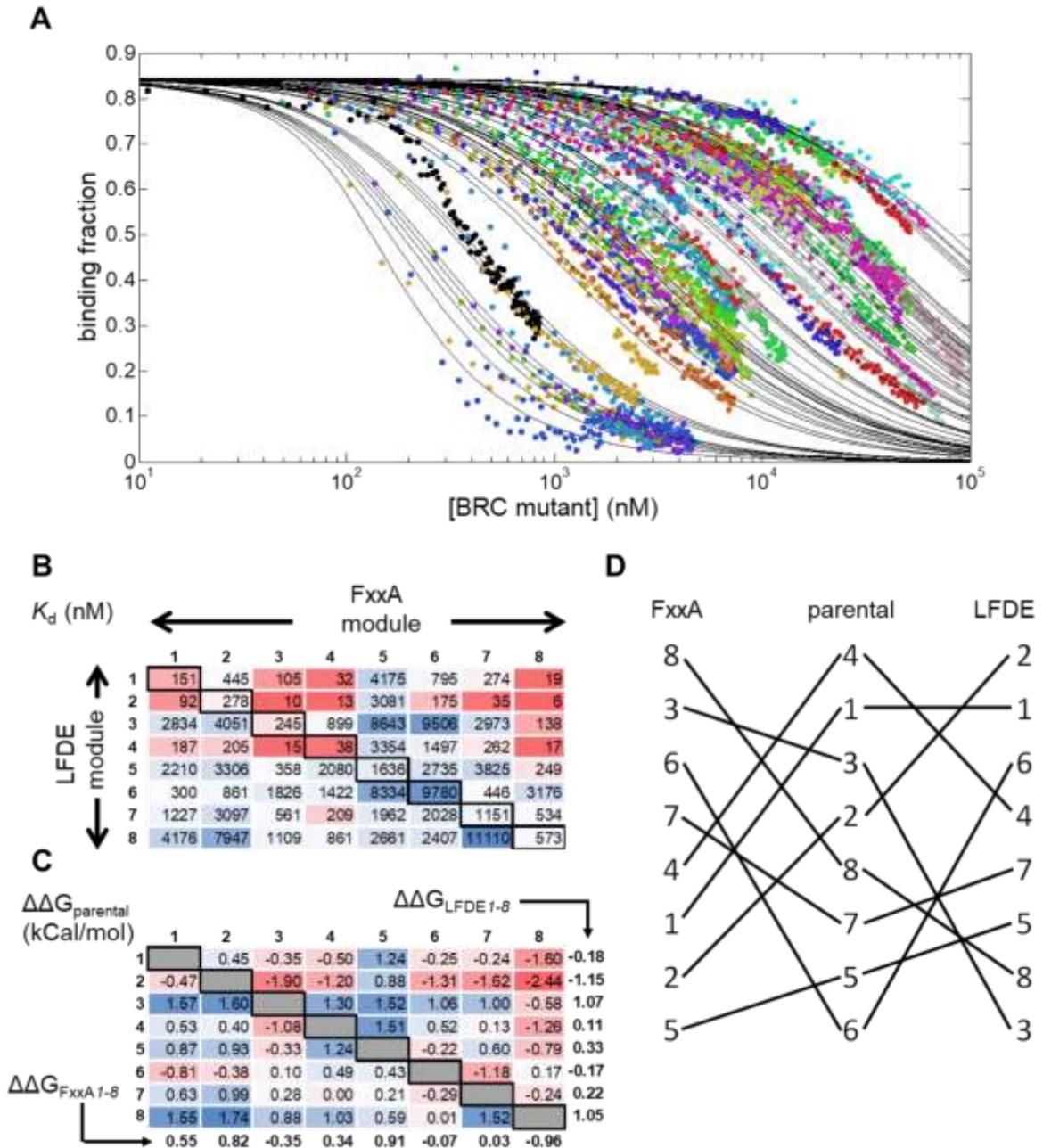
FIGURE 2

Figure 2. Affinity determination of 64 chimeric BRC4 repeats by a microfluidic droplet-on-demand system interfaced with fluorescence anisotropy detection. (**A**) Fraction of BRC4fl-peptide bound to monomeric RAD51 as a function of GB1-BRC peptide chimera concentration, measured using the droplet-on-demand anisotropy competition assay. Measurement conditions were 100 nM BRC4fl, 150 nM monomeric RAD51 in a buffer of 20 mM CHES (pH 9.5), 100 mM NaCl, 1 mM EDTA, at 20 °C. Note the starting binding fraction of 0.85, this may be explained by the affinity of BRC4fl peptide for monomeric RAD51 and the initial concentration used (47). (**B**) K_d values (in nM) determined for all 64 BRCA2 peptide chimeras using data in (A). (**C**) Analysis of the effect of recombination, expressed as the difference in ΔG (Supplementary Table S4) of each shuffled variant relative to the average of the two natural parental combinations (ΔΔG$_{parental}$, in kCal/mol) for each variant. The parental combinations

are depicted in grey (by definition, their $\Delta\Delta G_{parental}$ is always zero). The values indicated below each column and next to each row represent the average $\Delta\Delta G_{parental}$ value for FxxA and LFDE modules from each repeat, respectively and are referred to as $\Delta\Delta G_{FxxA\,1-8}$ and $\Delta\Delta G_{LFDE\,1-8}$, respectively. (**D**) Binding rank order of the parental BRC repeats and the individual FxxA and LFDE modules comprising the repeats. The contributions to binding of the 8 different repeat-derived modules are based on their $\Delta\Delta G_{LFDE\,1-8}$ and $\Delta\Delta G_{FxxA\,1-8}$ values.

FIGURE 3
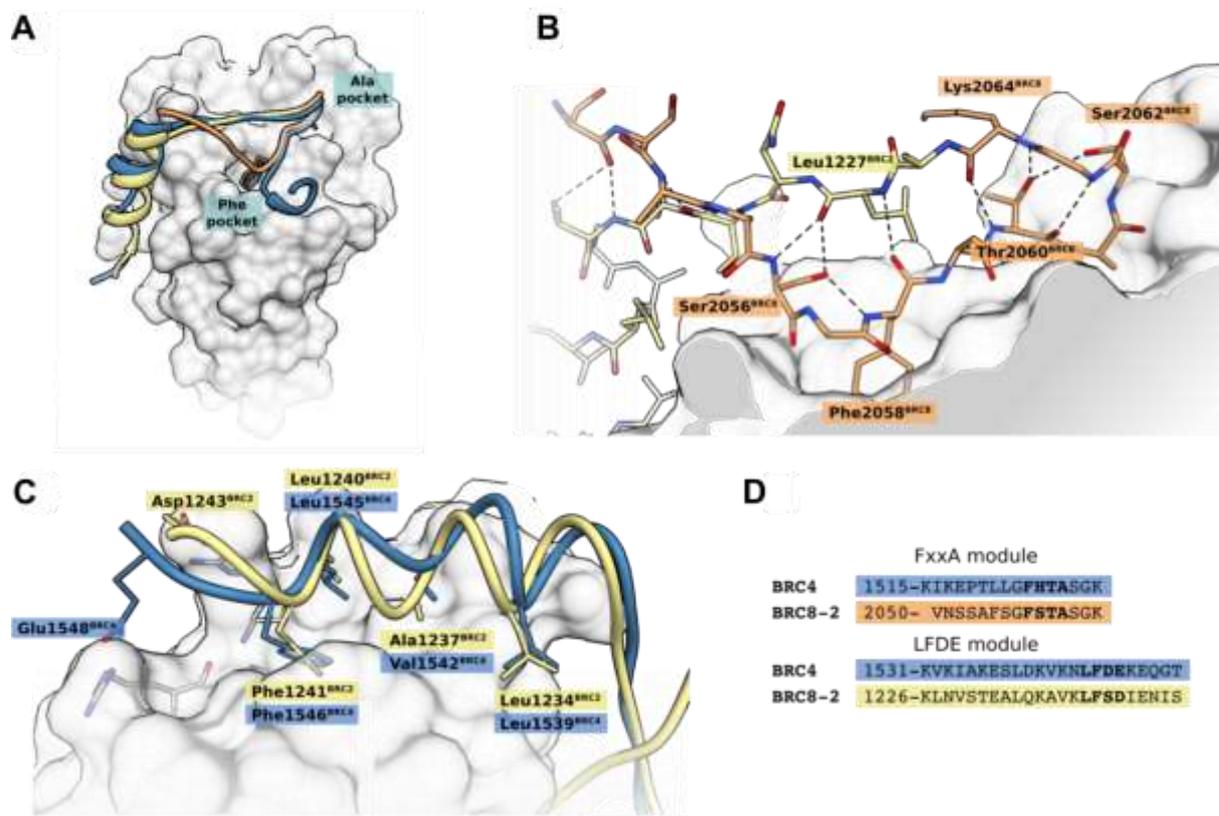
Figure 3. Crystal structure of the monomeric RAD51:BRC8-2 complex compared with RAD51:BRC4. BRC8-2 is depicted in orange and yellow, corresponding to BRC8 and BRC2 sequences, respectively. BRC4 is shown in blue. Peptides were superimposed by aligning the structures of their respective protein binding partners. Monomeric RAD51 is represented by a grey surface. Selected residues of the monomeric RAD51 are depicted in grey. (**A**) Overall topologies of the two peptides, with the Phe and Ala pockets of the FxxA site shown. (**B**) Hydrogen-bonding network of the BRC8-2 β-hairpin. (**C**) LFDE interface with side chains of crucial residues depicted. (**D**) Sequence alignment of BRC4 and BRC8-2 FxxA and LFDE modules.

FIGURE 4
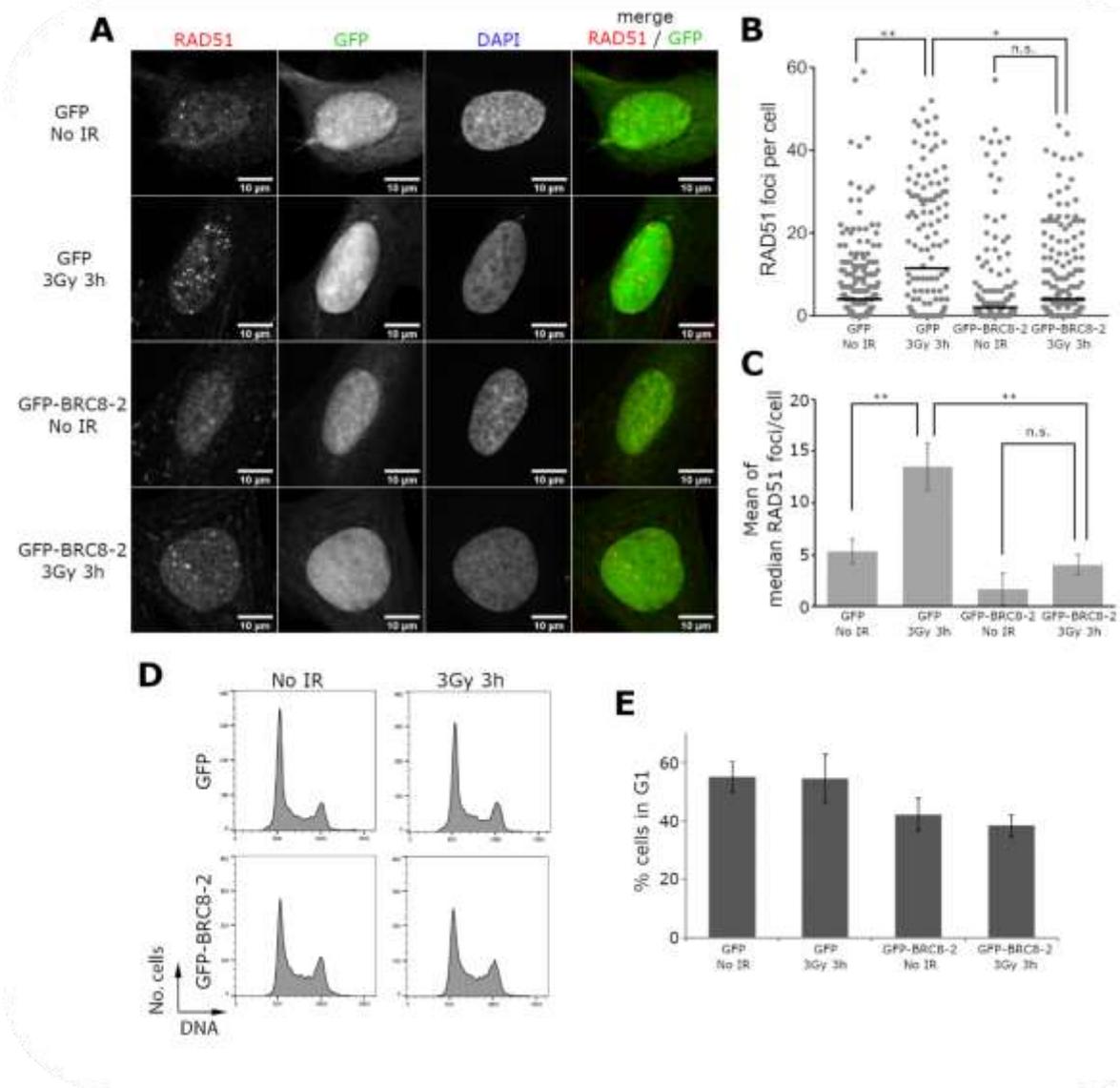
Figure 4. BRC8-2 impairs RAD51 foci formation in human U2OS cells. (**A**) Representative images of U2OS cells expressing empty GFP with a nuclear location signal (GFP) or GFP-BRC8-2 peptide (GFP-BRC8-2). Cells were monitored 3 hours after no treatment (No IR) or irradiation with 3 Gy (3Gy 3h) for GFP fluorescence or stained with RAD51 or DAPI as indicated. (**B**) Dot plot graph from one biological replicate plotting the number of RAD51 foci per GFP positive cell. Median values for each population are indicated with a bar. Outlier values were excluded from the graphical representation but included in the median calculation. More than 85 GFP positive cells were analyzed for each condition. Statistical analysis was done using Kruskal Wallis rank sum test followed by Dunn's procedure for pairwise comparison (*p <0.005, **p<0.001, n.s.= not significant). (**C**) Bar graph showing the average median RAD51 foci per GFP positive cell from three independent biological experiments. Data are presented as mean ± SEM, n = 3 biological repeats. **p<0.001, n.s.= not significant, using ANOVA test followed by Tukey's method. (**D**) Representative cell cycle profiles from GFP positive cells transfected with GFP or GFP-BRC8-2. Cells were analyzed by FACS 3 hours after no treatment (No IR) or irradiation with 3

Gy (3Gy 3h). (**E**) Bar graph showing the percentage of cells in G1 phase. Data are the mean values from three independent biological experiments ± SD.