ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE BIOTECNOLOGICHE, BIOCOMPUTAZIONALI, FARMACEUTICHE E FARMACOLOGICHE

Ciclo 38

Settore Concorsuale: 03/D1 - CHIMICA E TECNOLOGIE FARMACEUTICHE, TOSSICOLOGICHE E NUTRACEUTICO - ALIMENTARI

Settore Scientifico Disciplinare: CHIM/08 - CHIMICA FARMACEUTICA

RAD51 CONFORMATIONAL DYNAMICS EXPLORED THROUGH ADVANCED MOLECULAR SIMULATIONS

Presentata da: Veronica Bresciani

Coordinatore Dottorato

Maria Laura Bolognesi

Supervisore

Matteo Masetti

Co-supervisore

Mattia Bernetti

Esame finale anno 2026

Don't panic.

Table of Contents

Abstract

RAD51, a recombinase involved in DNA repair, is a promising drug target in the anticancer strategy known as Synthetic Lethality, a paradigm that leads a cancer cell to death by inhibiting two pathways simultaneously. RAD51 is found in fibril-like homo-oligomers; such structure is disassembled by the eight conserved BRC repeats of BRCA2, which recruit RAD51 monomers. RAD51 is then chaperoned to the damaged DNA site, where it catalyzes strand exchange during homologous recombination. The inhibition of the interaction between RAD51 and BRCA2 has proven to be synthetically lethal in association with PARP inhibition. Many drug discovery endeavors have been carried out to hinder the interaction between RAD51 and BRCA2, however, the details of such interface remain elusive because of the two proteins' high flexibility. To overcome this challenge, we combined experimental data from Small Angle X-ray Scattering and Crosslinking Mass Spectrometry with Molecular Dynamics simulations leveraging the Maximum Entropy principle to reconstruct the RAD51-BRC4 complex's conformational ensemble. Our results indicate that such complex assumes both compact and elongated conformations, and their interchange is guided by long range interaction between polar residues. We then approached the issue from a wider perspective, considering all eight RAD51-BRC complexes. We applied both end-point and alchemical perturbation based free energy estimation methods to identify residues that are possibly responsible for the different affinity of the BRC repeats for RAD51. Lastly, we tackled the modelling of a RAD51 filament assembly process. Our tool of election, given the availability of the filament's structure, was a Gō model, which allowed us to simulate the assembly in a coarse-grained setting, starting from RAD51 monomers.

List of Acronyms

C-ter	C-terminal domain
CV	Collective variable
DSB	Double strand break
HR	Homologous recombination
MAXENT	Maximum Entropy
MD	Molecular Dynamics
MetaD	Metadynamics
NHEJ	Non-homologous end joining
N-ter	N-terminal domain
PARP	Poly(ADP-ribose) polymerase
PARPi	Poly(ADP-ribose) polymerase inhibitor
Rg	Radius of gyration
SAXS	Small-angle X-ray scattering
SL	Synthetic Lethality
sMD	Steered Molecular Dynamics
SSB	Single strand break

1. Introduction

1.1 The Role of RAD51 in Homologous Recombination's Mechanisms

The recombinase RAD51 is among the proteins that have garnered attention by drug discovery groups over the years because of its implication in DNA repair. More specifically, RAD51 is involved in one of the two possible pathways of DNA double-strand break (DSB), namely homologous recombination (HR), as opposed to non-homologous end joining (NHEJ), in which there is no RAD51 involvement[1]. Moreover, while NHEJ happens during the entirety of the cell cycle, HR is restricted to S and G2 phases; the former is a rapid process which relies on DNA sequence to a lesser extent, resulting in joints with a lesser degree of homology, while the latter requires extensive homology between donor DNA and damaged DNA, but its results are virtually errorless. The two pathways only share their very first step, which entails the loading of the Ku70-Ku80 heterodimer on both ends of the DSB (**Figure 1**, step 1); after that the two processes diverge.

HR's second step entails the long-range resection of the damaged DNA, which begins with DNA nicking at the hands of the multifactorial end recognition complex (constituted by MRE11-RAD50-NBS1 and known as MRN). MRE11 detains both endonucleasic and 3'-5' exonucleasic activity; resection is supported by the phosphorylated cofactor CtIP which comes in contact with all three entities of the complex. First, the endonuclease nicks the DNA strand in proximity to its DSB 5' end (**Figure 1**, step 2); this process is followed by exonucleolytic degradation in the 3'-5' direction (**Figure 1**, step 3)[2].

The following step in DNA resection involves either DNA2, a protein with both nuclease and helicase activity, or the nuclease EXO1. Namely, DNA2 works together with helicases Bloom (BLM) or Werner (WRN), which unwind DNA by moving in 3'-5' direction and providing ssDNA to DNA2 (**Figure 1**, step 4a). The latter performs

resection through its nucleasic domain, while through its helicase domain translocates downstream of BLM/WRN. This process is supported by the Replication Protein A (RPA) binding to ssDNA. This final resection step involves different cofactors that support BLM/WRN and DNA2 activity, namely CtIP, which in turn interacts with BRCA1-BARD1, and MRN. On the other hand, if resection is carried out by EXO1 (**Figure 1**, step 4b), there are no helicases in play; however, RPA binding and BRCA1-BARD1 interaction promote the process[2].

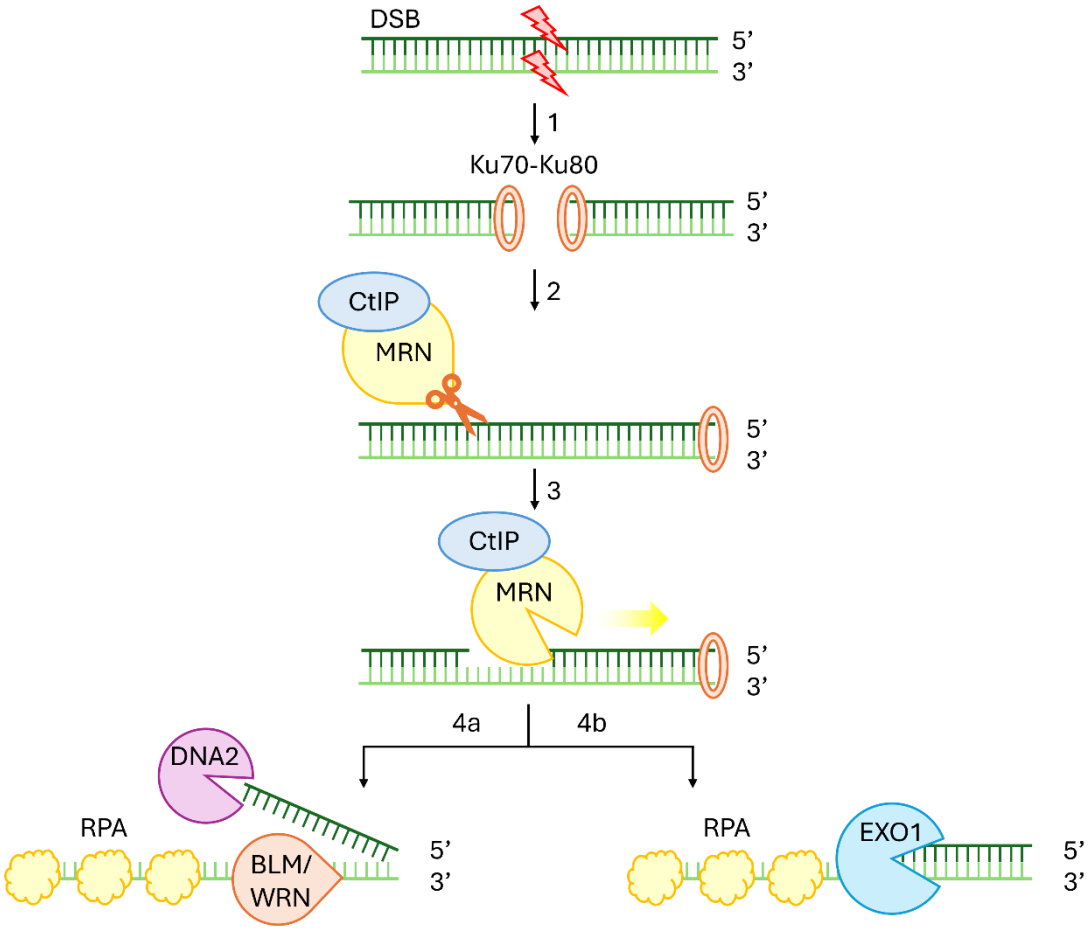

Figure 1. DNA long range end resection. Adapted from [1,2]

It is after this step that RAD51 comes into play. The ssDNA filament is inaccessible to RAD51 thanks to the binding with RPA; because of this, the latter is displaced from ssDNA by a recombination mediator, namely BRCA2. RAD51 is then allowed to come in contact with ssDNA, effectively forming a nucleoprotein filament

known as pre-synaptic complex and mediating strand exchange, homology search and promoting base pairing. After the ssDNA-RAD51 fibril invades a dsDNA donor strand, a three-stranded DNA intermediate (synaptic complex or synapsis) is formed; once enough homology between invading ssDNA and the complementary dsDNA strand is reached, the other, non-paired strand of the dsDNA is displaced. After RAD51 breaks contact with DNA, a DNA polymerase synthesizes the missing DNA starting at the 3' end of the invading strand using the invaded one as template[1,3].

1.2 RAD51 in Synthetic Lethality

Synthetic Lethality (SL) is a key concept for understanding RAD51's growing relevance as a pharmaceutical target. SL is defined as cell death, usually referring to a cancer cell specifically, which occurs when two genetic mutations are present simultaneously; however, both mutations must be present, otherwise one single genetic alteration is not per se sufficient to cause the lethal event[4]. The term can also be used to refer to the inhibition of two different proteins, usually the product of said genes: in this case, the more accurate term "fully small-molecule-induced SL" can be employed[5,6].

One of the most relevant examples of a synthetically lethal gene pairs consists of BRCA2 and poly(ADP-ribose) polymerase (PARP).[7] PARP is a family made up of several proteins, most of which catalyze the transfer of ADP-ribose to a given protein. More specifically, PARP1 and PARP2 are involved in single strand break (SSB) repair. However, were PARP to fail in its mechanism, a DSB would arise, and as such, HR would be the mechanism of choice for repairing DNA. Here is where BRCA2 comes into play, as a key factor in DSB repair through HR: were BRCA2 to fail too, the cell would no longer be viable (**Figure 2**). This is the reason behind the use and effectiveness of PARP-inhibitors (PARPi) such as olaparib, the first PARPi to be approved, in patients bearing a disruptive BRCA2 mutation[8].

However, not all patients bear the same, if any, BRCA mutations. Because of this, PARPi alone do not cause cell death in wild type BRCA2 individuals. As PARPi

hinder the effect of PARP, mimicking a disruptive mutation, the same can be done hindering BRCA2's function through modulators of its activity. Here is where RAD51 becomes a pharmaceutical target. In fact, RAD51 is itself a biomarker for synthetic lethality, as low expression of it indicate synthetically lethal effect of PARPi[9], while overexpression of RAD51 is linked to resistance to DNA damage and thus longer tumor survival[10]. This is supported by RAD51 being one of the primary partners of BRCA2 during HR. By this logic, were the RAD51-BRCA2 interaction to be disrupted, SL would be obtainable. This is the case of three inhibitors of the RAD51-BRCA2 interaction, namely ARN24089[11], CAM833[12] and ARN24992[13], which in combination with PARPi induce SL leveraging the synergic effect of PARP inhibition and RAD51-BRCA2 interaction disruption.

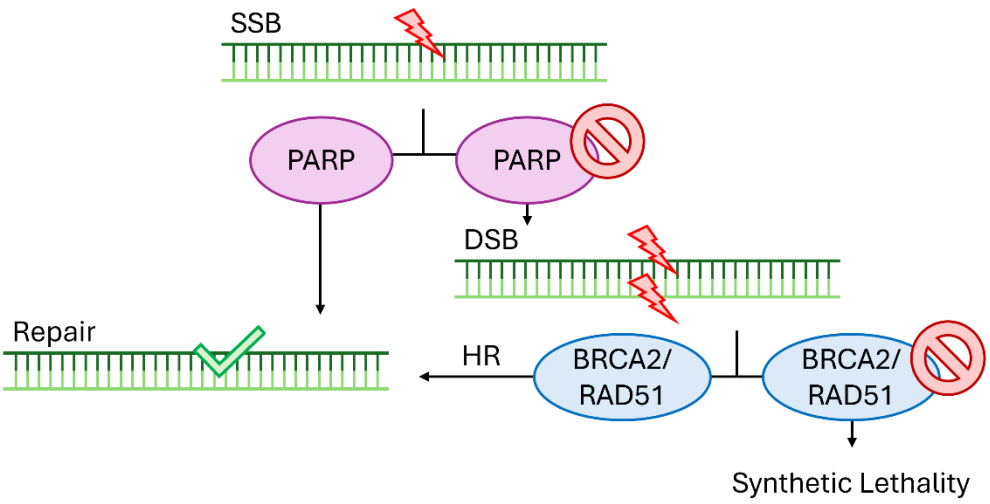

Figure 2. PARP and BRCA2 constitute a synthetically lethal pair. If PARP's pathway is inhibited, HR is responsible for DNA repair. If the BRCA2/RAD51 interaction is inhibited, the cell is no longer viable.

1.3 RAD51-BRCA2 Interaction

BRCA2 and RAD51 are essential partners in homologous recombination, yet their interaction remains poorly understood, primarily because of their intrinsic disorder, which complicates structural determination.

1.3 RAD51-BRCA2 Interaction

BRCA2 is a large, disordered protein, comprised of 3418 residues. During HR, it interacts with RAD51 through two different domains, with different functions. The first is a region containing eight repeat sequences with high homology, known as BRC repeats; they are involved in recruiting RAD51 to the DNA repair site. The second is BRCA2's C-terminal domain (C-ter), known as TR2, which protects the nucleoprotein filament from being disassembled.

RAD51 features a degree of intrinsic disorder, too. A smaller protein, compared to its partner BRCA2, it is made up of 339 residues and consists of two parts: the bigger, stabler, C-ter and the smaller, more flexible N-terminal domain (N-ter). The two are tethered by a flexible loop region.

C-ter is made up of around 100 residues, while N-ter consists of just over 200. Although the crystal structure of heptameric RAD51 has been obtained[14], the very beginning of its sequence, around 20 residues, could not be resolved. This is most likely due to such domain's high degree of intrinsic flexibility[15]. As a result, the available crystal structure for RAD51 in its filament form is missing 20 residues from its N-ter domain.

RAD51 arranges in fibril-like homo-oligomers; in the context of DNA repair, in order to reach the DSB site, RAD51 is recruited by BRCA2[16] (**Figure 3**), which makes contact with each monomer separately through the eight BRC repeats[17,18]. Once the DSB site is reached, the RAD51 monomers detach from BRCA2 and assemble on the ssDNA, giving origin to the nucleoprotein filament. At this stage, TR2 is dephosphorylated, allowing it to stabilize the formation of the synaptic complex, protecting it from disassembly by the BRC repeats[19–21]. After the complex has been formed, the strand exchange has been carried out and the post-synaptic complex has been disassembled, BRCA2's TR2 region is phosphorylated and TR2 detaches from the complex, allowing for RAD51 monomers to be once again recruited and chaperoned away from the repair site via the eight BRC repeats[22].

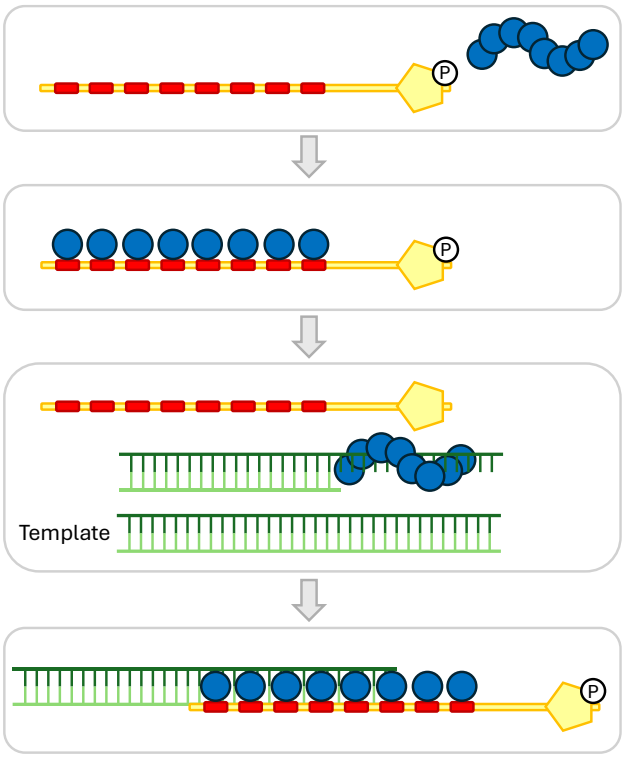

Figure 3. RAD51-BRCA2 interaction during HR. BRCA2 shown in yellow, BRC repeats in red, TR2 shown as a pentagon, RAD51 shown in blue.

The interaction between RAD51 and BRCA2 is still poorly understood. Part of the reason for this is BRCA2's intrinsic disorder. At present it has not yet been possible to obtain a full crystal structure of such a protein, which is also hefty in dimension. Partial crystals of the fourth BRC repeat (BRC4) in complex with monomeric RAD51[23] as well as the TR2 domain in complex with the RAD51-DNA nucleoprotein filament[24] are available on the protein data bank.

The available crystal structure for the RAD51-BRC4 complex is one of the few that detail RAD51-BRCA2 interaction (PDB code 1N0W[23]). The different repeats have different degrees of affinity for RAD51, and BRC4 is the one with the highest affinity of all. Such structure, however, only features RAD51's C-ter and BRC4, while it completely lacks information about N-ter.

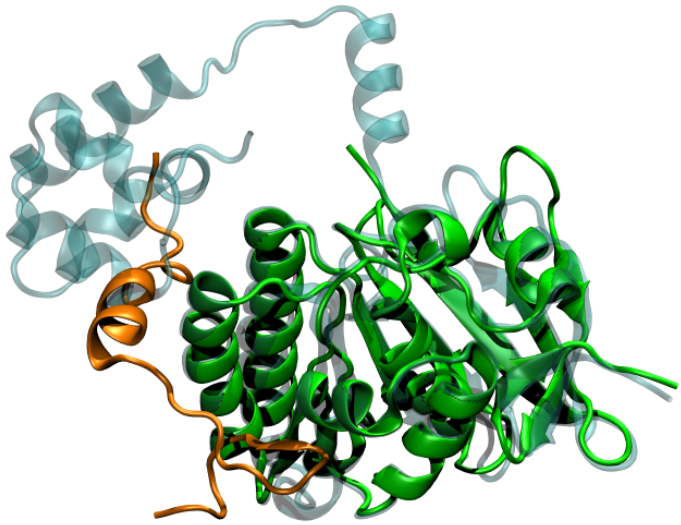

Figure 4. Superposition of RAD51's apo structure[14], shown in transparent cyan, and RAD51-BRC4's structure[23], with RAD51 C-ter shown in opaque green and BRC4 shown in opaque orange. Apo RAD51 N-ter would sterically clash with BRC4.

Moreover, by superposing a RAD51 monomer, taken from the deposited heptamer (PDB code 5NWL[14]), and the RAD51-BRC4 structure, a strong steric clash between RAD51 N-ter and BRC4 becomes evident. Indeed, the conformation of N-ter in RAD51 oligomer form is incompatible with the conformation of BRC4 in the complex: their respective positions largely overlap (**Figure 4**). Thus, N-ter needs to change is conformation to accommodate BRC4. It has been theorized that BRC4 binding strongly alters RAD51 N-ter's conformation, to the point of triggering a rearrangement so profound that the domain can be considered intrinsically disordered[15].

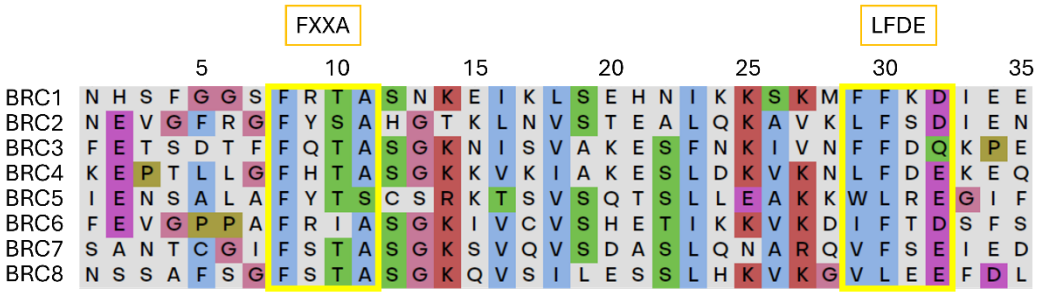

Figure 5. BRC repeats sequence alignment, showing the conserved FXXA and LFDE motifs.

At the same time, the structure of RAD51-BRC4, albeit incomplete, sheds light onto some key residues for RAD51-BRC4 interaction. Two motifs were identified and are relatively conserved in the eight repeats (**Figure 5**), which allow BRC4 to interact with RAD51's C-ter in a bidentate fashion. The more conserved motif FXXA engages a hydrophobic pocket which is also part of the self-assembly site of RAD51, the repeat then wraps around C-ter and the less conserved LFDE motif interacts with a second binding pocket (**Figure 6**).

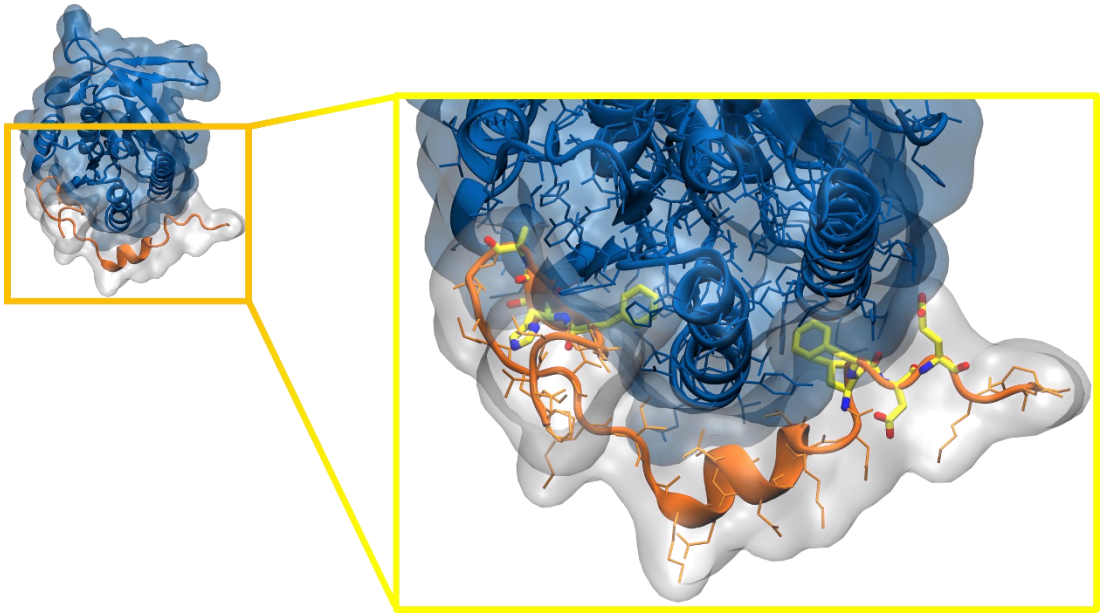

Figure 6. Crystal structure of BRC4 (in orange) bound to RAD51 C-ter (in blue). FXXA and LFDE motifs shown in yellow.

The affinity of the different BRC repeats has been experimentally found to be bipartite, with BRC1, -2, -3 and -4 being the ones with the highest affinity for RAD51, while BRC5, -6, -7 and -8 show lower affinity for RAD51's apo form. However, interestingly, the latter also show higher affinity for RAD51-ssDNA filament[25].

1.4 Aim of the Project

The project is articulated in three sections, more specifically one main project and two side projects. The first area that this work tackles is the interaction between

RAD51 and BRCA2, namely the RAD51-BRC repeats interface. The main project focuses on the RAD51-BRC4 complex, with the aim of characterizing its conformational ensemble making use of both computational tools and experimental data. Although several efforts have been made in towards drug discovery for SL centered on PARPi and RAD51-BRCA2 inhibition[11–13], they were only ever focused on the known BRC4 binding pockets, the FXXA and LFDE binding sites, both located on RAD51 C-ter. Given the lacking information regarding RAD51 N-ter, any other interactions involving BRC4 and RAD51 N-ter were difficult to model. This project stems from the need of a complete picture of the RAD51-BRC4 interaction, and detailing its conformational ensemble could help foster future drug discovery endeavors. Namely, we combined Molecular Dynamics (MD) and more specifically the enhanced sampling methods of Steered MD (sMD) and Metadynamics (MetaD) with experimental data obtained via Small-Angle X-ray Scattering (SAXS) and Crosslinking Mass Spectrometry (XL-MS). Tools specifically developed for in-silico modeling of SAXS data as well as the reweighting procedure following the Maximum Entropy (MAXENT) principle were employed to integrate experimental data with computational methods.

The first side project also insists on RAD51-BRCA2 interaction, but it focuses on the broader topic of RAD51-BRCs interfaces. Given that the only available structure for a RAD51-BRC complex is the aforementioned RAD51-BRC4 entry, little is known about the other seven repeats. Since experimental data shows that they differ in their affinity for RAD51, to the point of being split in two categories, we aimed at finding the structural reason behind these differences through computational methods, e.g. identifying key residues that could account for the gap in affinity. More specifically, we focused on relative binding free energy calculation using alchemical transformations to investigate the effect of individual residues on the binding free energy of each of the eight BRC repeats in complex with RAD51. This could be of interest for forthcoming drug discovery efforts, since all BRC repeats are thought to be involved in HR.

1.4 Aim of the Project

The second side project moves away from the RAD51-BRCA2 interaction and instead has its interest in RAD51 filament formation and its modeling via computational means. Since the crystal structure of a RAD51 filament is available, a Gō model was the method of choice for modelling RAD51 oligomer formation, given its nature as a structure-based approach for protein aggregation modeling. Although the RAD51 filament itself is currently not a pharmacological target, were the dynamics of its assembly to be computationally modeled, and hopefully in the future also the dynamics of RAD51-DNA nucleoprotein filament, they could provide helpful information for future pharmacological research.

2. Theory

2.1 Molecular Dynamics Simulations

MD is a popular technique for exploring a biological system's conformational space. Namely, MD mimics the physical evolution of said system, meaning how the atomic coordinates, the velocities at which its parts move, and the acceleration and force they are subjected to, change over time.

The starting point for any MD simulation in a set of coordinates and a set of velocities. Coordinates are stored in structure files that can be obtained experimentally or in silico; starting velocities for a system at a given temperature are generated via the Maxwell-Boltzmann distribution. The sampling engine behind MD is solving Newton's equation of motion (Equation 1), where $\boldsymbol{F}_i$ is the force acting on each atom i of the system. Here, a timestep dt must be chosen: it is an infinitesimal amount of time used to discretize time and split the integration in many, smaller steps. The timestep needs to have a value lower than 10 fs, since that is the C-H bond oscillation time, but usually does not need to be lower than 1 fs.

$$\frac{d^2 r_i(t)}{dt^2} = \frac{\boldsymbol{F}_i}{m_i} \tag{1}$$

Such force is calculated as the opposite of the gradient of the potential energy $\mathcal{V}$, as in Equation 2.

$$\boldsymbol{F}_i = -\frac{\partial \mathcal{V}(r_i, \ldots r_N)}{\partial r_i} \tag{2}$$

2.1.1 The Force Field

Force fields are sets of parameters and functions used to describe a system's potential energy. They are based on the concept of atom types, meaning the classification of each atom according to its element, hybridization state and bonding environment. Different atoms, even if they share the same element, will be assigned different sets of parameters according to their atom type. Such parameters define bonded energy terms, which describe the system's energy as a function of bond length and stretching, angle bending and torsion around bonds, as well as nonbonded terms, which correlate energy and interactions between atoms, meaning electrostatic and Van der Waals interactions. Equation 3 shows a generic force field function, with the different terms and the respective parameters.

$$\begin{aligned} \mathcal{V}(r^N) = & \sum_{\text{bonds}} \frac{1}{2} k_i \left(l_i - l_{i,0}\right)^2 \\ & + \sum_{\text{angles}} \frac{1}{2} \theta_i \left(\theta_i - \theta_{i,0}\right)^2 \\ & + \sum_{\text{dihedrals}} \frac{1}{2} V_n \left(1 + \cos(n\omega - \gamma)\right) \\ & + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \end{aligned} \tag{3}$$

Force field parameters are empirically derived sets of values whose accuracy strongly depends on the parametrization strategy, which often includes approximations and limited transferability. These limitations promoted optimization efforts over the years, aiming at improving force field accuracy. Moreover, the classical-mechanics nature of MD neglects quantum effects. Thus, electron or proton transfers, and consequently bond formation or breakage cannot be modelled[26,27].

2.1.2 Coarse-grained Representations

A core characteristic of a force field is the representation of a system that it employs. A force field in which to each physical atom corresponds one single virtual atom is called an all-atom force field. This type of force field strives to obtain a high accuracy in energy computation, but can result in long simulation time or hefty computational effort and need for computational resources in the case of longer or more complex simulations. For this reason coarse-grained approaches were developed. In a coarse-grained framework there is no one-to-one correspondence between atoms and their virtual representation. On the contrary, one coarse-grained entity, with its own set of parameters, corresponds to a multitude of physical atoms. In fact, such entities are called pseudoatoms or beads, given that they do not represent one atom, but two or more. Coarse-grained force fields and representations are useful for complex systems which would require extreme computational costs were they simulated using an all-atom force field. The reason for this is that the higher the number of atoms (or beads), the higher the number of interactions to be computed; by lowering the number of beads in the system, coarse-grained force fields allow for shorter simulation time and lower resource expense, at the cost of computational accuracy. Moreover, coarse-grained force fields are purposefully parametrized to allow for the use of much larger timesteps, i.e. 20 to 40 fs.

In this work we relied on different coarse-grained representation frameworks, as well as all-atom force fields, for different purposes. The MARTINI[28] coarse-grained force field is a popular choice. Originally developed for modeling lipids[29], its protein representation features residues being described by as few as one bead (e.g. alanine) up to five beads (e.g. tryptophane).

Single bead representations are available as well. As the name suggests, they model each amino acid using one single pseudoatom. In this work two of them were employed. The first is a Single Bead[30] representation tailored for modeling SAXS data, the details of which will be further discussed later.

2.1.3 Gō Models

The second single bead representation and force field that was used in this work is part of the Gō model family. A Gō model, named after its inventor[31], is a particular type of force field, characterized by being coarse-grained and structure based in nature. In the context of a structure based force field the energy minimum corresponds to the native state of the system, which needs to be available in order for the force field to be developed. This stems from the concept that, in such a force field, a negative (thus, favorable) contact interaction energy is assigned only to pairs of beads that are in contact in the system's native conformation, otherwise, no energy contribution is assigned. Originally, Gō models were developed for protein folding, but they can be used for multiprotein complex assembly as well.

The Gō model we relied on is called GoCa[32], whose name stems from the coarse graining paradigm it employs: indeed, it is a single bead representation, in which the bead representing a residue is placed in the location corresponding to the Cα atom (**Figure 7**).

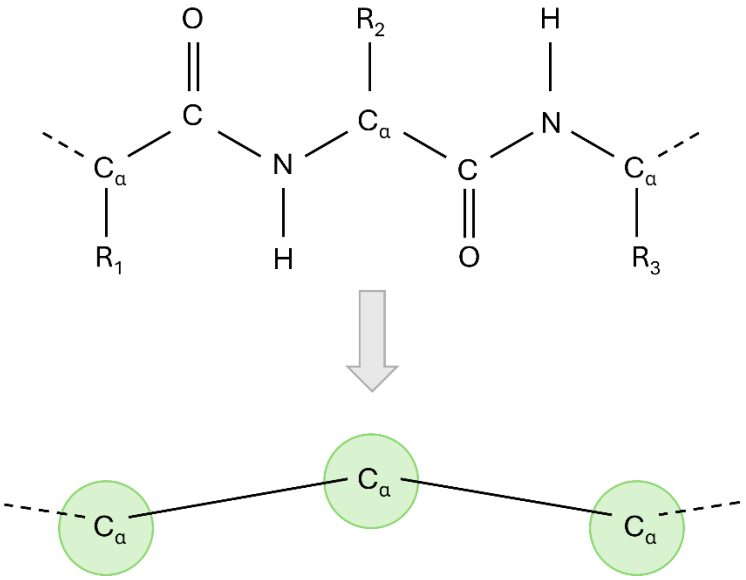

Figure 7. Schematic representation of the GoCa coarse graining approach, adapted from [32]. In a single bead coarse-grained force field each residue, independently from its size and number

of atoms, is described by one bead. In this particular case, the bead's coordinates match those of the Cα atom.

Equation 4 describes the GoCa forcefield, which comprises bonded interactions (harmonic bond length, harmonic bond angle as well as dihedral torsion) and nonbonded ones (12-6 Lennard-Jones potential). Electrostatic terms are purposefully left out of the picture: in fact, by virtue of being a structure-based force field, thus based on the native protein's structure, the force field implicitly takes into account the electrostatic contribution, since electrostatic effects affect the native conformation. However, since equilibrium values for distances and angles for each interaction is calculated as the distance or angle present in the native conformation, the force field will be valid only for the target structure it has been tailored on. A new force field must then be generated for each target that needs to be modeled.

$$
\begin{aligned}
V_{GoCa} = & \sum_{bonds} \frac{1}{2} k^{b} \left(r_{ij} - b_{ij}\right)^{2} \\
& + \sum_{angles} \frac{1}{2} k^{\theta} \left(\theta_{ijk} - \alpha_{ijk}\right)^{2} \\
& + \sum_{dihedrals} \sum_{n \in \{1,3\}} k_{n}^{\phi} \left(1 + cos\left(n\left(\phi_{ijkl} - d_{ijkl}\right)\right)\right) \\
& + \sum_{intra.\ native\ pairs} \epsilon_{inter} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \\
& + \sum_{non-native\ pairs} \epsilon_{intra} \left(\frac{\sigma_{R}}{r_{ij}}\right)^{12} \\
& + \sum_{inter.\ native\ pairs} \epsilon_{inter} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)
\end{aligned}
\tag{4}
$$

In GoCa, the sum over all bonds takes into account the beads (*i,j*) whose corresponding residues in the native structure are linked by a peptide bond (b_{ij} is the equilibrium distance value). In a similar fashion, the sum over all angles considers the

beads (*i,j,k*) that in their atomistic counterparts have a bond between the two pairs (*i,j* and *j,k*), while torsion terms consider the beads (*i,j,k,l*) that correspond in the all atom reference to amino acids that are pairwise connected by direct bonds (pairs are *i,j*, *j,k* and *k,l*). The equilibrium value d_{ijkl} is also calculated from the native conformation.

Nonbonded terms are divided in native and non-native contributions. Namely, a Van der Waals interaction-based cutoff is used to categorize amino acid pairs as either native or non-native: the pairs that, in the native structure, are closer than the cutoff are considered native and assigned an attractive potential, while non-native pairs, further away from each other than the cutoff, are assigned a repulsive contribution.

2.2 Enhanced Sampling Methods

Plain MD is often not enough on its own to fully explore the conformational landscape of a given system. Some events are difficult to sample because they are either rare or slow: they either happen on a timescale (e.g. folding events can take milliseconds, ligand recognition can require seconds) which is several orders of magnitude larger than the timescale that is currently possible to simulate via MD in reasonable amounts of time (e.g. microseconds), or they are difficult to sample because of high energy barriers. Enhanced sampling methods are born out of the necessity of sampling a system's conformational landscape in a shorter amount of time and without risking the sampling of only one set of low-energy configurations because of high energy barriers.

Enhanced sampling methods can be divided into two categories: Collective Variable (CV) based methods and non-CV-based methods. A CV is a reaction coordinate which describes a slow degree of freedom of the system that is used to expand the sampling of the configurational space. An example of non-CV-based approaches is parallel tempering (also called replica exchange)[33], while the methods used in this work fall under the CV-based umbrella, namely sMD[34] and MetaD[35].

2.2.1 Steered Molecular Dynamics

sMD was originally developed[34] to mimic the Atomic Force Microscopy experiments to study ligand binding[36], and consisted in the application of an external force, namely a harmonic potential, to restrain the ligand in a certain point in space and subsequently moving said point along a certain direction to allow for the sampling of different conformations at accessible timescales[37]. Since its development, its scope has broadened, and said harmonic potential can also be applied to other, arbitrary CVs.

Two scenarios are possible: Equation 5 describes an external potential given a single CV (x), where x_0 is the initial position of the restraint point moving at constant velocity (v) and K the restraint's stiffness.

$$U = \frac{K(x - x_0)^2}{2} \tag{5}$$

In this case, the external force applied to the system is described as in Equation 6, as a harmonic spring with constant stiffness.

$$F = K(x_0 + vt - x) \tag{6}$$

In the second scenario, the spring's stiffness is not constant, but increases with time ($K = \alpha t$); on the opposite hand, the point of restraint is fixed instead of moving. In this second case the force exerted on the system is described as in Equation 7.

$$F = \alpha t(x_0 - x) \tag{7}$$

2.2.2 Metadynamics

MetaD[38] is a CV-based enhanced sampling method developed with the purpose of exploring a system's configurational landscape without being stuck in a certain free energy minimum. As such, the method consists of an external, history-dependent biasing potential, a function of the CVs[35].

Such potential is applied on the system as a series of Gaussians, which allow the overcoming of high energy barriers by discouraging the re-sampling of configuration that have already been visited. **Figure 8** is a popular visual representation of the MetaD concept.

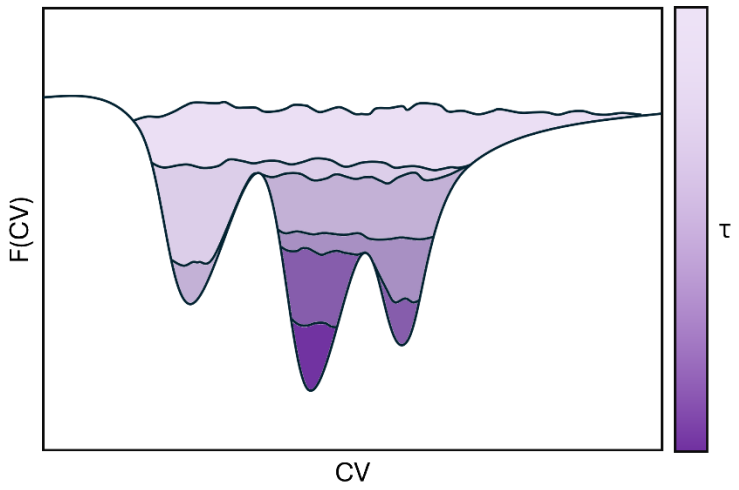

Figure 8. Visual representation of the MetaD process. The free energy surface with respect to a CV is shown. Gaussians are deposited over time τ (from dark to light purple), gradually fill energy wells and are able to overcome high energy barriers. The simulation starts in the central basin, where it would remain were the biasing potential non enforced. While the simulation time progresses, and as Gaussians are deposited, the basin on the right is filled. The last basin to be visited is the one on the left; after it is filled completely, the simulation progresses freely, akin to a random walk on the flattened free energy surface.

Given a set (S) of d functions of the coordinates R, the potential V_G corresponding to the ith CV at time t is described in Equation 8. This potential is the sum of Gaussians of width σ that are deposited along the trajectory in the CVs space.

$$V_G(S,t) = \int_0^t dt' \omega e^{-\sum_{i=1}^{d} \frac{\left(S_i(R) - S_i\left(R(t')\right)\right)^2}{2\sigma_i^2}}$$

(8)

Here, ω is the energy rate, described in Equation 9. It depends on the deposition stride of the Gaussians, τ_G, and the height of the Gaussians, W.

$$\omega = \frac{W}{\tau_G}$$

(9)

The free energy surface can then be reconstructed as in Equation 10.

$$V_G(S, t \to \infty) = -F(S) + C$$

(10)

A limitation of this implementation of MetaD is the assessment of its convergence: in fact, after the free energy wells have been visited, and more bias keep being deposited, the system is driven towards visiting high-energy conformations. This process can be seen as an instance of overfilling of the free energy basins. Ideally, a simulation should be stopped as soon as they are filled, but assessing how is not trivial.

Well-tempered MetaD was developed as a response to this problem[39]. In this framework, the height W of the Gaussians is not kept at a constant value, as in non-well-tempered MetaD, but it decreases over time according to Equation 11, where k_B is the Boltzmann constant and ΔT is a parameter with a temperature's dimension which represents the maximum temperature at which CVs are sampled.

$$W = \omega \tau_G e^{-\frac{V_G(S,t)}{k_B \Delta T}}$$

(11)

This entails that when the system visits a new free energy basin, the Gaussian height is restored to its original value, only to start decreasing again. This means that the potential converges in the long run, as shown in Equation 12.

$$V_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \tag{12}$$

However, Gaussian height decrement must be controlled in order to avoid not gathering enough bias in a given basin to overcome free energy barriers. This is done through a parameter called bias factor, γ in Equation 13.

$$\gamma = \frac{T + \Delta T}{T} \tag{13}$$

2.3 Integrating Experimental Data with Molecular Simulations

MD can be used in synergy with experimental data, integrating the latter into simulations to obtain results that comply with a given experimental observation. Two possible approaches can be envisaged: an ensemble of structures (in this case, from a MD trajectory) can be postprocessed through a reweighting procedure or it can be generated through experiment-biased simulation in an on-the-fly fashion[40].

When reweighting is performed, the original weights assigned to each state of an initial ensemble, called prior, are recalculated so that the weighted-averaged of a given observable over the ensemble matches an experimentally-measured reference value. The resulting ensemble, generated via such reweighting procedure, is called posterior. Since different distributions can result in the same average value, different approaches can be used[40,41]. Strategies based on the MAXENT principle aim at maximizing the amount of information from the original ensemble by introducing the

minimal perturbation that ensures agreement with the experimental data. Conversely, Maximum Parsimony approaches generate a posterior featuring the least amount of states that fit experimental data. A third paradigm, called Bayesian Inference or MaxPrior, can be combined with wither MAXENT or Maximum Parsimony by leveraging Bayesian statistics[40,42].

Agreement with experimental data can also be enforced on-the-fly during the simulations. To achieve this, an additional bias is added to the force field, and for this purpose several approaches can be employed. For instance, experimental restraints can be introduced as penalty terms, either applied iteratively or through a multi-replica approach. MAXENT and Bayesian Inference (in this context called Metainference[43]) can also be employed on-the-fly.

Both ensemble reweighting and on-the-fly enforcement have advantages and drawbacks. A reweighting protocol starts from an ensemble and generates a new, experimental data-compliant one; if new experimental information were to emerge, a new reweighting instance could be carried out on the same original ensemble without the need of performing the simulations again. This is not the case for experiment-biased simulations, whose protocol is inherently specific to the experimental data: because of this, new data requires a new simulation, making it a less flexible approach.

However, reweighting approaches strongly depend on the quality of the initial ensemble, as a poor prior ensemble will not make this procedure successful. In practice, this is typically the case when the sampling strategy is unable to explore states compatible with the experimental ones. This can be the case of poor force fields, which tend to sample unrealistic states, or of states that are particularly challenging to sample, even via advanced enhanced-sampling strategies. Conversely, with on-the-fly enforcement the system adapts to comply with the reference experimental data during the simulation, directly generating a compatible ensemble. However, it should be also considered that frequently computing the target observables during the simulations, i.e. repeatedly applying the forward models to enforce experimental restraints on the fly,

may degrade simulation performance, to an extent that depends on the computational cost of the forward model used.

Finally, we note that the calculation of experimental observables from structures sampled by MD simulations presents intrinsic limitations, since forward models are, by definition, based on approximations, and that experimental measurements are also subject to errors.

2.3.1 SAXS Spectra Modeling

SAXS is a popular experimental technique used for gaining non-atomistic insights regarding the conformation of a system, namely its overall size and shape. Despite its limited structural resolution, SAXS is a valuable tool, particularly when higher-resolution approaches, e.g. X-ray crystallography or cryo-EM, are difficult to apply. In addition, compared to these other approaches, SAXS features a relatively simpler setup, is more robust against experimental conditions (e.g. ionic strength), and, most importantly, measurements are performed directly in solution, thus providing information on biological systems under near physiological conditions. Moreover, the method does not impose a size limit on the system being analyzed[44].

The protocol for conducting a SAXS analysis is schematized as in **Figure 9**. To account for the solvent contributions, the measurement is conducted twice, once on the analyte's solution, and once on the buffer alone. Since the measurement is performed directly in solution at a defined analyte concentration, the signal obtained by the detector is the result of the contribution of the possibly multiple configurations adopted by the biomolecular system. The resulting scattering pattern therefore represents an average of many scattering instances generated by each possible configuration of the analyte, and the resulting SAXS spectrum provides ensemble-averaged structural information.

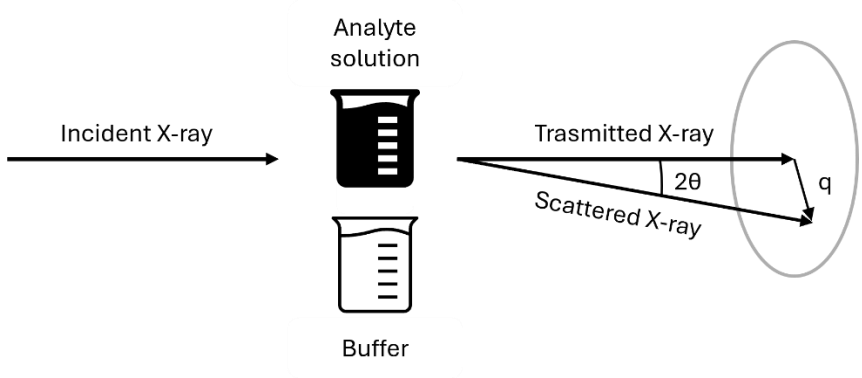

Figure 9. Schematic representation of a SAXS experiment. The process is carried out both on the solution and on the buffer.

A SAXS experiment consists of a monochromatic X-ray beam incident on the sample, which generates a transmitted and a scattered beam. The latter is imaged by a 2D detector. The scattering vector (or momentum transfer, q) indicates the change in direction, with respect to the incident beam, of the scattered one[44]. q is defined according to Equation 14, where λ is the wavelength of the incident beam and 2θ is the scattering angle.

$$q = 4\pi \frac{sin\theta}{\lambda} \tag{14}$$

Since the orientation of the biomolecules in solution is random, the recorded scattering intensity (I) can be radially averaged. It is then plotted against q to obtain a one-dimensional scattering profile. To highlight some features of this spectrum, particularly useful for protein analysis, $q^2I(q)$ can be plotted against q: the resulting representation is called a Kratky plot (**Figure 10B**), which gives qualitative information about the overall shape of the analyte. A bell-shaped curve with a well-defined maximum is a characteristic of a more compact, globular protein, whereas a curve lacking such a peak and displaying a plateau at higher q values indicates an unfolded or flexible protein[45].

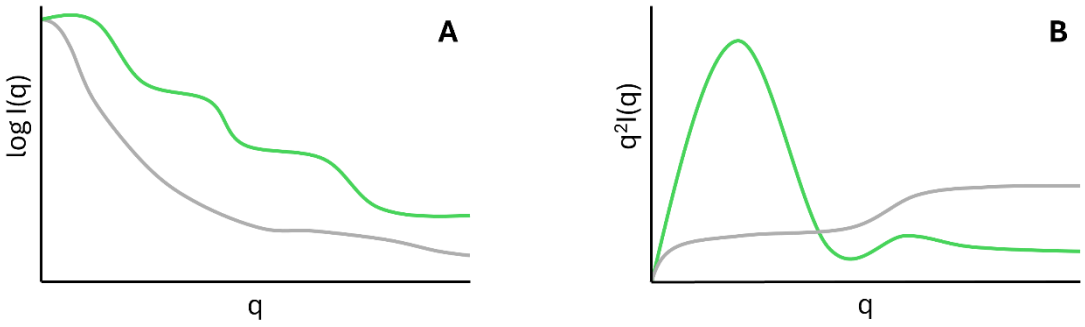

Figure 10. Pictorial representation of SAXS spectra of a globular (shown in green) and an unfolded protein (shown in grey). A) Logarithmic plot. B) Kratky plot.

The Radius of Gyration (Rg) of a biomolecule, which quantifies its degree of structural compactness, can be obtained via the SAXS spectrum. This is achieved via the Guinier approximation[46], according to which for low values of q ($q < 1/Rg$) the intensity I depends on two parameters only, as shown in Equation 15.

$$I(q) = I(0)e^{-\frac{q^2 {R_g}^2}{3}} \tag{15}$$

From a computational standpoint the SAXS spectrum of a given three-dimensional structure can be calculated according to Equation 16:

$$I(q) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(q) f_j(q) \frac{sin(qr_{ij})}{qr_{ij}} \tag{16}$$

where the scattered intensity I at a certain value of the momentum transfer q is obtained from the pairwise distance r_{ij} between all pairs of the N atoms in the biomolecules and the corresponding atomic scattering factors $f_i(q)$ and $f_j(q)$. Scattering factors are computed starting from experimentally available atom-type specific parameters (a_k, b_k, c) as shown in Equation 17.

$$f_i(q) = f_i^{atomic}(q) = \sum_{k=1}^{4} a_k e^{-b_k\left(\frac{q}{4\pi}\right)^2} + c \tag{17}$$

However, the influence of the solvent on the system cannot be overlooked. It can be either included through explicit solvent models, that include the solvent molecules that surround the protein in the calculation of pairwise distances, or modeled through implicit solvent approaches. The latter expedite the process by adding modifications to the atom's form factors, without computing distances for solvent molecules.

One first additional parameter accounts for the effect of the displaced solvent: it depends, for an atom *i*, on the volume of solvent displaced by it (v_i) and the solvent's electron density (ρ_0), as shown in Equation 18.

$$f'_i(q) = f_i^{atomic}(q) - \rho_0 f_i^{solvent}(q) = f_i(q) - v_i \rho_0 e^{-\frac{q^2 v_i^{\frac{2}{3}}}{4\pi}} \tag{18}$$

Another parameter that needs to be taken care of is the effect of the solvation shell: in an implicit solvent framework a third term is added to the scattering factor.

Calculating the pairwise distance between each pair of atoms (Equation 16) in a macromolecular system at atomistic resolution is computationally demanding. This becomes particularly critical if the calculation has to be repeated many times, e.g. on a large structural ensemble or on-the-fly during MD simulations. For this reason, a coarse-grained representation can improve the efficiency of the calculation. This results in a system representation comprising *M* beads, with typically $M \ll N$, and the computed scattering intensity can be expressed as Equation 19.

$$I(q) = \sum_{i=1}^{M}\sum_{j=1}^{M} F_i(q)F_j(q)\frac{sin(qR_{ij})}{qR_{ij}}$$

(19)

Here, R_{ij} is the distance between pairs of beads. Notably, this implies that scattering factors F associated with the pseudoatoms are available, i.e. have been purposely parameterized. Most interestingly, a great advantage of this strategy is that it can be exploited in a hybrid coarse-grain/all-atom fashion. Specifically, the MD simulations can be conducted at a fully atomistic resolution, while resorting to a coarse-grained representation for the sole purpose of computing efficiently the SAXS spectra (hySAXS scheme).

The MARTINI representation is a popular choice for this purpose. Scattering factors of MARTINI beads have been parameterized for both protein[47] and nucleic acid[48] systems, allowing for hySAXS scheme simulations exploiting MARTINI as the coarse-grained model (MT-hySAXS).

$$F_j(q) = \left[\sum_{i\in j}\sum_{k\in j} f'_i(q)f'_k(q)\frac{sin(qr_{ik})}{qr_{ik}}\right]^{\frac{1}{2}}$$

(20)

Equation 20 describes the MARTINI scattering factors for proteins: the scattering factor for a bead (j) depends on the factors of the atoms (i, k) that are part of said bead.

Recently, an alternative coarse-grained model for SAXS spectra calculation was introduced, named Single-Bead[30,49]. In the protein context, this representation replaces each amino acid with a single bead, hence the name. Despite being more simplified than the MARTINI representation, thus further improving the efficiency of SAXS spectra calculation, this model retains comparable accuracy, particularly for q values lower than 0.3 A^{-1} [30]. The Single-Bead approach (SB-hySAXS) features a

solvent layer contribution term[30] in the definition of scattering factor F that allows accounting for the effect of solvation on solvent-exposed atoms in the biomolecule.

$$
\begin{aligned}
F_i(q) = \left[F_i^{atomic} + \rho^2 F_i^{solvent} - \rho F_i^{mixed}\right]^{\frac{1}{2}} \\
= \left[\sum_{k\in i}\sum_{l\in i} f_k^{atomic}(q) f_l^{atomic}(q) \frac{sin(qr_{kl})}{qr_{kl}}\right. \\
+ \rho^2 \sum_{k\in i}\sum_{l\in i} f_k^{solvent}(q) f_l^{solvent}(q) \frac{sin(qr_{kl})}{qr_{kl}} \\
- \rho \sum_{k\in i}\sum_{l\in i} \left[f_k^{atomic}(q) f_l^{solvent}(q)\right. \\
\left.\left. + f_k^{solvent}(q) f_l^{atomic}(q)\right] \frac{sin(qr_{kl})}{qr_{kl}}\right]^{\frac{1}{2}}
\end{aligned}
\tag{21}
$$

Equation 21 details the scattering factors of the Single Bead[30] framework as the sum of three terms. The method is particularly flexible, as it allows for the customization of the ρ parameter to both account for buffering agents different from water and to permit the use of modified solvation densities to different beads to mimic the solvation layer's effect according to $\rho = \rho_0 - SLC$, where SLC is the solvent layer contribution parameter.

A parameter that is typically employed to assess overfitting of a predicted SAXS spectrum with respect to an experimental one is the reduced χ^2:

$$
\chi^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\frac{I_{fit}(q_i) - I_{exp}(q_i)}{SE\big(I(q_i)\big)}\right)^2
\tag{22}
$$

where $I_{fit}(q_i)$ and $I_{exp}(q_i)$ are the predicted and experimental intensities, and $SE(I(q_i))$ the standard error on the intensities at each of the m values q_i of the SAXS spectrum. χ^2 values close to 0 indicate overfitting, while values around 1 indicate

optimal fitting of the experimental data. In this work, $SE(I(q_i))$ included both the experimental and predicted uncertainties after applying error propagation.

2.3.2 Maximum Entropy Principle

The MAXENT principle is a theoretical framework that can be used to integrate experimental data with *in silico* molecular simulations. In particular, the approach can be applied to either bias an MD simulation or to reweight an already performed simulation. This is done in the spirit of ensuring compliance with experiments while enforcing the least possible alteration on the ensemble's original distribution, the prior.

According to the principle, for a system consisting in a set of states, the optimal distribution of said states that is compatible with observed data is the one that maximizes its Shannon's entropy, later expanded to relative entropy[50]. This distribution, obtained after the input of experimental knowledge, is the posterior.

Given a set of continuous variables $\boldsymbol{x}$, entropy is defined, for a distribution P and relative to a prior distribution (P_0), as described in Equation 23.

$$S[P \parallel P_0] = -\int d\boldsymbol{x} P(\boldsymbol{x}) ln \frac{P(\boldsymbol{x})}{P_0(\boldsymbol{x})} \tag{23}$$

Equation 24 shows that the posterior ($P_{ME}(\boldsymbol{x})$) is obtained by maximizing the entropy (S) (first line), while the ensemble average of M observables s_i calculated on the distribution $P(\boldsymbol{x})$ is constrained to match the experimental values s_i^{exp} (second line) and the distribution is normalized (third line).

$$\begin{cases} P_{ME}(\boldsymbol{x}) = arg \max_{P(\boldsymbol{x})} S[P \parallel P_0] \\ \int dx s_i(\boldsymbol{x}) P(\boldsymbol{x}) = \langle s_i(\boldsymbol{x}) \rangle = s_i^{exp};\ i = 1, \ldots, M \\ \int d\boldsymbol{x} P(\boldsymbol{x}) = 1 \end{cases} \tag{24}$$

This process generates a posterior in agreement with experimental data which is as similar as possible to the prior: this is due to the fact that the Kullback-Leibler divergence ($D_{KL}[P \parallel P_0]$), which describes the information that is gained by substituting $P_0(\boldsymbol{x})$ with $P(\boldsymbol{x})$, is equivalent to the negative of the relative entropy ($S[P \parallel P_0]$).

The entropy maximization problem is solved by identifying stationary points of the Lagrange function as shown in Equation 25, where λ_i and μ are Lagrangian multipliers.

$$\mathcal{L} = S[P \parallel P_0] - \sum_{i=1}^{M} \lambda_i \left(\int d\boldsymbol{x} s_i(\boldsymbol{x}) P(\boldsymbol{x}) - s_i^{exp} \right) - \mu \left(\int d\boldsymbol{x} P(\boldsymbol{x}) - 1 \right)$$

(25)

Equation 26 shows the derivative of $\mathcal{L}$ with respect to $P(\boldsymbol{x})$.

$$\frac{\partial \mathcal{L}}{\partial P(x)} = -ln \frac{P(\boldsymbol{x})}{P_0(\boldsymbol{x})} - 1 - \sum_{i=1}^{M} \lambda_i s_i(\boldsymbol{x}) - \mu$$

(26)

By having such derivative equal zero, $\frac{\partial \mathcal{L}}{\partial P(x)} = 0$, the posterior is described by Equation 27.

$$P_{ME}(\boldsymbol{x}) \propto e^{-\sum_{i=1}^{M} \lambda_i s_i(x)} P_0(\boldsymbol{x})$$

(27)

In the context of a reweighting procedure on an already available MD trajectory comprising N_s structures, the prior ensemble, new weights are assigned to each configuration (x_l). First, choosing the values of the Lagrangian multipliers λ requires minimizing the $\Gamma(\lambda)$ function in Equation 28.

$$\Gamma(\lambda) = ln\left(\frac{1}{N_s}\sum_{t=1}^{Ns} e^{-\sum_{i=1}^{M}\lambda_i s_i(x_t)}\right) + \sum_{i=1}^{M}\lambda \cdot s^{exp} + \Gamma_{err}(\lambda)$$

(28)

The last term in the equation is introduced to model experimental error, acting as a regularization term to reduce over-fitting, and can be defined as shown in Equation 29, where a different σ_i value can be specified for each observable s_i.

$$\Gamma_{err}(\lambda) = \frac{1}{2}\sum_i \lambda_i^2 \sigma_i^2$$

(29)

Once the Lagrangian multipliers are selected, new weights (w_t) are computed and assigned to each configuration (x_t) as in Equation 30.

$$w_t = e^{-\sum_i \lambda_i^* \cdot s_i(x_t)}$$

(30)

Weights are then normalized so that $\sum_{t=1}^{N_S} w_t = 1$.

In this work, the ensemble to be reweighed was generated via MetaD, thus it featured uneven weights for each conformation, associated with MetaD bias. Equation 28 and 30 can then be rewritten as Equation 31 and 32, where $V_b(x_t)$ is the MetaD bias at the end of the simulation recomputed for the coordinates of the x_t^{th} frame, and k_BT is the thermal energy.

$$\Gamma(\lambda) = ln\left(\frac{1}{N_s}\sum_{t=1}^{Ns} e^{-\sum_i \lambda_i^* \cdot s_i(x_t) + \frac{V_b(x_t)}{k_B T}}\right) + \lambda \cdot s^{exp} + \Gamma_{err}(\lambda)$$

(31)

$$w_t = e^{-\sum_i \lambda_i^* \cdot s_i(x_t) + \frac{V_b(x_t)}{k_B T}}$$

(32)

Typically, only a fraction of the structures from the prior ensemble contribute effectively to the reweighted ensemble. The number of such structures, that carry significant weight, can be approximately estimated by calculating the Kish effective sample size, shown in Equation 33.

$$K = \frac{\left(\sum_{t=1}^{N_s} w_t\right)^2}{\sum_{t=1}^{N_s} w_t^2}$$

(33)

2.4 Free Energy Estimation

The interaction between two proteins L (for the word ligand) and R (for receptor) is described by Equation 34.

$$L + R \rightleftharpoons RL$$

(34)

Their binding constant, K_D, is related to the binding free energy ΔG^0_{bind} of the complex (i.e. the difference between the free energy of the bound and unbound state). The binding free energy can be computed as a function of the partition functions *Z* of the bound and unbound state, as shown in Equation 35, where *R* is the gas constant and *T* the temperature[51]. Computing a system's ΔG^0_{bind} is a process known as absolute binding free energy estimation.

$$\Delta G^0_{bind} = RTln(K_D) = -RTln\left(\frac{Z_{bound}}{Z_{unbound}}\right)$$

(35)

The difference in binding affinity between congeneric ligands with respect to the same receptor is their $\Delta\Delta G^0_{bind}$; its computation is known as relative binding free energy estimation.

Different computational paradigms are available to estimate absolute and relative binding free energy, namely end point methods, alchemical perturbation methods and physical pathway methods. The latter are not discussed in this work, as they were not employed.

2.4.1 MM-PB(GB)SA

End point methods are free energy estimation techniques based on the sampling of the initial and final states of the binding process, unbound ligand and receptor, and ligand-receptor complex. Although they can be carried out on single structures, they are usually applied to MD trajectories.

Namely, in the MM-PB(GB)SA framework, the free energy of a molecule is described by Equation 36, where E_{MM} is its gas phase energy, $G^{PB}_{solv} + G^{SA}_{solv}$ is its solvation free energy and TS is the solvent's configurational entropy[52].

$$G = E_{MM} + G^{PB}_{solv} + G^{SA}_{solv} - TS \tag{36}$$

The gas phase energy E_{MM} is computed via a molecular mechanics force field as shown in Equation 3. The solvation energy is split in a polar and a nonpolar term, modelled through implicit solvent methods. In the original version of the method, named MM-PBSA, the polar term G^{PB}_{solv} is computed, for the transfer of a molecule from gas phase to the solvent, via the Poisson Boltzmann equation (Equation 37), where k is the Boltzmann constant, T the temperature and the ε, φ, ρ, and κ are functions of the position vector $\boldsymbol{r}$. Namely, ε is the dielectric constant, φ is the electrostatic potential in units of $\frac{kT}{q}$, where q is the proton's charge and ρ^f is the fixed charge density

in units of q. The term κ, the Debye-Hückel screening parameter, is the reciprocal of the Debye length λ, where $\lambda = \sqrt{\frac{\varepsilon_0 kT}{8\pi q^2 I}}$, where I is the bulk solution's ionic strength[53].

$$\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla \cdot \phi(\boldsymbol{r})] - \varepsilon(\boldsymbol{r})\kappa(\boldsymbol{r})^2 \sinh[\phi(\boldsymbol{r})] + \frac{4\pi\rho^f}{kT} = 0$$

(37)

In the more recent MM-GBSA version of the technique, the polar term G_{solv}^{GB} is computed from an extension[54] of the Generalized Born equation[55] (Equation 38), where $f^{GB} = \sqrt{{r_{ij}}^2 + {\alpha_{ij}}^2 e^{-D}}$. Here r_{ij} is the distance between atoms i and j, $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$, where α_i and α_j are atomic radii, and $D = \frac{{r_{ij}}^2}{\left(2\alpha_{ij}\right)^2}$

$$G_{solv}^{GB} = -166\left(1 - \frac{e^{-\kappa f^{GB}}}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f^{GB}}$$

(38)

The nonpolar term, G_{solv}^{SA}, is a sum of the favorable interactions between solute and solvent (ΔG_{att}) and unfavorable energy required for cavity formation[56] (ΔG_{rep}), in Equation 39. Here, the unfavorable energy depends on the molecule's solvent accessible surface area ($\Delta G_{rep} = \gamma SAS + c$), where γ is the surface tension coefficient. On the other hand, the favorable energy can be approximated by the Van der Waals attractive interaction potential energy as in $\Delta G_{att} \approx U_{att}^{uv} = \sum_{a=1}^{N_s}\int \rho_{aw}(\boldsymbol{r}_{aw})V_{att}(\boldsymbol{r}_{aw})d\boldsymbol{r}_{aw}$, where N_s is the number of atoms involved, r_{aw} is the solvent-solute distance, ρ is a solvent distribution function and V_{att} is the Van dee Waals potential[56].

$$G_{solv}^{SA} = \Delta G_{rep} + \Delta G_{att}$$

(39)

Lastly, the entropic term is the most difficult to compute, to the point of being often ignored.

To compute a complex's binding free energy a thermodynamic cycle is considered (**Figure 11**) A thermodynamic cycle is a set of reactions that starts and ends in the same state: by virtue of being a state function (meaning that its change only depends on the initial and final state) the total change in free energy considering the whole cycle is null. Thus, in the depicted cycle describing solvation and binding of receptor and ligand, the free energy of binding is shown in Equation 40 as the sum of energy and configurational entropy associated with complex formation and the difference between complex solvation and solvation of standalone receptor and ligand[52].

$$\Delta G_{bind} = E_{MM} + \left(\Delta G_{solvation}^{complex} - \Delta G_{solvation}^{ligand} - \Delta G_{solvation}^{receptor}\right) - T\Delta S \quad (40)$$

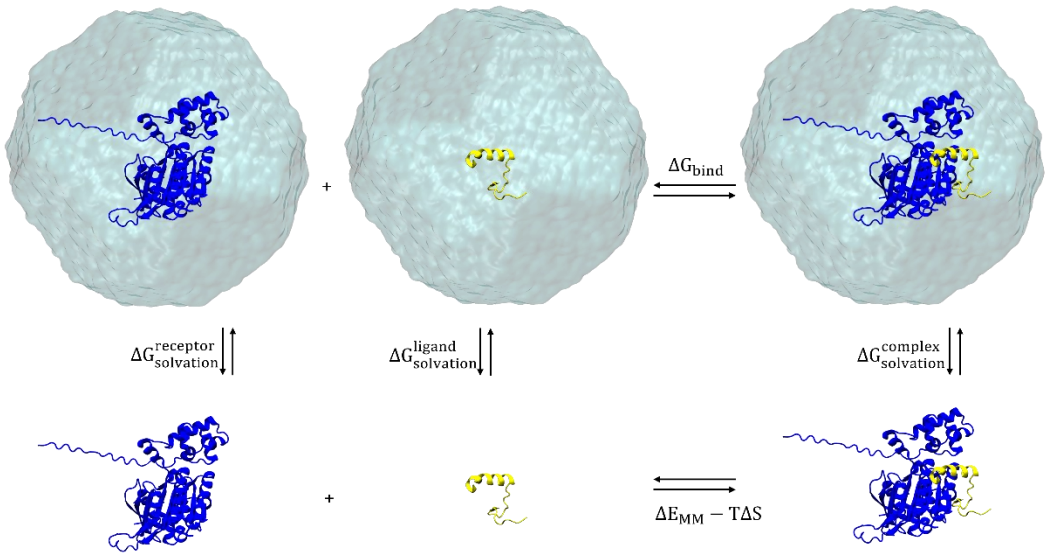

Figure 11. Schematic representation of a MM-PB(GB)SA thermodynamic cycle.

Two approaches are available to carry out these calculations, namely the single trajectory and three trajectories approach. In the single trajectory version, only the

complex is simulated and parameters relative to ligand and receptor alone are computed by ignoring the other partner. In the three trajectories version, three simulations are carried out, one for the complex and one for each of its components. The single trajectory approach is usually employed for systems where great conformational change upon binding is not expected[52].

2.1.1.1 Computational alanine scanning

Computational alanine scanning is a technique used to assess the contribution of a particular amino acidic residue to the overall binding free energy of a protein-protein complex[57]. More specifically, given a protein consisting of N residues, N mutants are generated by substituting, separately, an alanine to each residue (**Figure 12**).

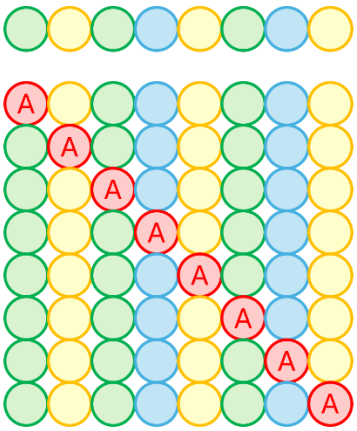

Figure 12. Pictorial representation of the alanine scanning setup. Original residues shown in blue, green and yellow, alanine shown in red.

From a practical point of view, this is identical to truncating the residue's side chain at C_γ and substituting C_γ with a H. After that, the binding free energy of each system is estimated. The difference in binding free energy between the wild type protein and each mutant depends on the contribution of the original residue's side chain, with respect to alanine's methyl side chain.

2.4.2 Alchemical Perturbations

Alchemical perturbation based methods, albeit more computationally expensive than the MM-PB(GB)SA approach, strive to obtain either absolute or relative binding free energies for a complex or a congeneric series of ligands with respect to the same receptor without the need to explicitly sample the binding process. Here, relative binding free energy estimation through alchemical perturbation methods will be described.

This class of methods relies on a thermodynamic cycle as well. Such cycle (**Figure 13**) features the binding event of the two ligands in question to the same receptor, both physical reactions, and two non-physical reactions, namely the transformation of each ligand into the other, in bound and unbound state. These transformations, albeit non-physical, are needed to construct the cycle and to be able to more efficiently compute binding free energies without explicitly simulating the binding event.

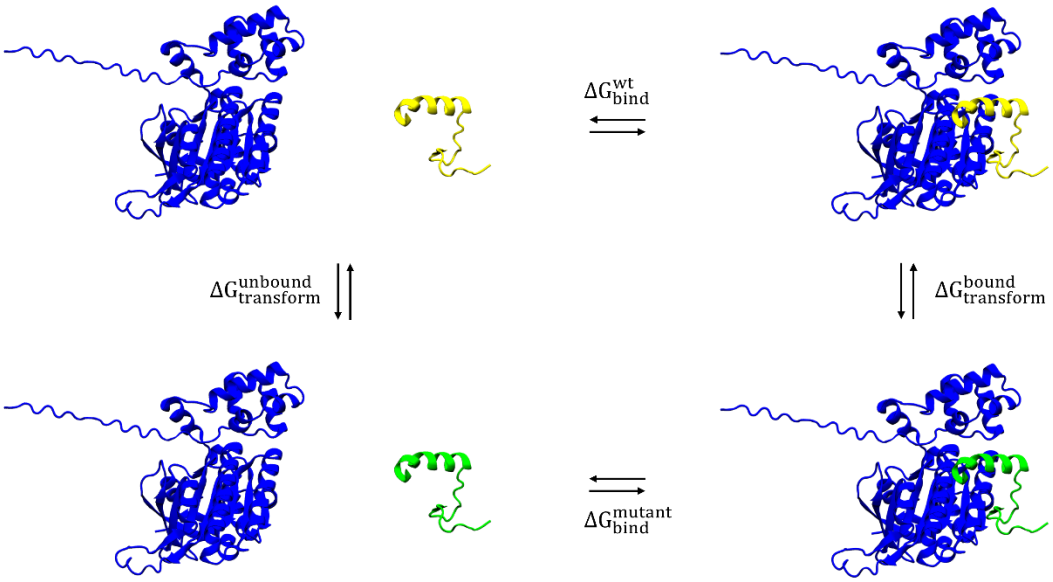

Figure 13. Schematic representation of a thermodynamic cycle. The receptor is shown in blue, the native ligand in yellow and the mutated ligand in green. Horizontal reactions represent binding events, while vertical reactions are alchemical transformations.

In fact, the binding events themselves are particularly computationally demanding to simulate: because of this, to obtain $\Delta\Delta G_{bind}$, computing ΔG_{bind}^{wt} and ΔG_{bind}^{mutant}, following what is known as the direct path, is a very expensive strategy. Conversely, following the alchemical path, meaning computing the free energies associated with the transformations, $\Delta G_{transform}^{unbound}$ and $\Delta G_{transform}^{bound}$, requires less computational effort. Indeed, thanks to the properties of the thermodynamic cycle, the difference between the two sets of free energies have the same value. This concept is shown in equation 41, by applying Equation 35 to a relative binding free energy computation.

$$\Delta\Delta G_{mutant,wt} = \Delta G_{bind}^{wt} - \Delta G_{bind}^{mutant}$$

$$= -RT\ ln\left(\frac{Z_{bound}^{wt} Z_{unbound}^{mutant}}{Z_{bound}^{mutant} Z_{unbound}^{wt}}\right)$$

$$= \Delta G_{transform}^{bound} - \Delta G_{transform}^{unbound}$$

(41)

The perturbation itself consists of a series of intermediate states between one ligand and the other; in this case, since the ligand is a protein, wild type and mutant. These intermediate states are generated by connecting the systems through a virtual coordinate, λ, with values between 0 and 1, where its change from 0 to 1 represents the transformation between mutant and wild type. Namely, the transformation is usually a linear interpolation of the two ligands through λ, the coupling parameter; the result is a modified potential function shown in Equation 42, where V_0 is the original potential and V_1 is the perturbed one.

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1$$

(42)

Two main approaches are available to set up alchemical relative binding free energy calculations. Namely, simulations can be carried out using either a single

topology or a dual topology. In a single topology framework, there is only one set of coordinates, where atom types and coordinates change to reflect the end states[58]. Dummy atoms, meaning atoms that are free from nonbonded interactions, might be present in order to have atoms appear and disappear. Conversely, in a dual topology framework two complete sets of coordinates exist, and the atoms of the group subject to change exist in their two end states simultaneously, albeit not interacting with each other.

An additional tool used in this context are soft core potentials[59]. They are used to have nonbonded interactions belonging to appearing and disappearing atoms smoothly switched on and off as shown in Equations 43 and 44; here, the α parameter is added to adjust the potential's softness. They were developed from the need of removing instabilities that arise when switching on and off such interactions[60].

$$V_{V_1,appearing} = 4\varepsilon\lambda\left[\frac{1}{\left[\alpha(1-\lambda)+\left(\frac{r_{ij}}{\sigma}\right)^6\right]^2}-\frac{1}{\alpha(1-\lambda)+\left(\frac{r_{ij}}{\sigma}\right)^6}\right] \tag{43}$$

$$V_{V_0,disappearing} = 4\varepsilon(1-\lambda)\left[\frac{1}{\left[\alpha\lambda+\left(\frac{r_{ij}}{\sigma}\right)^6\right]^2}-\frac{1}{\alpha\lambda+\left(\frac{r_{ij}}{\sigma}\right)^6}\right] \tag{44}$$

2.4.2.1 Thermodynamic Integration

Thermodynamic integration is one of the most popular frameworks to compute relative binding free energies through alchemical perturbations[59]. Namely, the free energy change associated with two states is obtained as shown in Equation 45.

$$\Delta G_{TI} = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

(45)

To perform such integration, different simulations with different λ values ranging from 0 to 1 are separately carried out. The choice of λ values and the spacing between them can significantly affect the accuracy and convergence properties of the integral, especially for transformations involving large structural changes or the appearance/disappearance of atoms. In the context of relative binding free energy calculations, two sets of simulations are carried out: one in the unbound state and the other in the bound state, both describing the alchemical transformation on the ligand.

3 Applications

3.1 Characterization of RAD51-BRC4 Interaction

3.1.1 Methods

3.1.1.2 Steered Molecular Dynamics

All MD simulations were carried out at in all-atom mode using the GROMACS MD engine[61], using the Amber ff19SB force field[62] for the proteins, and the OPC model for water molecules[63]. The complete structure of the RAD51-BRC4 complex predicted with AlphaFold2 was inserted in a dodecahedron-shaped box, with edges distant 15 Å from the biomolecules. The box was then filled with OPC waters and the system was neutralized and brought to physiological ionic concentration (0.15 M) with NaCl using Joung and Cheatham ions[64]. The system was energy minimized using the steepest descent method and subsequently equilibrated via a 1.2 ns simulation in the NVT ensemble with position restraints of 239 kcal/mol on protein heavy atoms (of which, 400 ps were conducted at 100 K, 400 ps at 200 K, and the last 400 ps at 300 K) and 800 ps in the NPT ensemble (of which, 400 ps were conducted with protein heavy atoms restraints, while the last 400 ps only featured alpha carbon atoms restraints) using the V-rescale thermostat[65] with time constant 0.1 for temperature control and C-rescale barostat[66] with time constant of 0.1 for pressure control.

sMD simulations were carried out via the MOVINGRESTRAINTS directive in PLUMED 2.9[67,68]. Three replicates of 10 ns each were performed using the Martini coarse-grained representation for SAXS spectra calculations, and three replicates of 10 ns using the single-bead model.

A set of 19 SAXS intensities computed at different values of the momentum transfer q, in the range 0.00-0.30 $Å^{-1}$, was used as CV. During the 10 ns of sMD, the positions of the harmonic restraints were linearly interpolated from the values of the SAXS intensities computed on the initial structure, i.e. the AlphaFold2 model, to the experimental ones. The SAXS experimental spectrum for the RAD51-BRC4 complex

was taken from the Small Angle Scattering Biological Data Bank (SASBDB), under accession code SASDQT9. To reduce the influence of experimental noise, the experimental values were taken after performing a 51-point running average on the experimental SAXS spectrum. The force constant was kept constant at a value of 10^6 kJ/(mol*a.u.^{2}) for the entire simulation.

3.1.1.2 Metadynamics

A 100 ns Well-Tempered MetaD[38] run in the NVT ensemble was carried out via PLUMED 2.9, using the Rg as CV. Gaussians of width 0.05 Å and height 0.500 kcal/mol were deposited every 500 steps, using a bias factor of 10. The distances between Cα atoms of the four pairs of cross-linked lysine residues, as indicated by the XL-MS experiments, were restrained through a flat-bottom restraining potential, as implemented in the UPPER_WALLS PLUMED directive, positioned at a value of 30 Å [69] with a force constant of 23.90 kcal/(mol*Å^2). Specifically, the restrained distances were the ones between RAD51's K59 and BRC4's K24, RAD51's K59 and BRC4's K29, RAD51's K65 and BRC4's K29, RAD51's K71 and BRC4's K29. MetaD weights were computed a posteriori using the final bias.

3.1.1.3 MAXENT reweighting

MAXENT reweighting was carried out using a regularization term σ of 0.03. Such figure was chosen among values between 0.005 and 0.5 because of the χ^2 between reweighted and experimental SAXS spectra associated with it, the closer to 1 among the ones taken into consideration. This is typically recommended for a model that adequately describes the experimental data while reducing potential overfitting[70–73].

3.1.1.4 Trajectory analysis

The trajectory obtained from the MetaD simulation was analyzed through principal component analysis (PCA) carried out on the Cartesian coordinates of the system's α carbons after aligning the snapshots on the α carbons of the C-ter domain. The first 5 PCs, explaining a cumulative variance of 93%, were used to compute a

distance matrix on which we performed a weighted cluster analysis using the Quality Threshold (QT) algorithm, with a cutoff of 3 and the maxent weights. The cluster analysis, performed using the py-bussilab python package (https://github.com/bussilab/py-bussilab), was done separately on compact and extended structures, using an Rg value of 2.47 Å to discriminate between the two groups. All reported statistical errors were computed through standard bootstrapping with 400 iterations, after dividing the snapshots from the MetaD-generated trajectory into 10 blocks.

The ensemble of conformations was summarized and visualized by constructing a conformational space network, where the representative structure of each cluster served as a node. Edges in the network were assigned using a distance cutoff of 5, whereas the size of the nodes is proportional to the MAXENT weights. To better visualize the most populated clusters in the network, the weights were rescaled using a logistic function. The Pyvis-0.1.3.1 software[74] was used to represent the network, employing the BarnesHut[75] layout algorithm.

The contact analysis on the reweighted ensemble was performed using the compute_contacts module in MDTraj[76], using the closest-heavy scheme to compute minimum distances between pairs of residues. To define a contact, we employed a distance cutoff of 5 Å. The weighted average of the number of contacts was computed using the weights of the maxent-reweighted ensemble. To highlight the residues mainly involved in the interfacial interaction, we used a 2.5% probability threshold in the reweighted ensemble, and any contacts exceeding this value were deemed as persistent.

3.1.2 Results

3.1.2.1 Probing the Interaction between RAD51's N-ter and BRC4 by Combining AlphaFold2 and XL-MS

As no high-resolution structure of the full RAD51 in complex with BRC4 is available, we relied on an AlphaFold2 model generated by our colleagues, which

would provide an initial reasonable guess of the complex structure[77–79]. To ensure compatibility with experimental data the BRC4 sequence used for generating the model features 39 residues (instead of the aforementioned 35). The additional residues and located at the N-ter of the peptide. In line with previous findings, we observed that AlphaFold predicted a conformational rearrangement of the RAD51 N-ter to allow the BRC4 peptide binding[14,15,23] (**Figure 14)**. Moreover, we noted the presence of a network of polar contacts stabilizing the interaction of the BRC4 C-ter with the RAD51 N-ter, thus suggesting another important interface for the peptide binding, as also highlighted by a previous work[80]. To test the existence of this putative interaction between BRC4 and the protein N-ter, we relied on XL-MS, performed by our colleagues. Indeed, this technique would provide us with useful structural information, overcoming the difficulties of achieving a 3D crystal structure, as we observed the presence of different lysine residues at the predicted interface between the BRC4 peptide and the RAD51 N-ter.

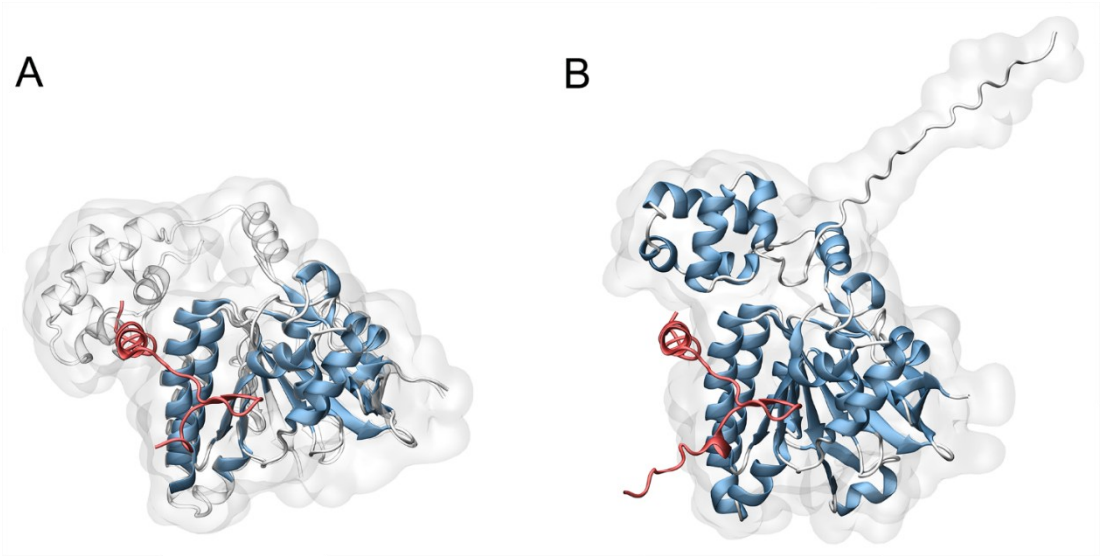

Figure 14. Structure of the RAD51-BRC4 complex. A) PDB entry 1N0W, featuring RAD51's C-ter (in blue) and BRC4 (in red), superposed with PDB entry 5NWL, featuring RAD51's C-ter and N-ter (in white) and lacking the BRC4 repeat. The superposition highlights how the N-ter domain and BRC4 are in overlap. B) The generated AlphaFold2 model, featuring RAD51's C-ter and N-ter (in blue) and the BRC4 repeat (in red). In the prediction, the N-ter is shifted to preserve BRC4 binding and avoid overlap with the repeat.

Crosslinking with bis(sulfosuccinimidyl)suberate (BS3) and XL-MS analysis led to the identification of four high-quality crosslinks between BRC4 and RAD51 (**Figure 15A**, and **Table 1**). Specifically, we observed that Lys1536 and Lys1541, in proximity of the BRC4 LFDE domain, were found to be crosslinked with Lys59, Lys65 and Lys71 located on the RAD51 N-ter in a region encompassing a cluster of alpha-helices, which support RAD51 interaction with the DNA (**Figure 15A**, and **Table 1**) [81].

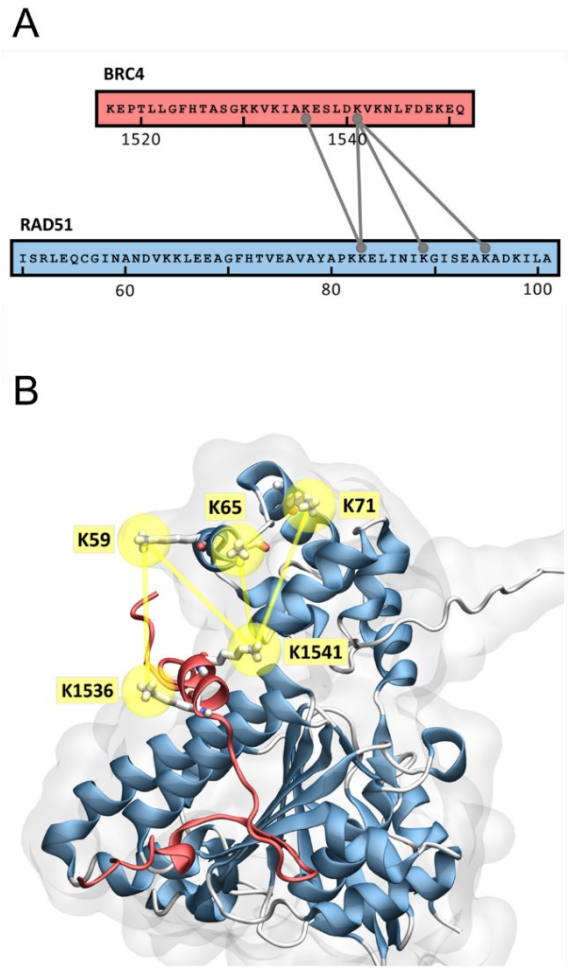

Figure 15. XL-MS data show crosslinks between RAD51 and BRC4. A) Sequence overview of the BRC4 repeat of BRCA2 and His - tagged RAD51 illustrating the detected inter-protein crosslinks filtered for MS quality score > 30 and Cα-Cα distances of 10 – 35 Å. Connecting lines point towards the cross-linked residues. B) Mapping of the identified crosslinks in the AlphaFold2 prediction of the RAD51-BRC4 complex.

Table 1. Table of detected inter-protein crosslinks and their measured Cα-Cα distances in the AlphaFold2 prediction of the His-RAD51 / BRC4 complex (+His), exploited for XL-MS experiments, and the RAD51 / BRC4 complex (-His), utilized in SAXS experiments.

Position in BRCA2	Position in RAD51 (+His)	Cα-Cα distance (Å) (+His)	Position in RAD51 (-His)	Cα-Cα distance (Å) (-His)
1541	89	16.80	65	16.04
1541	95	21.24	71	21.11
1541	83	12.39	59	21.76
1536	83	18.84	59	30.38

This result highlighted that the BRC4 C-ter is crucial to allow the binding of the peptide to RAD51, since it directly interacts with the RAD51 N-ter, displacing it. Then, we aimed to verify the compliance with crosslinking data of His-RAD51/BRC4 and RAD51/BRC4 AlphaFold2 models, respectively employed for XL-MS or SAXS experiments (**Figure 15**). Notably, in both predictions the identified pairs of cross-linked lysine residues displayed Cα-Cα distance compatible with the experimental XL-MS results (**Table 1**) and within the limits of typical distance constraints, further corroborating our initial observation, made on the RAD51-BRC4 AlphaFold2 model, that a hydrogen bond network amidst the BRC4 C-ter and the RAD51 N-ter exists[69]. Additional validation of the generated AlphaFold2 predictions was provided by distances of identified intra-RAD51 crosslinks. Indeed, the majority of identified crosslinks matched permissive distances in the predicted structure, with only two exhibiting long distances > 45 Å. While we cannot a priori exclude that the latter are experimental XL-MS false identifications, they may also derive from RAD51 N-ter flexibility in solution when BRC4 is bound[15]. Indeed, the conformational rearrangements of this domain could potentially bring the lysine residues in closer proximity to each other, thus allowing the crosslinking reaction by BS3.

3.1.2.2 Including the Solvent Contribution is Necessary for Generating a SAXS-Consistent Single-Structure Model of the RAD51-BRC4 Complex

Having validated the RAD51-BRC4 AlphaFold2 prediction by XL-MS, we then compared its computed SAXS spectrum with the experimental one recently determined for the RAD51-BRC4 complex in solution, already available in SASBDB [15]. Nevertheless, we observed a significant discrepancy between the initial AlphaFold2 guess and experimental SAXS data (**Figure 16A**), as the predicted spectrum indicated an overly compact configuration compared to experiments[45].

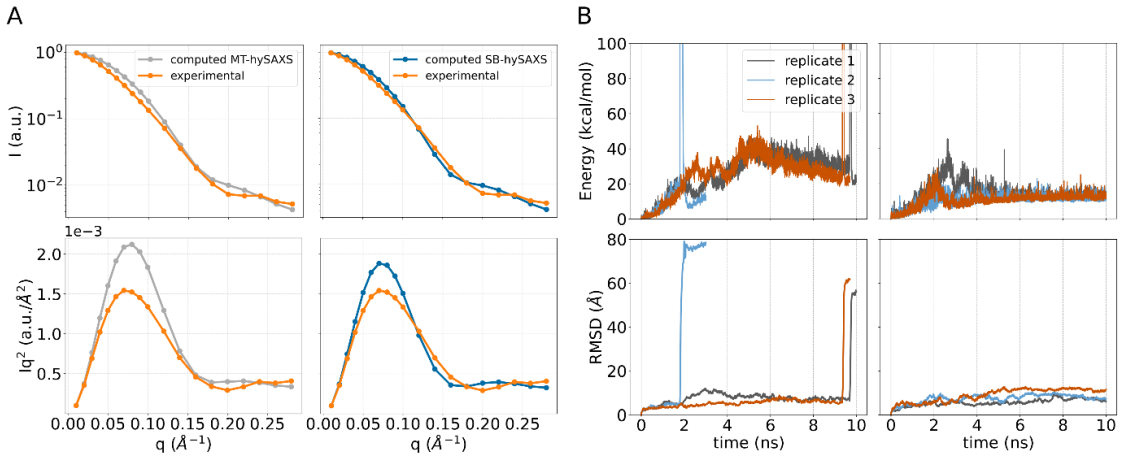

Figure 16. A) Logarithmic scale (top) and Kratky (bottom) plots of the SAXS spectrum computed on the AlphaFold model without (MT-hySAXS scheme, left) and with (SB-hySAXS scheme, right) solvent contribution, compared with the experimentally measured SAXS spectrum of the complex. B) Timeseries of the sMD bias (top) and BRC4 heavy atoms RMSD after alignment on the C-ter (bottom), carried out without (left) and with (right) solvent contribution.

To guide the initial guess towards a configuration in better agreement with experimental data, we took advantage of sMD. Here, the system was driven towards the target state, defined by the experimental SAXS spectrum, through the hySAXS scheme. Unexpectedly, in all three replicates, the simulation invariably resulted in the detachment of the BRC4 repeat from RAD51, as indicated by a rapid rise in the energy and quantified by a marked increase in the RMSD of the complex (**Figure 16B**, left

top and bottom panels, respectively). Notably, this behavior is in contrast with the experimental SAXS data, which refer to the formed complex, as it has been formerly demonstrated[15]. We note that this set of calculations relied on a forward model that does not take into account the contribution of the protein's solvation layer to the SAXS spectra (MT-hySAXS scheme). Remarkably, repeating the procedure by taking into account solvation effects (SB-hySAXS scheme) led to the preservation of the RAD51-BRC4 complex, as indicated by the absence of sudden rises in the energy as well as in RMSD (**Figure 16B**, right panels).

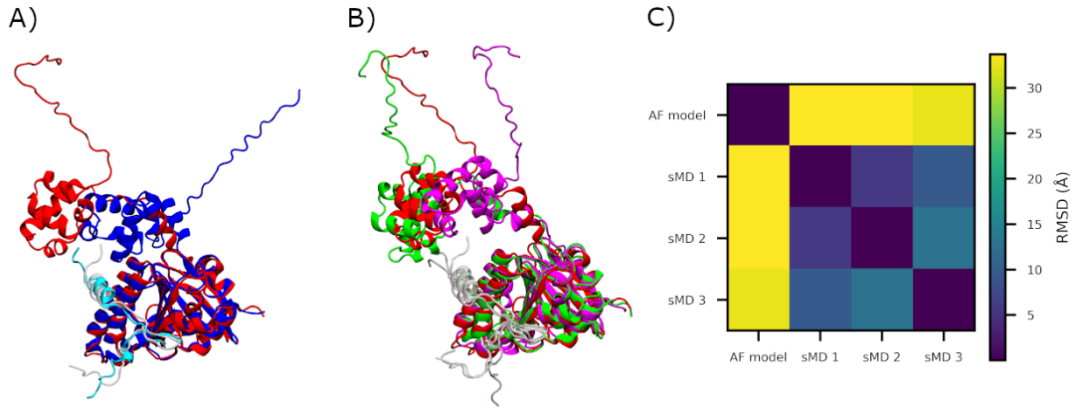

Figure 17 A) One model from the single structure model from replicate 1 of the sMD simulations via the SB-hySAXS scheme (red) superposed to the AlphaFold model (blue), and B) single structure models obtained at the end of the three replicates of sMD (red, green, and purple for replicates 1, 2, and 3, respectively). All structures are aligned on the heavy atoms of RAD51's C-ter domain. C) RMSD matrix between the three single structure models obtained at the end of the sMD replicates and the AlphaFold model; the RMSDs were computed on the system's α carbons after optimal alignment on the α carbons of the C-ter domain.

Importantly, most of the conformational rearrangements required to match the experimental SAXS spectrum involved the N-ter (**Figure 17** and **18**). We stress that this result was achieved by the simulations without any prior information about flexible regions in RAD51, nor by introducing restraint on the BRC4 repeat to prevent detachment. Thus, in the effort to satisfy the experimental SAXS data, the system was able to naturally adapt based on its physical properties. We refer to the configurations

attained by the system at the target state of the sMD simulation, obtained using the SB-hySAXS scheme, as single-structure models. Three replicates of sMD with the SB-hySAXS scheme resulted in comparable single structure models (**Figure 17** and **18**). These models were instrumental for providing a SAXS-consistent initial state for the subsequent XL-MS-informed MetaD simulations. Here, we used the single-structure model from the first replicate of SB-hySAXS sMD for the subsequent stage.

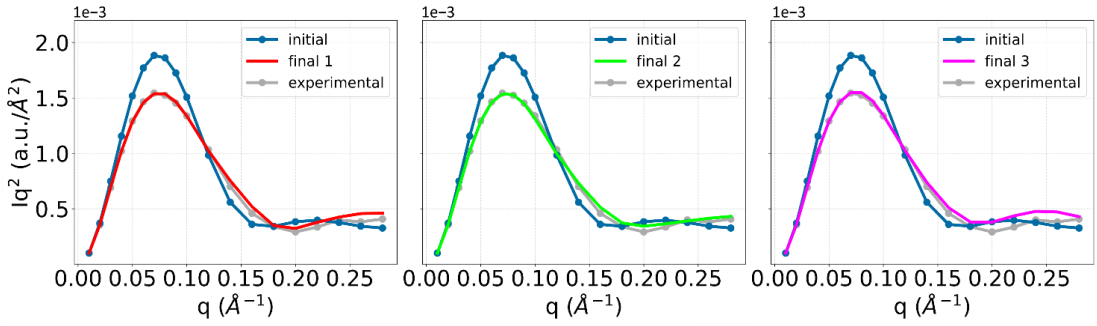

Figure 18 SAXS spectra in the Kratky form for the three replicates of sMD using the SB-hySAXS scheme. The three panels report the spectrum computed at the end of the simulation (labelled final), compared with the starting structure used for the simulations, i.e. the AlphaFold model (in blue), and the experimental SAXS spectrum (in grey).

3.1.2.3 XL-MS-Informed Simulations Enable Determination of a SAXS-Compliant Conformational Ensemble of the RAD51-BRC4 Complex

As previously suggested, binding of the BRC4 peptide to RAD51 could trigger a conformational re-arrangement of RAD51's N-ter, which could behave as an intrinsically disordered domain[15]. Given this scenario, an ensemble of conformations is better suited to achieve a more realistic description of the system[42,82]. Therefore, we aimed to identify a conformational ensemble of structures that would be compatible with the experimental SAXS data[15]. To this end, we first generated a heterogeneous ensemble of RAD51-BRC4 configurations, then the population weights were refined through a reweighting procedure in order to match experimental data. For the first stage, MD simulations did not include any information about the experimental SAXS spectrum, except for the use of the SAXS-consistent single-structure of the RAD51-BRC4 complex obtained from the first replicate of the sMD simulations as a starting

configuration. Specifically, we carried out MetaD simulations; in this case, to promote the exploration of RAD51-BRC4 configurations with varying degrees of structural compactness, we used the Rg of the complex as a CV (**Figure 19A** and **19B**). Additionally, MetaD simulations were integrated with information from the XL-MS experimental data in the form of distance restraints between RAD51's N-ter and the BRC4 repeat (**Figure 19C**). Including such information avoided sampling of irrelevant states that would be incompatible with the maximum distances as suggested by the XL-MS data. This favored a more effective exploration of the system's configurational space and, in turn, optimized the efficiency of the simulation. Therefore, the RAD51-BRC4 structures produced in this way were used as a prior ensemble for the subsequent reweighting procedure according to MAXENT, using the experimental SAXS spectrum as ground truth.

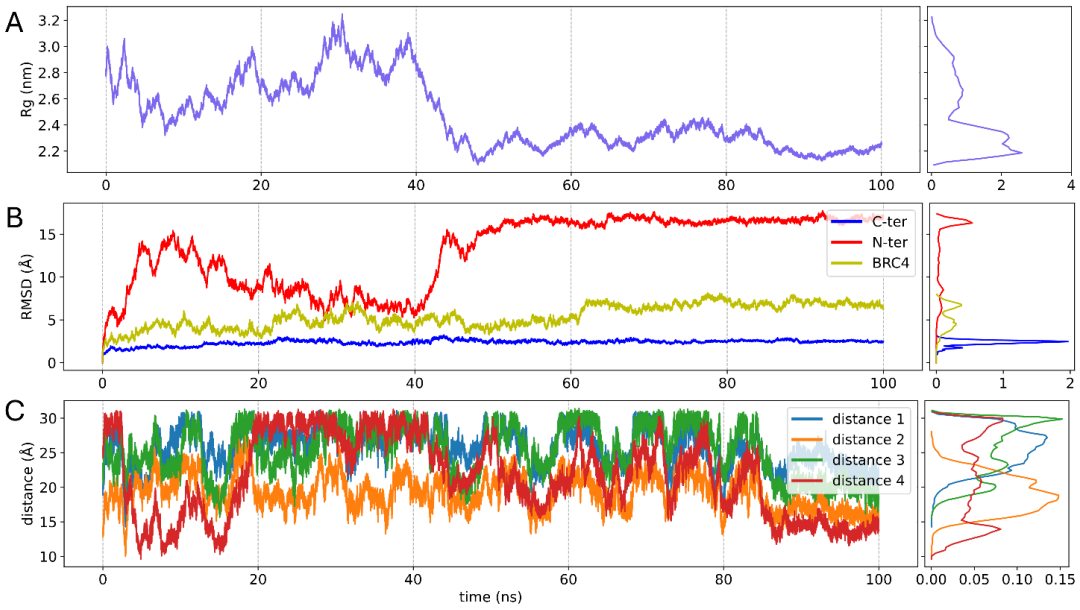

Figure 19 Timeseries and corresponding distributions along the MetaD simulation for A) the Rg of the system, B) RMSDs of RAD51's N-ter and C-ter, and the BRC4 repeat, and C) distances between lysine pairs from the XL-MS experiments. RMSDs were computed on heavy atoms, after aligning separately on each of the three characters (N-ter, C-ter, and BRC4).

As a result, we were able to identify an ensemble of structures with a calculated SAXS spectrum in agreement with the experimental one within statistical error

(**Figure 20**). This reweighting stage resulted in a Kish effective sample size of 7391, out of the 20001 structures used in the analysis. This indicates that a significant portion of structures from the prior ensemble, derived from the MetaD simulation, were retained in the MAXENT-reweighted one and effectively contributed to the final computed SAXS spectrum.

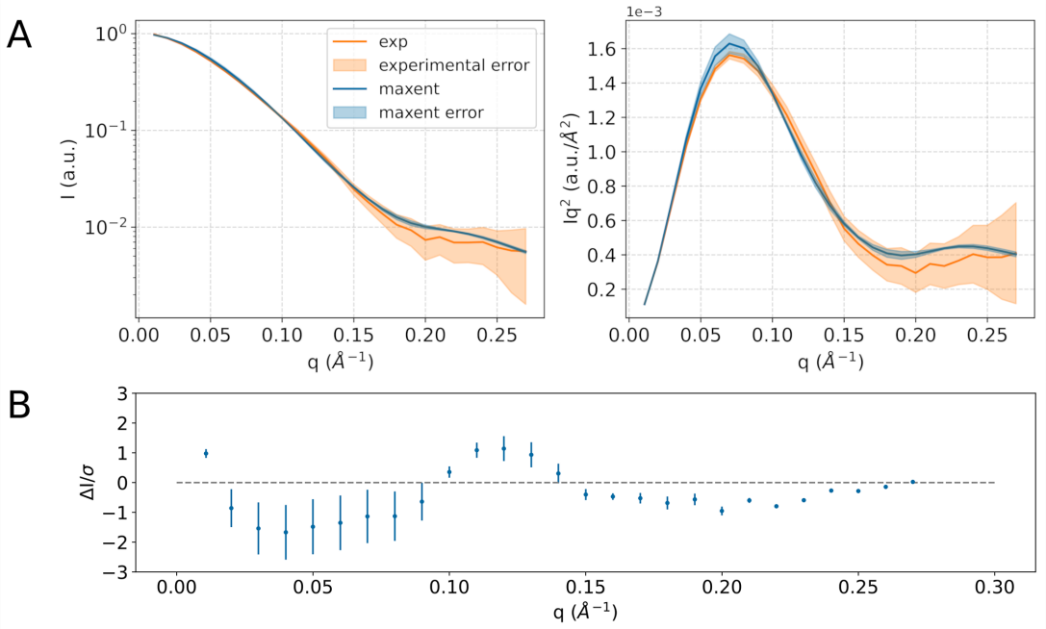

Figure 20. A) Comparison of the reweighted SAXS spectrum through MAXENT with the experimental one, in logarithmic scale (left) and Kratky form (right), for the RAD51-BRC4 complex. B) Residuals plot of the computed spectrum with respect to the experimental one.

3.1.2.4 Charged Residues at the N-ter-BRC4 Interface Bridge Compact and Extended Structures in the Reweighted Ensemble

A closer inspection of the reweighted ensemble revealed the presence of compact and extended conformations, with low and high values of the Rg, respectively. The prior ensemble from MetaD showed a deeper energy well centered at Rg ≈ 23 Å, while maxent reweighting shifted the population by assigning higher weights to the more elongated structures in the shallower energy well at Rg ≈ 29 Å (**Figure 21**). The MAXENT reweighting, therefore, resulted in a more balanced ensemble, with 65%

extended and 35% compact structures. In contrast, the MetaD population was almost entirely composed of compact conformations (~99%).

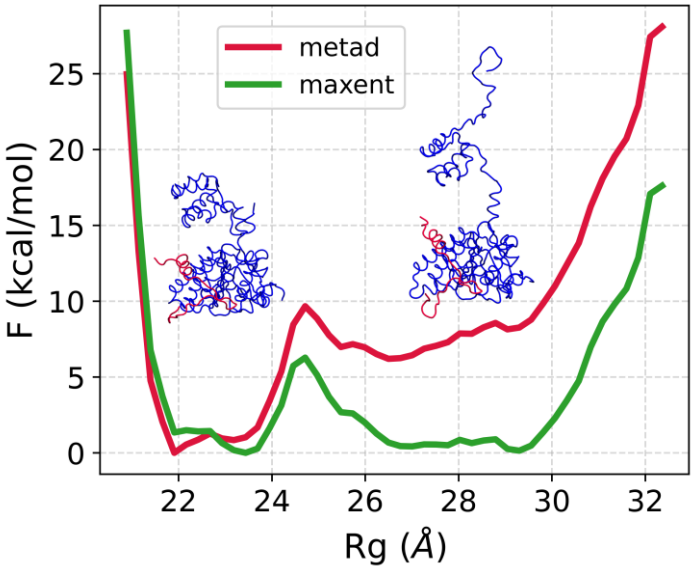

Figure 21. Free energy as a function of the Rg for the MetaD ensemble (red) and the reweighted ensemble via MAXENT (green).

To better appreciate the observed MAXENT conformational heterogeneity of the RAD51-BRC4 complex in solution, we decided to represent it as a conformational space network as reported in **Figure 22**. Nodes in the network represent distinct clusters obtained from the MetaD simulations, with their size proportional to the corresponding MAXENT weights. The presence of edges indicates substantial similarity among different cluster representatives. Therefore, the network topology reflects the composition of the reconstructed conformational ensemble, allowing us to appreciate the distinctive features of the compact and extended subsets of structures (shown as orange and light blue nodes in **Figure 22**, respectively) and their mutual relationships.

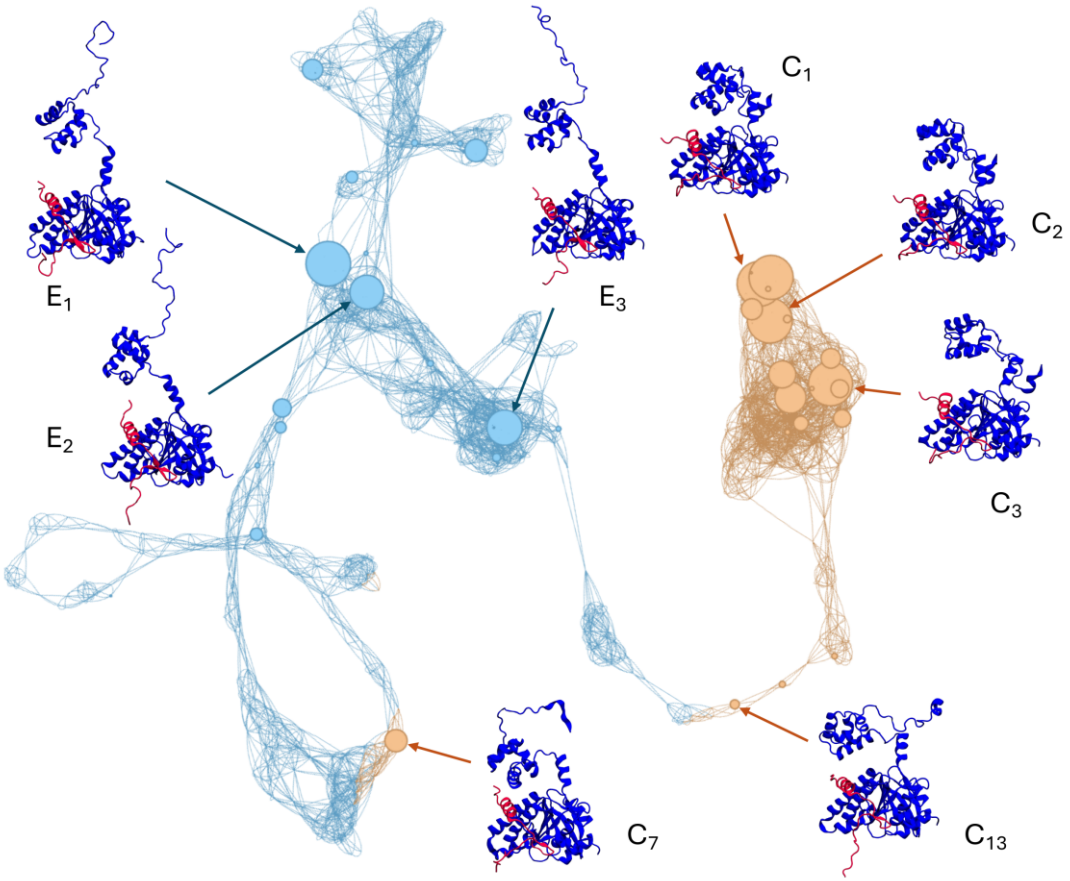

Figure 22. Conformational space network showing the reconstructed RAD51-BRC4 ensemble. Each node represents a cluster, whose size is proportional to the MAXENT reweighted population. Nodes are colored based on the degree of structural compactness, as indicated by the reweighted free energy profile in Figure 21 (orange for compact, light blue for extended). The representative structures of selected nodes, corresponding to either the most populated clusters or regions with peculiar network topology, are shown to highlight the conformational heterogeneity of the reconstructed ensemble.

Most compact structures are located in a dense, highly interconnected region of the plot, highlighting a striking conformational similarity among the representative members of different clusters. Nodes C_1, C_2, and C_3 represent compact states with high statistical weight, showing an increasing and progressive separation between RAD51's N-ter and BRC4. Conversely, the regions of the network representing extended structures are more scattered, reflecting greater dissimilarity due to a higher conformational freedom associated with the full detachment of RAD51's N-ter and

increased disorder. Nodes E_1, E_2, and E_3, are examples of clusters with high statistical weight belonging to the extended subset of conformations. Additionally, node C_{13} emerges as an intermediate structure connecting the compact and extended states, showing a substantially compact conformation with a partly disordered RAD51's N-ter. While marginal in statistical weight, isolated compact states also appear in different regions of the network, with node C_7 being the most populated. As revealed by the conformational space network, the difference in structural compactness is mainly governed by configurational rearrangements of the N-ter domain with respect to the complex between BRC4 and RAD51's C-ter (**Figure 22**). This is also reflected by diverse distributions of the distances between the cross-linked residues within the different clusters, with generally broader distributions shifted to lower values for the clusters of compact structures.

To gain deeper mechanistic insights into RAD51-BRC4 complex dynamics, we analyzed the interaction at the interface between BRC4 and RAD51's N-ter in the MAXENT-reweighted ensemble (**Figure 23**). To this end, we performed a contact analysis between BRC4 and the N-ter, which revealed that the interaction is predominantly mediated by charged and polar residues. In particular, the analysis highlighted the involvement of several charged and polar residues at the interface, including three lysines, three glutamates, one glutamine, and one serine, forming the most persistent contacts. These findings underscore a notable enrichment of lysine and glutamate side chains in driving the interaction. A few hydrophobic residues, such as one leucine and one alanine, also contributed to the interface. Lys59 of RAD51's N-ter, specifically, emerged as a key contact residue, frequently engaging with Glu38 and Gln39 from the BRC4 repeat. These residues form a localized electrostatic and hydrogen-bonding hotspot at the interface, suggesting that their interaction may act as a molecular anchor stabilizing the compact subset of structures. The spatial distribution and contact frequency of the residues suggest that electrostatic interactions between RAD51's N-ter and BRC4 can have a central role in regulating the complex's structural dynamics. From an evolutionary perspective, the residues involved in such interaction are conserved to different degrees. BRCA2 has homologs in all eukaryotic kingdoms,

but has no prokaryotic homolog[83]. BRC4 is one of the most conserved BRC repeats in mammals[84]; among the residues underlined here, Glu38 is conserved in dog BRCA2, while it is substituted by lysine in monkey and pig BRCA2 and by glutamine in hamster and mouse BRCA2. Gln39 on the other hand is conserved across human, dog, pig and monkey BRCA2 and substituted by tyrosine in hamster and mouse BRCA2. On the opposite hand, RAD51 is conserved in eukaryotes and prokaryotes, albeit the latter featuring no BRCA2 counterpart. In mammals, the residues we identified as making contact with BRC4 are conserved across human, mouse, monkey, pig and dog RAD51.

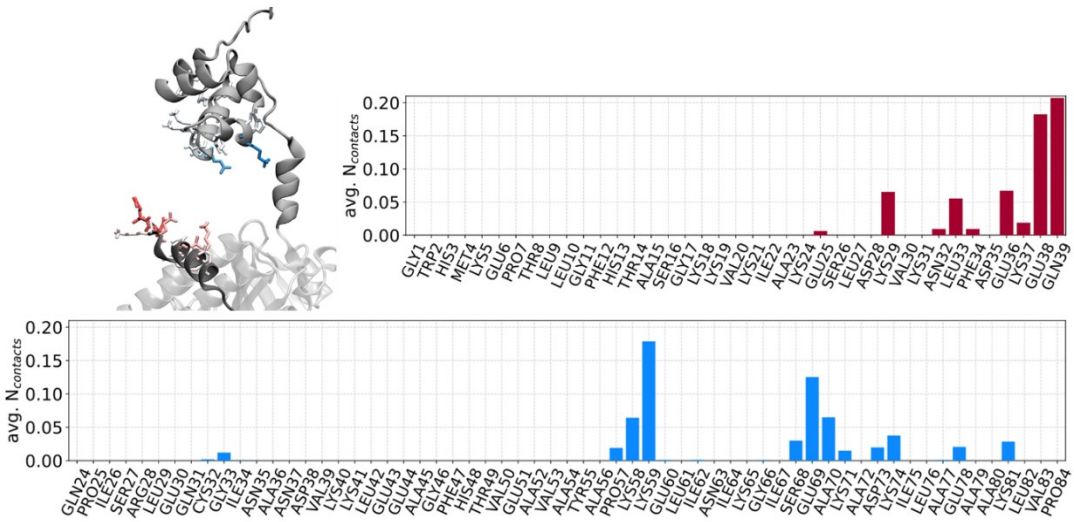

Figure 23. Residue-wise average number of contacts for the structures in the maxent-reweighted ensemble. The analysis focused on contacts for residues from the BRC4 repeat (in red) and RAD51's N-ter (in blue). In the 3D structure, residues with non-zero contacts are displayed in licorice, colored consistently with the plots and with color ranging from white to full solid, indicating low to maximum number of observed contacts, respectively; the C-ter, N-ter and BRC4 are displayed in VMD's NewCartoon representation, colored white, grey and dark grey, respectively.

Whether the presence of both compact and elongated conformations of RAD51-BRC4 in the reweighted ensemble points towards different mechanisms of complex formation is hard to tell. While our analysis shed light on the complex's conformational dynamics in solution, a recognition mechanism between RAD51 and

BRC4 is difficult to infer. Moreover, the presence of the other BRC repeats as well as the whole BRCA2 protein cannot be overlooked, and any biologically relevant claim about a recognition pathway between the two should take such complexity into account.

From a drug discovery point of view, the identification of the RAD51 N-ter-BRC4 interface could be of interest. The available BRCA2-RAD51 disruptors have been developed to make contact with the FXXA and LFDE pockets on RAD51 C-ter[11,12], which are known interaction sites for BRC4; however, given RAD51 N-ter's extreme flexibility, there is no evidence of such compounds making contact with it, or of the opposite[13].

In summary, this project[85] relied on different computational techniques, namely MD enriched with enhanced sampling and experimental coupling techniques to model a complex characterized by intrinsic disorder. As mentioned in the Theory section, these tools are not free from limitations and drawbacks. Plain MD would not necessarily guarantee an efficient sampling of the conformational space; hence our reliance on MetaD to visit a diverse ensemble of conformations. To ensure experimental data compliance we purposefully selected the MAXENT reweighting paradigm, as to retain as many conformations as possible from the original, MetaD-generated ensemble. This reweighting strategy would have failed had the force field been unreliable and had the MetaD protocol generated a conformational ensemble from which no states could account for the experimental SAXS spectrum. However, the Kish size analysis proved that a relevant share of the original ensemble could fit the experimental data, thereby enabling the assessment of the free energy as a function of the degree of compactness of the complex and further interpretation of the conformational ensemble.

3.2 Alchemical Transformations on the RAD51-BRCn Complexes

3.2.1 Methods

The MD simulations were carried out at in all-atom mode using the Amber MD engine, the Amber ff19SB force field[62] for the proteins, the OPC model for water molecules[63] and Li and Merz ions.

3.2.1.1 MM-GBSA and alanine scanning

The structures of all the RAD51-BRC complexes predicted with AlphaFold2 were inserted in a truncated octahedron-shaped box, with edges distant 10 Å from the biomolecules. The box was then filled with OPC waters and the system was neutralized with NaCl.

The system was energy minimized using the steepest descent and conjugate gradient methods with heavy atoms position restraints of 10 kcal/(mol*Å^2). Then, the system was equilibrated in three phases: first in the NVT ensemble during 400 ps, with heavy atoms position restraints of 1 kcal/(mol*Å^2), gradually bringing the system's temperature from 0 to 300 K. The second part of the equilibration was carried out in the NPT ensemble during 200 ps, with heavy atoms position restraints of 0.1 kcal/(mol*Å^2). Finally, the last equilibration phase was carried out unrestrained, in the NPT ensemble, during 200 ps. Production runs in the NPT ensemble were carried out for 20 ns.

MM-GBSA and alanine scanning calculations were carried out using Amber's MMPBSA.py module.

3.2.1.2 Thermodynamic integration

Point mutants of different BRC repeats were generated and their topologies created through Amber's Leap. Then, for each mutant, the wild type and mutant topologies were merged in a single topology, and inserted in a truncated octahedral

box with edges distant 10 Å from the system. Water was added and the system was neutralized and brough to physiological concentration (0.15 mol/L) with NaCl. The topology was then made non redundant by removing shared atoms using Amber's parmed module. 21 equally spaced λ values between 0 and 1 were then selected. Thus, for each mutation, two sets (one concerning the single repeat in water, the other the RAD51-BRC complex) of 21 simulations were carried out, each in three replicas.

Each system was thus minimized using the steepest descent method, then heated to 300 K in the NVT ensemble during 20 ps and then equilibrated in the NPT ensemble for 200 ps. The production simulation was conducted in the NPT ensemble for 20 ns for each selected λ value.

Soft core potentials were applied to the atoms subject to change, initially with the default parameters, namely, $\alpha = 0.5$, then switched to $\alpha = 0.2$.

3.2.2 Results

3.2.2.1 MM-GBSA and alanine scanning

Given the lack of experimental structures, we once again relied on eight RAD51-BRC AlphaFold models, generated by our colleagues. Alanine scanning was carried out on all eight complexes (**Figure 24**, **S1-7**) relying on the MM-GBSA method for binding free energy estimation.

Remarkably, in all eight cases, the phenylalanine residue belonging to the FXXA motif (Phe 8) showed one of the highest contributions to the overall binding free energy of the peptide.

Interestingly, the last acidic residue belonging to the motif corresponding to the LFDE motif in BRC4 (residue 32) showed great contribution to the binding of BRC1 (Asp 32 in **Figure 24**) as well as BRC2 (**Figure S1**), while the residue occupying the same position in BRC4 (**Figure S3**), albeit being acidic as well (Glu 32) contributes to the binding to a much lesser extent. On the other hand, position 32 is occupied by

an aspartate in BRC6 (**Figure S5**), but its contribution is not nearly as relevant as its counterparts in BRC1 or BRC2.

Similarly, the serine occupying position 12 in BRC1 (**Figure 24**), as well as in BRC3, -4, -6, -7, and -8 (**Figure S2, S3, S5, S6, S7**) shows a significant contribution, while in BRC5 (**Figure S4**), where the same position is occupied by a cysteine, its contribution is negligible.

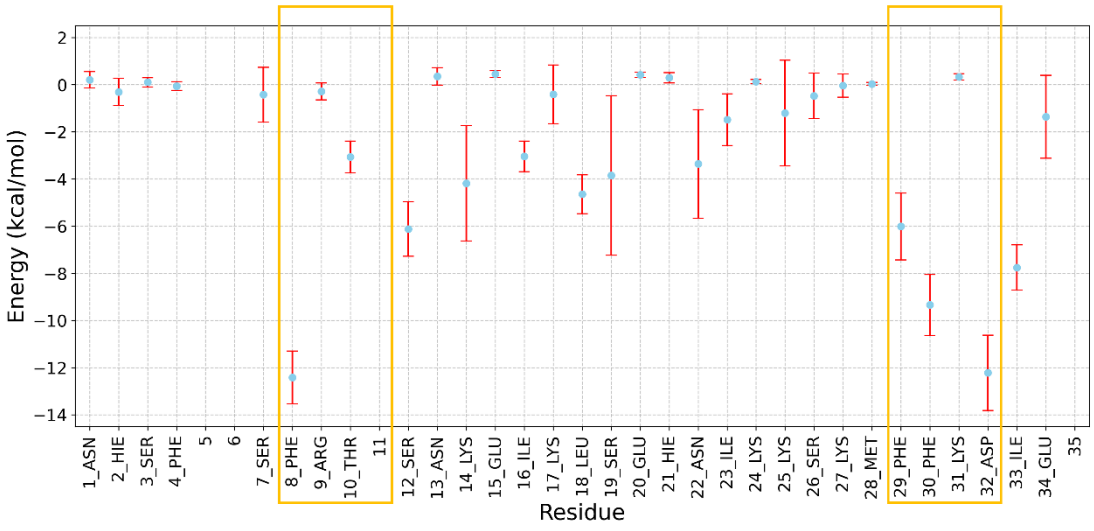

Figure 24. Alanine scanning of BRC1 in the RAD51-BRC1 complex. The strong contribution of the phenylalanine in the FXXA and of the phenylalanine and aspartate of the LFDE (here FFKD) motifs is clearly visible. The residues that are not shown on the x axis are the ones for which no contribution was computed (e.g. alanines)

3.2.2.2 Thermodynamic Integration

Since we identified these peculiar residues, namely the last acidic residue in the LFDE-similar motif (residue 32), and Ser12, adjacent to the FXXA motif, we focused the effort in dissecting their influence on the binding of a single repeat. We started with BRC1 and set up three point mutations: S12C, D32E and K31T. The latter was chosen because position 31 is occupied by a lysine residue in BRC1 and by a threonine in BRC6, neither having much relevance for binding; however, both repeats feature an aspartate in position 32, but it seems to contribute differently to the binding.

We thus suppose that a change in the neighboring residue could affect the aspartate's behavior.

We first carried out the simulations using the default α parameter (α = 0.5). The first mutation, S12C, yielded coherent and smooth potential energy curves (**Figure S8**). This transformation is the less problematic to simulate, since there is no change in charge between serine and cysteine. However, issues arose simulating the transformation relative to the D32E mutation: even though the residues share the same charge, the energy profile looks jagged (**Figure S9**). The same problem, to a much greater extent, resulted from the transformation relative to the K31T mutation. Here the transformation entails net charge change as well: a counterion was then added to the selection of disappearing atoms to balance the net charge. Despite this countermeasure, this transformation's potential energy profile is completely jagged (**Figure S10**).

To overcome this issue we changed the α parameter from its default value of 0.5 to α = 0.2, carrying out only one replica so far. More replicas are needed, but fine tuning the parameter seems a promising strategy (**Figure 25**).

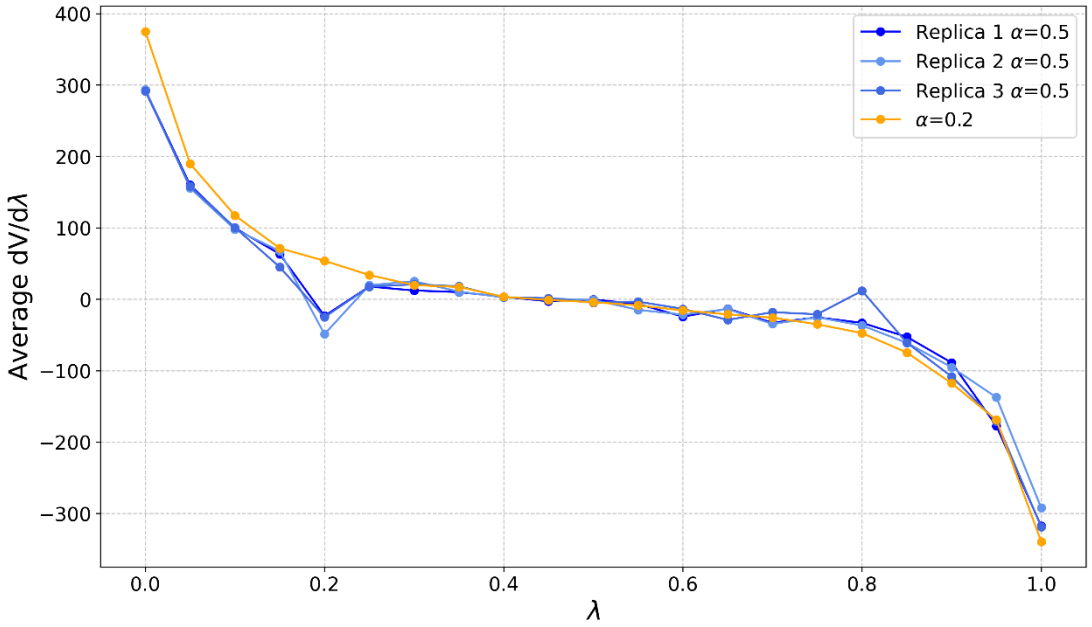

Figure 25. Potential energy curves for the TI simulation involving wild type BRC1 and D32E BRC1, in complex with RAD51, comparing default and custom α parameter.

3.3 RAD51 Filament Modeling

3.3.1 Methods

The structure of the RAD51 heptamer was taken from the Protein Data Bank and its missing 20 residue domain on N-ter was modelled using Modeller. Such structure also features seven ATP units. Given that GoCa currently only supports coarse-graining of amino acids, to account for ATP's contribution to the filament, we treated ATP as if it were a tiny pentapeptide (**Figure 26**).

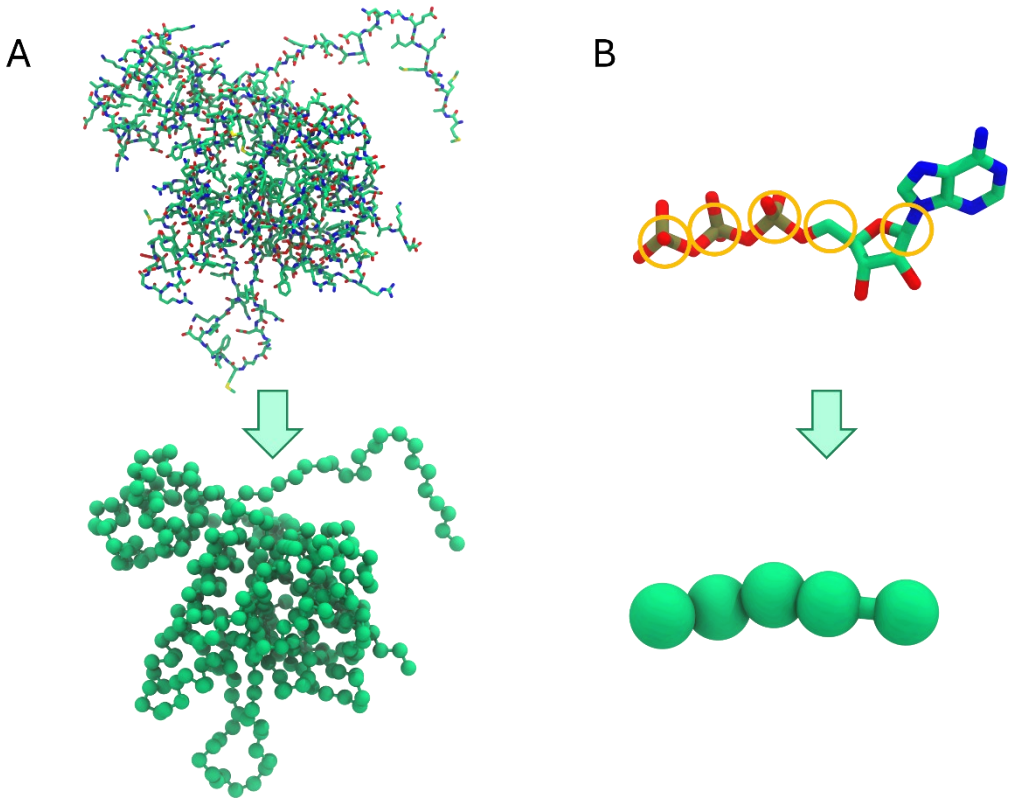

Figure 26. A) Coarse-graining of RAD51; the beads are positioned in the coordinates corresponding to Cα. B) coarse-graining of ATP; the beads are positioned in the coordinates corresponding to the atoms circled in orange.

Hydrogen atoms were added to the structure via Amber's Leap. GoCa's toolbox was used to generate coarse-grained topology and structure files; then, the starting structure was generated using Gromacs, by randomizing the positions of the seven monomers. A single MD instance was then carried out during 80 ns of simulation time with a time step of 4 fs, using the Gromacs MD engine.

3.3.2 Results

The GoCa framework was successful in simulating the assembly of the RAD51 filament. To monitor any misfolding events we kept track of each monomer's internal RMSD; a few instances arose, but quickly resolved. At the same time, the permutation invariant RMSD of the heptamer was computed (**Figure 27**). Since the filament is a

homo-heptamer, a permutation agnostic RMSD calculation would often yield a misleading result, since the monomers would not necessarily assemble in the specific order in which they were indexed. This is by virtue of the GoCa architecture, which allows for any permutation of monomers if they all share the same sequence and folding.

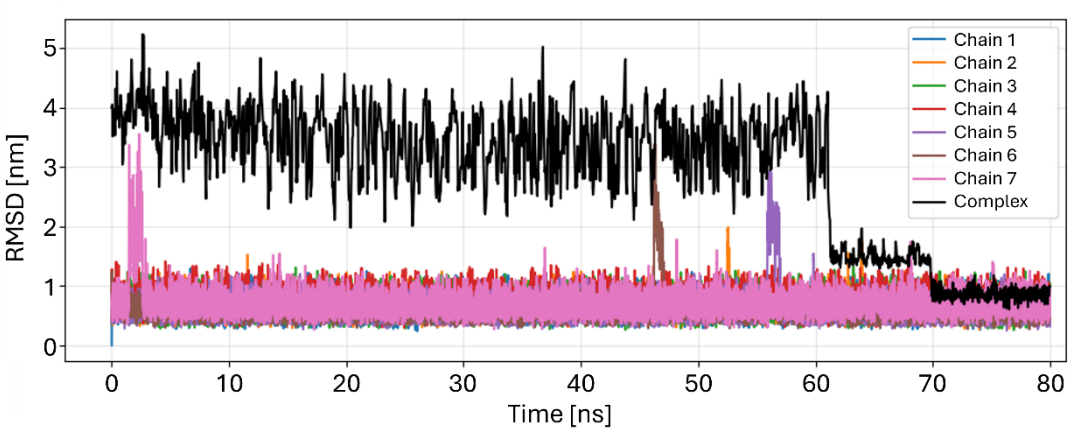

Figure 27. Internal RMSD of each RAD51 monomer (in different colors), and permutation invariant RMSD of the heptamer to be assembled. Each monomer's RMSD is relatively stable at low value. The permutation invariant RMSD has two steep drops that correspond to the formation of a intermediate hexamer and complete heptamer.

From a structural standpoint, the assembly happened stepwise (**Figure 28**), first with the formation of two different oligomers that later fused giving origin to a heptamer that finally came in contact with the last loose monomer.

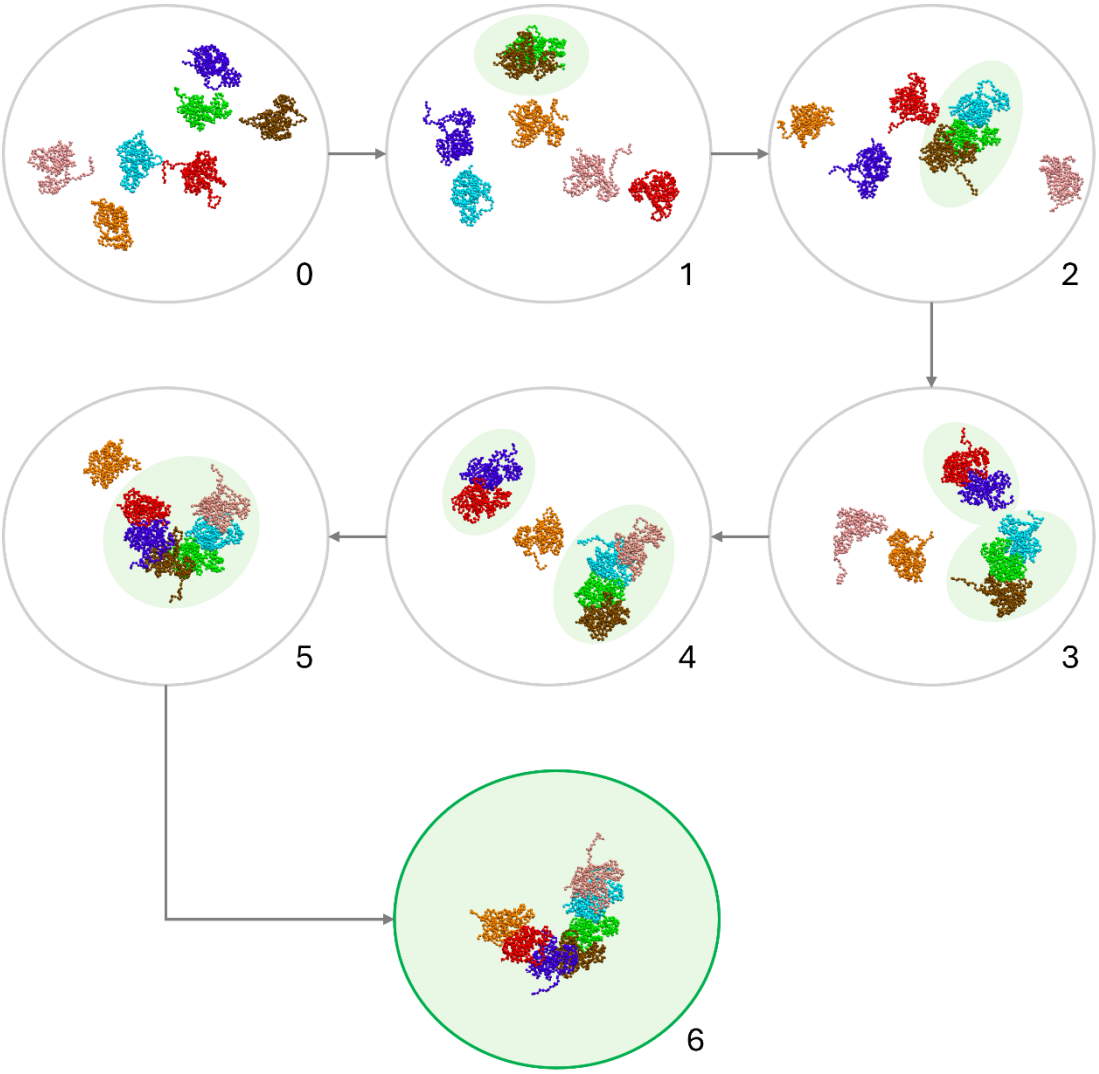

Figure 28. Schematic depiction of RAD51's oligomer assembly. Step 0 shows the random, initial coordinates; then a dimer forms (1), then joined by a third monomer forming a trimer (2), followed by the formation of a second dimer (3). The trimer comes in contact with an additional monomer forming a tetramer (4), that then binds the dimer forming a hexamer (5) that finally binds the last free monomer completing the heptamer (6).

4. Conclusions

These projects stemmed from the need of inquiring into the conformational dynamics of the RAD51-BRCA2 complex, a particularly challenging target given the intrinsic flexibility of both partners. Since the only available structure that details such interaction features RAD51's C-ter and BRC4, the starting point of our project was an AlphaFold prediction of the full RAD51-BRC4 complex; we thus combined MD simulations and experimental techniques (XL-MS and SAXS) to reconstruct the complex's conformational ensemble.

Albeit compatible with XL-MS data, the structure's computed SAXS spectrum did not match experimental results. We applied two different enhanced sampling techniques: first, a sMD instance was carried out through the hySAXS scheme. Here, the CV was defined as the complex's SAXS spectrum, and the structure was driven towards a conformation whose spectrum matched the experimental one. Interestingly, when ignoring the protein's solvation layer contribution to the spectrum, the simulation resulted in BRC4 detachment from RAD51. Conversely, when taking such solvation effect into account, the simulation was able to reach a compatible spectrum without resorting to complex disassembly. A structure generated through sMD was used to carry out MetaD simulations applying positional restraints to crosslinked residues to comply with XL-MS data. Once obtained, the resulting ensemble was reweighted using the MAXENT principle enforcing compliance with SAXS data while preserving the original ensemble's distribution to the maximum possible extent.

The reweighting results highlighted that, in order to match the experimental SAXS spectrum, both compact and elongated structures need to be taken into account. Further contact analysis between BRC4 and RAD51's N-ter shed light on the interaction between the two from a mechanistic standpoint. The interchange between elongated and compact conformations, mainly due to N-ter displacement, seems to be guided by long range interactions between charged residues on RAD51 N-ter and in proximity of the LFDE motif on BRC4.

To gain more insight on RAD51-BRCA2 dynamics we broadened our scope to include all eight BRC repeats. We tackled the issue from a relative binding free energy perspective: indeed, all repeats have a degree of homology, but they differ significantly in their affinity for RAD51. We first carried out computational alanine scanning to estimate the contribution of each residue to the overall binding free energy. Remarkably, the well-known FXXA motif always showed a great contribution corresponding to its phenylalanine residue. Conversely, residues belonging to the LFDE-like motif seemed to vary in their contribution within the repeats. From this result, we inferred the FXXA motif, namely its Phe residue, to be the pivot of RAD51-BRCs interaction, while the LDFE-like motif could be responsible for the differences in affinity between the eight repeats, together with other residues in proximity of the two binding domains. To better inquire into this topic we moved to alchemical transformations of single residues to verify what their influence is on the binding free energy; we focused on residues in the LFDE-like motif and on residues close to the FXXA motif. Fine tuning of the simulation parameters is ongoing and we aim at carrying out the simulations for the selected mutations and possibly more with a solid pipeline.

RAD51 does not only interact with BRCA2, in fact, it often interacts with itself as part of a homo-oligomer. We thus focused on reconstructing the process of RAD51 filament assembly. To do so, we relied on a Gō model, given the availability of a RAD51 heptamer crystal structure. The simulation showed how the filament assembly proceeds stepwise via the formation of smaller oligomers that then take contact with each other to form the full heptamer. The project is ongoing and more replicates are needed to gain relevant knowledge on the assembly process.

References

(1) Scully, R.; Panday, A.; Elango, R.; Willis, N. A. DNA Double-Strand Break Repair-Pathway Choice in Somatic Mammalian Cells. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 698–714. https://doi.org/10.1038/s41580-019-0152-0.

(2) Ceccaldi, R.; Cejka, P. Mechanisms and Regulation of DNA End Resection in the Maintenance of Genome Stability. *Nat. Rev. Mol. Cell Biol.* **2025**, *26* (8), 586–599. https://doi.org/10.1038/s41580-025-00841-4.

(3) Stok, C.; Kok, Y. P.; van den Tempel, N.; van Vugt, M. A. T. M. Shaping the BRCAness Mutational Landscape by Alternative Double-Strand Break Repair, Replication Stress and Mitotic Aberrancies. *Nucleic Acids Res.* **2021**, *49* (8), 4239–4257. https://doi.org/10.1093/nar/gkab151.

(4) Rinaldi, F.; Girotto, S. Structure-Based Approaches in Synthetic Lethality Strategies. *Curr. Opin. Struct. Biol.* **2024**, *88*, 102895. https://doi.org/10.1016/j.sbi.2024.102895.

(5) Falchi, F.; Giacomini, E.; Masini, T.; Boutard, N.; Di Ianni, L.; Manerba, M.; Farabegoli, F.; Rossini, L.; Robertson, J.; Minucci, S.; Pallavicini, I.; Di Stefano, G.; Roberti, M.; Pellicciari, R.; Cavalli, A. Synthetic Lethality Triggered by Combining Olaparib with BRCA2–Rad51 Disruptors. *ACS Chem. Biol.* **2017**, *12* (10), 2491–2497. https://doi.org/10.1021/acschembio.7b00707.

(6) Roberti, M.; Schipani, F.; Bagnolini, G.; Milano, D.; Giacomini, E.; Falchi, F.; Balboni, A.; Manerba, M.; Farabegoli, F.; De Franco, F.; Robertson, J.; Minucci, S.; Pallavicini, I.; Di Stefano, G.; Girotto, S.; Pellicciari, R.; Cavalli, A. Rad51/BRCA2 Disruptors Inhibit Homologous Recombination and Synergize with Olaparib in Pancreatic Cancer Cells. *Eur. J. Med. Chem.* **2019**, *165*, 80–92. https://doi.org/10.1016/j.ejmech.2019.01.008.

(7) Myers, S. H.; Ortega, J. A.; Cavalli, A. Synthetic Lethality through the Lens of Medicinal Chemistry. *J. Med. Chem.* **2020**, *63* (23), 14151–14183. https://doi.org/10.1021/acs.jmedchem.0c00766.

(8) Kim, G.; Ison, G.; McKee, A. E.; Zhang, H.; Tang, S.; Gwise, T.; Sridhara, R.; Lee, E.; Tzou, A.; Philip, R.; Chiu, H.-J.; Ricks, T. K.; Palmby, T.; Russell, A. M.; Ladouceur, G.; Pfuma, E.; Li, H.; Zhao, L.; Liu, Q.; Venugopal, R.; Ibrahim, A.; Pazdur, R. FDA Approval Summary: Olaparib Monotherapy in Patients with Deleterious Germline BRCA-Mutated Advanced Ovarian Cancer Treated with Three or More Lines of Chemotherapy. *Clin. Cancer Res.* **2015**, *21* (19), 4257–4261. https://doi.org/10.1158/1078-0432.CCR-15-0887.

(9) Cruz, C.; Castroviejo-Bermejo, M.; Gutiérrez-Enríquez, S.; Llop-Guevara, A.; Ibrahim, Y. H.; Gris-Oliver, A.; Bonache, S.; Morancho, B.; Bruna, A.; Rueda, O. M.; Lai, Z.; Polanska, U. M.; Jones, G. N.; Kristel, P.; de Bustos, L.; Guzman, M.; Rodríguez, O.; Grueso, J.; Montalban, G.; Caratú, G.; Mancuso, F.; Fasani, R.; Jiménez, J.; Howat, W. J.; Dougherty, B.; Vivancos, A.; Nuciforo, P.; Serres-Créixams, X.; Rubio, I. T.; Oaknin, A.; Cadogan, E.; Barrett, J. C.; Caldas, C.; Baselga, J.; Saura, C.; Cortés, J.; Arribas, J.; Jonkers, J.; Díez, O.; O'Connor, M. J.; Balmaña, J.; Serra, V. RAD51 Foci as a Functional Biomarker of Homologous Recombination Repair and PARP Inhibitor Resistance in Germline BRCA-Mutated Breast Cancer. *Ann. Oncol.* **2018**, *29* (5), 1203–1210. https://doi.org/10.1093/annonc/mdy099.

(10) Previtali, V.; Bagnolini, G.; Ciamarone, A.; Ferrandi, G.; Rinaldi, F.; Myers, S. H.; Roberti, M.; Cavalli, A. New Horizons of Synthetic Lethality in Cancer: Current Development and Future Perspectives. *J. Med. Chem.* **2024**, *67* (14), 11488–11521. https://doi.org/10.1021/acs.jmedchem.4c00113.

(11) Bagnolini, G.; Milano, D.; Manerba, M.; Schipani, F.; Ortega, J. A.; Gioia, D.; Falchi, F.; Balboni, A.; Farabegoli, F.; De Franco, F.; Robertson, J.; Pellicciari, R.; Pallavicini, I.; Peri, S.; Minucci, S.; Girotto, S.; Di Stefano, G.; Roberti, M.; Cavalli, A. Synthetic Lethality in Pancreatic Cancer: Discovery of a New RAD51-BRCA2 Small Molecule Disruptor That Inhibits Homologous Recombination and Synergizes with Olaparib. *J. Med. Chem.* **2020**, *63* (5), 2588–2619. https://doi.org/10.1021/acs.jmedchem.9b01526.

(12) Scott, D. E.; Francis-Newton, N. J.; Marsh, M. E.; Coyne, A. G.; Fischer, G.; Moschetti, T.; Bayly, A. R.; Sharpe, T. D.; Haas, K. T.; Barber, L.; Valenzano, C. R.; Srinivasan, R.; Huggins, D. J.; Lee, M.; Emery, A.; Hardwick, B.; Ehebauer, M.; Dagostin, C.; Esposito, A.; Pellegrini, L.; Perrior, T.; McKenzie, G.; Blundell, T. L.; Hyvönen, M.; Skidmore, J.; Venkitaraman, A. R.; Abell, C. A Small-Molecule Inhibitor of the BRCA2-RAD51 Interaction Modulates RAD51 Assembly and Potentiates DNA Damage-Induced Cell Death. *Cell Chem. Biol.* **2021**, *28* (6), 835-847.e5. https://doi.org/10.1016/j.chembiol.2021.02.006.

(13) Myers, S. H.; Poppi, L.; Rinaldi, F.; Veronesi, M.; Ciamarone, A.; Previtali, V.; Bagnolini, G.; Schipani, F.; Ortega Martínez, J. A.; Girotto, S.; Di Stefano, G.; Farabegoli, F.; Walsh, N.; De Franco, F.; Roberti, M.; Cavalli, A. An 19F NMR Fragment-Based Approach for the Discovery and Development of BRCA2-RAD51 Inhibitors to Pursuit Synthetic Lethality in Combination with PARP Inhibition in Pancreatic Cancer. *Eur. J. Med. Chem.* **2024**, *265*, 116114. https://doi.org/10.1016/j.ejmech.2023.116114.

(14) Brouwer, I.; Moschetti, T.; Candelli, A.; Garcin, E. B.; Modesti, M.; Pellegrini, L.; Wuite, G. J.; Peterman, E. J. Two Distinct Conformational States Define the Interaction of Human RAD 51- ATP with Single-stranded DNA. *EMBO J.* **2018**, *37* (7), e98162. https://doi.org/10.15252/embj.201798162.

(15) Rinaldi, F.; Schipani, F.; Balboni, B.; Catalano, F.; Marotta, R.; Myers, S. H.; Previtali, V.; Veronesi, M.; Scietti, L.; Cecatiello, V.; Pasqualato, S.; Ortega, J. A.; Girotto, S.; Cavalli, A. Isolation and Characterization of Monomeric Human RAD51: A Novel Tool for Investigating Homologous Recombination in Cancer. *Angew. Chem.* **2023**, *135* (51), e202312517. https://doi.org/10.1002/ange.202312517.

(16) Davies, A. A.; Masson, J.-Y.; McIlwraith, M. J.; Stasiak, A. Z.; Stasiak, A.; Venkitaraman, A. R.; West, S. C. Role of BRCA2 in Control of the RAD51 Recombination and DNA Repair Protein. *Mol. Cell* **2001**, *7* (2), 273–282. https://doi.org/10.1016/S1097-2765(01)00175-7.

(17) Jensen, R. B.; Carreira, A.; Kowalczykowski, S. C. Purified Human BRCA2 Stimulates RAD51-Mediated Recombination. *Nature* **2010**, *467* (7316), 678–683. https://doi.org/10.1038/nature09399.

(18) Shahid, T.; Soroka, J.; Kong, E. H.; Malivert, L.; McIlwraith, M. J.; Pape, T.; West, S. C.; Zhang, X. Structure and Mechanism of Action of the BRCA2 Breast Cancer Tumor Suppressor. *Nat. Struct. Mol. Biol.* **2014**, *21* (11), 962–968. https://doi.org/10.1038/nsmb.2899.

(19) Davies, O. R.; Pellegrini, L. Interaction with the BRCA2 C Terminus Protects RAD51–DNA Filaments from Disassembly by BRC Repeats. *Nat. Struct. Mol. Biol.* **2007**, *14* (6), 475–483. https://doi.org/10.1038/nsmb1251.

(20) Esashi, F.; Galkin, V. E.; Yu, X.; Egelman, E. H.; West, S. C. Stabilization of RAD51 Nucleoprotein Filaments by the C-Terminal Region of BRCA2. *Nat. Struct. Mol. Biol.* **2007**, *14* (6), 468–474. https://doi.org/10.1038/nsmb1245.

(21) Halder, S.; Sanchez, A.; Ranjha, L.; Reginato, G.; Ceppi, I.; Acharya, A.; Anand, R.; Cejka, P. Double-Stranded DNA Binding Function of RAD51 in DNA Protection and Its Regulation by BRCA2. *Mol. Cell* **2022**, *82* (19), 3553-3565.e5. https://doi.org/10.1016/j.molcel.2022.08.014.

(22) Lord, C. J.; Ashworth, A. RAD51, BRCA2 and DNA Repair: A Partial Resolution. *Nat. Struct. Mol. Biol.* **2007**, *14* (6), 461–462. https://doi.org/10.1038/nsmb0607-461.

(23) Pellegrini, L.; Yu, D. S.; Lo, T.; Anand, S.; Lee, M.; Blundell, T. L.; Venkitaraman, A. R. Insights into DNA Recombination from the Structure of a RAD51–BRCA2 Complex. *Nature* **2002**, *420* (6913), 287–293. https://doi.org/10.1038/nature01230.

(24) Appleby, R.; Joudeh, L.; Cobbett, K.; Pellegrini, L. Structural Basis for Stabilisation of the RAD51 Nucleoprotein Filament by BRCA2. *Nat. Commun.* **2023**, *14* (1), 7003. https://doi.org/10.1038/s41467-023-42830-1.

(25) Carreira, A.; Kowalczykowski, S. C. Two Classes of BRC Repeats in BRCA2 Promote RAD51 Nucleoprotein Filament Function by Distinct Mechanisms. *Proc. Natl. Acad. Sci.* **2011**, *108* (26), 10448–10453. https://doi.org/10.1073/pnas.1106971108.

(26) González, M. A. Force Fields and Molecular Dynamics Simulations. *Éc. Thématique Société Fr. Neutron.* **2011**, *12*, 169–200. https://doi.org/10.1051/sfn/201112009.

(27) van Gunsteren, W. F.; Oostenbrink, C. Methods for Classical-Mechanical Molecular Simulation in Chemistry: Achievements, Limitations, Perspectives. *J. Chem. Inf. Model.* **2024**, *64* (16), 6281–6304. https://doi.org/10.1021/acs.jcim.4c00823.

(28) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.

(29) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834. https://doi.org/10.1021/ct700324x.

(30) Ballabio, F.; Paissoni, C.; Bollati, M.; De Rosa, M.; Capelli, R.; Camilloni, C. Accurate and Efficient SAXS/SANS Implementation Including Solvation Layer Effects Suitable for Molecular Simulations. *J. Chem. Theory Comput.* **2023**, *19* (22), 8401–8413. https://doi.org/10.1021/acs.jctc.3c00864.

(31) Taketomi, H.; Ueda, Y.; Gō, N. STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION: I. The Effect of Specific Amino Acid Sequence Represented by Specific Inter-unit Interactions. *Int. J. Pept. Protein Res.* **1975**, *7* (6), 445–459. https://doi.org/10.1111/j.1399-3011.1975.tb02465.x.

(32) Walter, L. J.; Quoika, P. K.; Zacharias, M. Structure-Based Protein Assembly Simulations Including Various Binding Sites and Conformations. *J. Chem. Inf. Model.* **2024**, *64* (8), 3465–3476. https://doi.org/10.1021/acs.jcim.4c00212.

(33) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.

(34) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271* (5251), 997–999. https://doi.org/10.1126/science.271.5251.997.

(35) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1* (5), 826–843. https://doi.org/10.1002/wcms.31.

(36) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations. *Living J. Comput. Mol. Sci.* **2022**, *4* (1). https://doi.org/10.33011/livecoms.4.1.1583.

(37) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A. E., Reich, S., Skeel, R. D., Eds.; Springer: Berlin, Heidelberg, 1999; pp 39–65. https://doi.org/10.1007/978-3-642-58360-5_2.

(38) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(39) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603.

(40) Orioli, S.; Larsen, A. H.; Bottaro, S.; Lindorff-Larsen, K. Chapter Three - How to Learn from Inconsistencies: Integrating Molecular Simulations with Experimental Data. In *Progress in Molecular Biology and Translational Science*; Strodel, B., Barz, B., Eds.; Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly; Academic Press, 2020; Vol. 170, pp 123–176. https://doi.org/10.1016/bs.pmbts.2019.12.006.

(41) Ravera, E.; Sgheri, L.; Parigi, G.; Luchinat, C. A Critical Assessment of Methods to Recover Information from Averaged Data. *Phys. Chem. Chem. Phys.* **2016**, *18* (8), 5686–5701. https://doi.org/10.1039/C5CP04077A.

(42) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116. https://doi.org/10.1016/j.sbi.2016.12.004.

(43) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian Inference Method for Heterogeneous Systems. *Sci. Adv.* **2016**, *2* (1), e1501177. https://doi.org/10.1126/sciadv.1501177.

(44) Skou, S.; Gillilan, R. E.; Ando, N. Synchrotron-Based Small-Angle X-Ray Scattering of Proteins in Solution. *Nat. Protoc.* **2014**, *9* (7), 1727–1739. https://doi.org/10.1038/nprot.2014.116.

(45) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19, Part A), 2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.

(46) Guinier, A. La diffraction des rayons X aux tres petits angles : applications a l'etude de phenomenes ultramicroscopiques. *Ann. Phys.* **1939**, No. 11. Sér. 12, 161–237.

(47) Niebling, S.; Björling, A.; Westenhoff, S. MARTINI Bead Form Factors for the Analysis of Time-Resolved X-Ray Scattering of Proteins. *J. Appl. Crystallogr.* **2014**, *47* (4), 1190–1198. https://doi.org/10.1107/S1600576714009959.

(48) Paissoni, C.; Jussupow, A.; Camilloni, C. Martini Bead Form Factors for Nucleic Acids and Their Application in the Refinement of Protein–Nucleic Acid Complexes against SAXS Data. *J. Appl. Crystallogr.* **2019**, *52* (2), 394–402. https://doi.org/10.1107/S1600576719002450.

(49) Yang, S.; Park, S.; Makowski, L.; Roux, B. A Rapid Coarse Residue-Based Computational Method for X-Ray Solution Scattering Characterization of Protein Folds and Multiple Conformational States of Large Protein Complexes. *Biophys. J.* **2009**, *96* (11), 4449–4463. https://doi.org/10.1016/j.bpj.2009.03.036.

(50) Cesari, A.; Reißer, S.; Bussi, G. Using the Maximum Entropy Principle to Combine Simulations and Solution Experiments. *Computation* **2018**, *6* (1), 15. https://doi.org/10.3390/computation6010015.

(51) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911–2937. https://doi.org/10.1021/acs.jcim.7b00564.

(52) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31* (2), 114–122. https://doi.org/10.1002/minf.201100135.

(53) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268* (5214), 1144–1149. https://doi.org/10.1126/science.7761829.

(54) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003**, *330* (4), 891–913. https://doi.org/10.1016/S0022-2836(03)00610-7.

(55) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127–6129. https://doi.org/10.1021/ja00172a038.

(56) Tan, C.; Tan, Y.-H.; Luo, R. Implicit Nonpolar Solvent Models. *J. Phys. Chem. B* **2007**, *111* (42), 12263–12274. https://doi.org/10.1021/jp073399n.

(57) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121* (36), 8133–8143. https://doi.org/10.1021/ja990935j.

(58) Pearlman, D. A. *A Comparison of Alternative Approaches to Free Energy Calculations*. ACS Publications. https://doi.org/10.1021/j100056a020.

(59) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. *J. Chem. Phys.* **2007**, *127* (21), 214108. https://doi.org/10.1063/1.2799191.

(60) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222* (6), 529–539. https://doi.org/10.1016/0009-2614(94)00397-1.

(61) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; Vondrumen, R.; Vanderspoel, D.; Sijbers, A.; Keegstra, H.; Renardus, M. GROMACS - A PARALLEL COMPUTER FOR MOLECULAR-DYNAMICS SIMULATIONS: 4th International Conference on Computational Physics (PC 92). *Phys. Comput. 92* **1993**, 252–256.

(62) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(63) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.

(64) Joung, I. S.; Cheatham, T. E. I. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. https://doi.org/10.1021/jp8001614.

(65) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.

(66) Bernetti, M.; Bussi, G. Pressure Control Using Stochastic Cell Rescaling. *J. Chem. Phys.* **2020**, *153* (11), 114107. https://doi.org/10.1063/5.0020514.

(67) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A.; The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16* (8), 670–673. https://doi.org/10.1038/s41592-019-0506-8.

(68) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.

(69) Merkley, E. D.; Rysavy, S.; Kahraman, A.; Hafen, R. P.; Daggett, V.; Adkins, J. N. Distance Restraints from Crosslinking Mass Spectrometry: Mining a Molecular Dynamics Simulation Database to Evaluate Lysine–Lysine Distances. *Protein Sci.* **2014**, *23* (6), 747–759. https://doi.org/10.1002/pro.2458.

(70) Franke, D.; Jeffries, C. M.; Svergun, D. I. Correlation Map, a Goodness-of-Fit Test for One-Dimensional X-Ray Scattering Spectra. *Nat. Methods* **2015**, *12* (5), 419–422. https://doi.org/10.1038/nmeth.3358.

(71) Trewhella, J.; Duff, A. P.; Durand, D.; Gabel, F.; Guss, J. M.; Hendrickson, W. A.; Hura, G. L.; Jacques, D. A.; Kirby, N. M.; Kwan, A. H.; Pérez, J.; Pollack, L.; Ryan, T. M.; Sali, A.; Schneidman-Duhovny, D.; Schwede, T.; Svergun, D. I.; Sugiyama, M.; Tainer, J. A.; Vachette, P.; Westbrook, J.; Whitten, A. E. 2017 Publication Guidelines for Structural Modelling of Small-Angle Scattering Data from Biomolecules in Solution: An Update. *Acta Crystallogr. Sect. Struct. Biol.* **2017**, *73* (9), 710–728. https://doi.org/10.1107/S2059798317011597.

(72) Pedersen, J. S. Analysis of Small-Angle Scattering Data from Colloids and Polymer Solutions: Modeling and Least-Squares Fitting. *Adv. Colloid Interface Sci.* **1997**, *70*, 171–210. https://doi.org/10.1016/S0001-8686(97)00312-6.

(73) Andrae, R.; Schulze-Hartung, T.; Melchior, P. Dos and Don'ts of Reduced Chi-Squared. arXiv December 16, 2010. https://doi.org/10.48550/arXiv.1012.3754.

(74) Perrone, G.; Unpingco, J.; Lu, H. Network Visualizations with Pyvis and VisJS. arXiv June 2, 2020. https://doi.org/10.48550/arXiv.2006.04951.

(75) Barnes, J.; Hut, P. A Hierarchical O(N Log N) Force-Calculation Algorithm. *Nature* **1986**, *324* (6096), 446–449. https://doi.org/10.1038/324446a0.

(76) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.

(77) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(78) Aithani, L.; Alcaide, E.; Bartunov, S.; Cooper, C. D. O.; Doré, A. S.; Lane, T. J.; Maclean, F.; Rucktooa, P.; Shaw, R. A.; Skerratt, S. E. Advancing Structural Biology through Breakthroughs in AI. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102601. https://doi.org/10.1016/j.sbi.2023.102601.

(79) Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (11), 681–697. https://doi.org/10.1038/s41580-019-0163-x.

(80) Subramanyam, S.; Jones, W. T.; Spies, M.; Spies, M. A. Contributions of the RAD51 N-Terminal Domain to BRCA2-RAD51 Interaction. *Nucleic Acids Res.* **2013**, *41* (19), 9020–9032. https://doi.org/10.1093/nar/gkt691.

(81) Aihara, H.; Ito, Y.; Kurumizaka, H.; Yokoyama, S.; Shibata, T. The N-Terminal Domain of the Human Rad51 Protein Binds DNA: Structure and a DNA Binding Surface as Revealed by NMR. *J. Mol. Biol.* **1999**, *290* (2), 495–504. https://doi.org/10.1006/jmbi.1999.2904.

(82) Fisher, C. K.; Stultz, C. M. Constructing Ensembles for Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2011**, *21* (3), 426–431. https://doi.org/10.1016/j.sbi.2011.04.001.

(83) Pfeffer, C. M.; Ho, B. N.; Singh, A. T. K. The Evolution, Functions and Applications of the Breast Cancer Genes BRCA1 and BRCA2. *Cancer Genomics Proteomics* **2017**, *14* (5), 293–298.

(84) Bignell, G.; Micklem, G.; Stratton, M. R.; Ashworth, A.; Wooster, R. The BRC Repeats Are Conserved in Mammalian BRCA2 Proteins. *Hum. Mol. Genet.* **1997**, *6* (1), 53–58. https://doi.org/10.1093/hmg/6.1.53.

(85) Bresciani, V.; Rinaldi, F.; Franco, P.; Girotto, S.; Cavalli, A.; Langer, J. D.; Masetti, M.; Bernetti, M. Dissecting the RAD51–BRC4 Interaction Landscape through Integrative Molecular Simulations and Experimental Biophysics. *J. Chem. Inf. Model.* **2025**. https://doi.org/10.1021/acs.jcim.5c01639.

Supporting Information

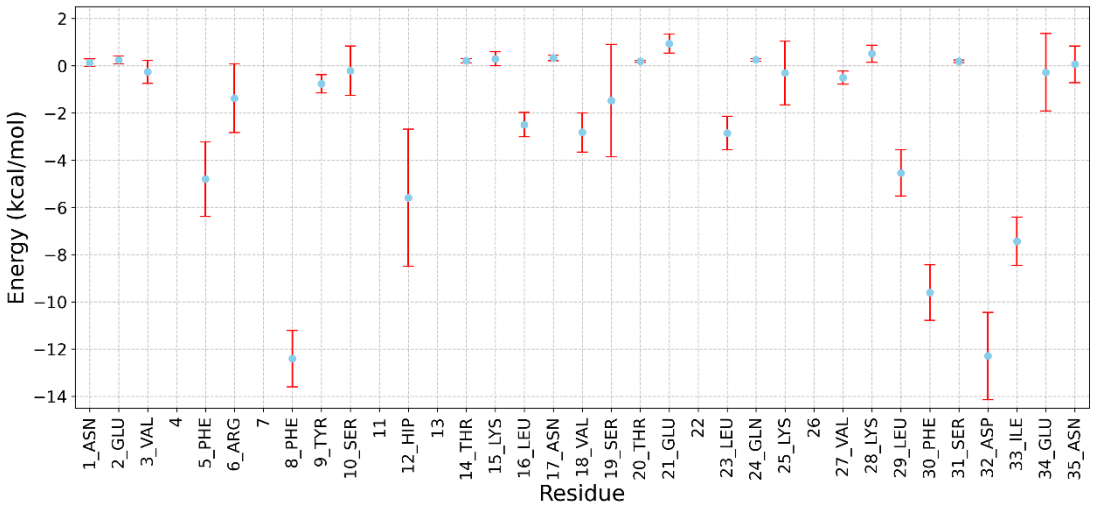

Figure S1. Alanine scanning of BRC2 in the RAD51-BRC2 complex.

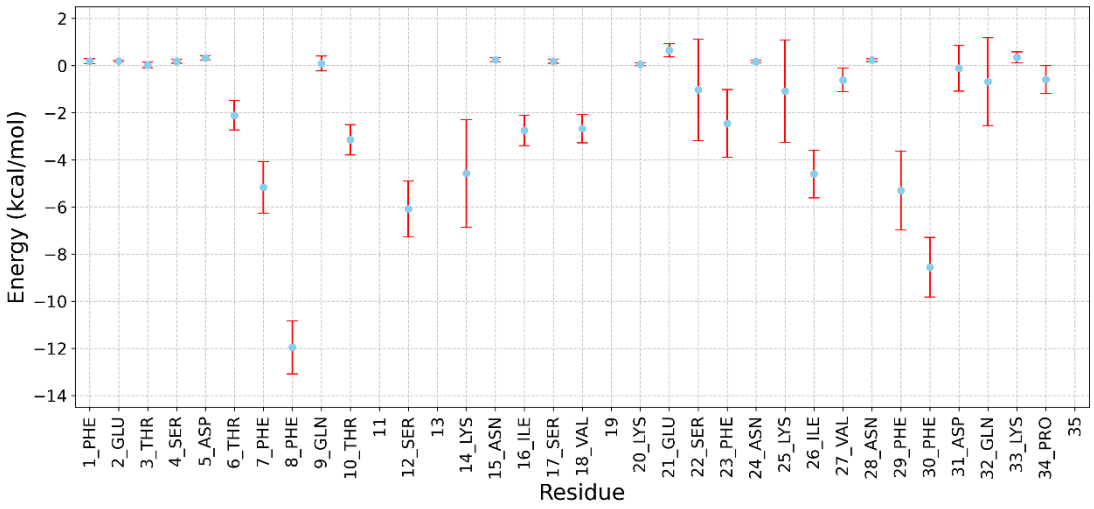

Figure S2. Alanine scanning of BRC3 in the RAD51-BRC3 complex.

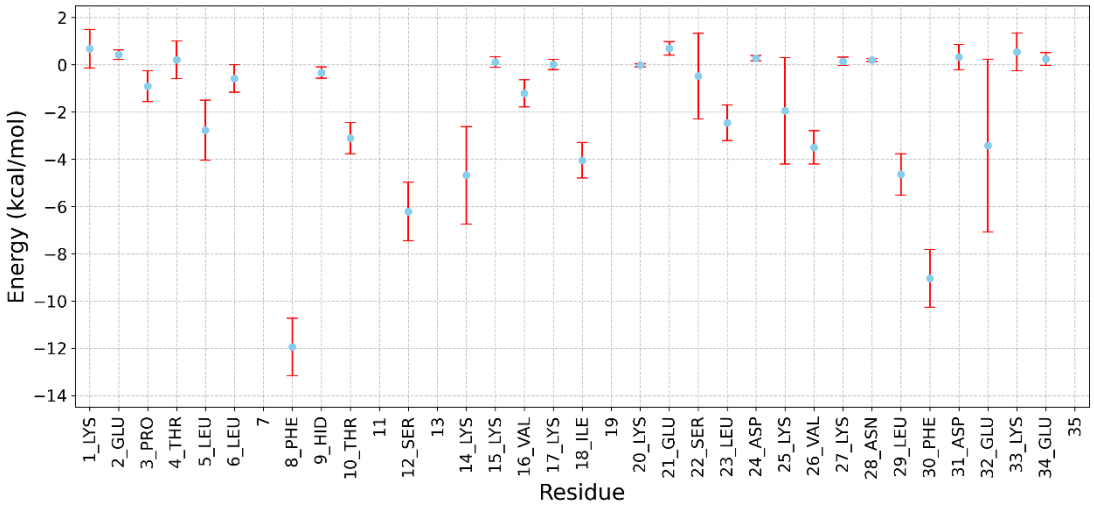

Figure S3. Alanine scanning of BRC4 in the RAD51-BRC4 complex.

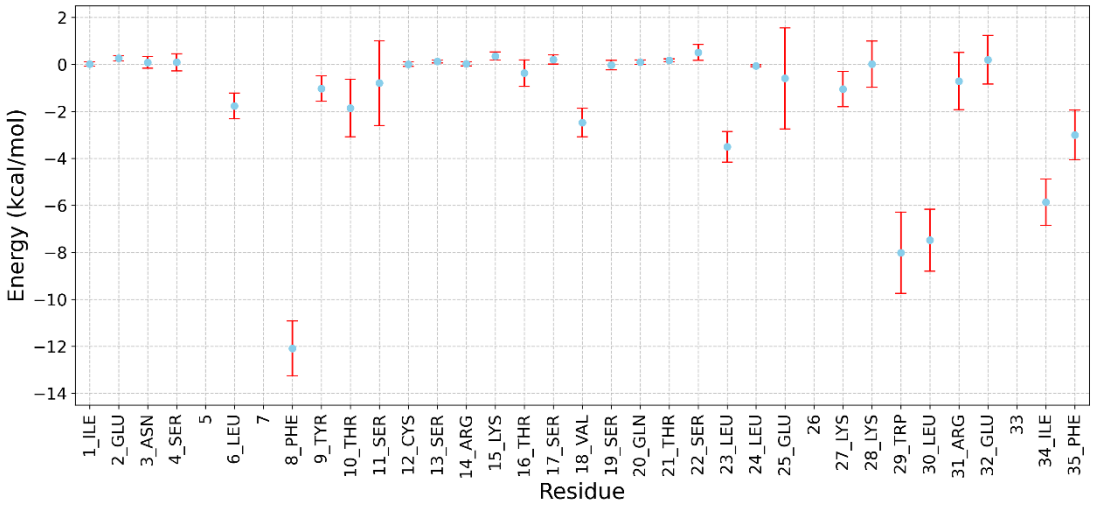

Figure S4. Alanine scanning of BRC5 in the RAD51-BRC5 complex.

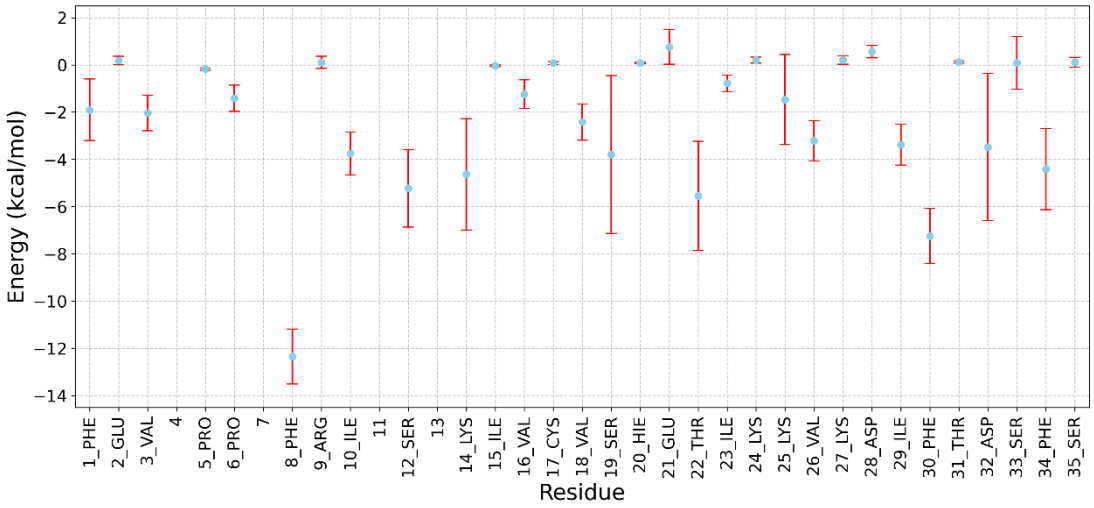

Figure S5. Alanine scanning of BRC6 in the RAD51-BRC6 complex.

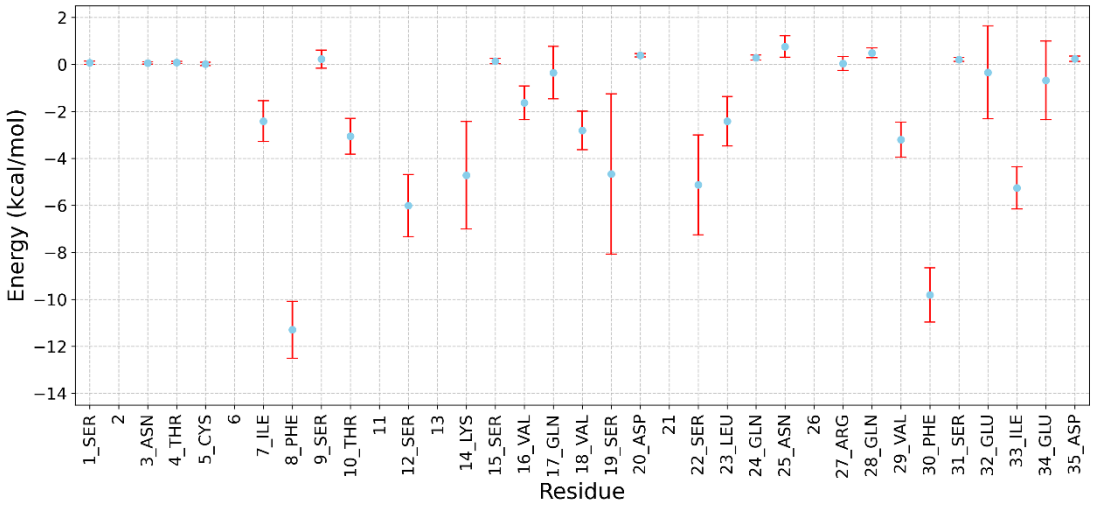

Figure S6. Alanine scanning of BRC7 in the RAD51-BRC7 complex

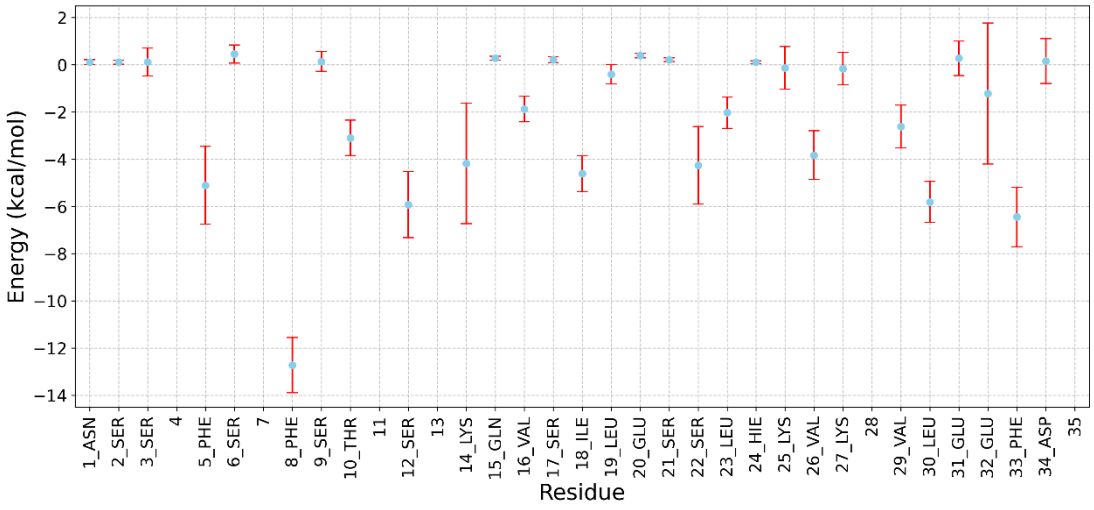

Figure S7. Alanine scanning of BRC8 in the RAD51-BRC8 complex.

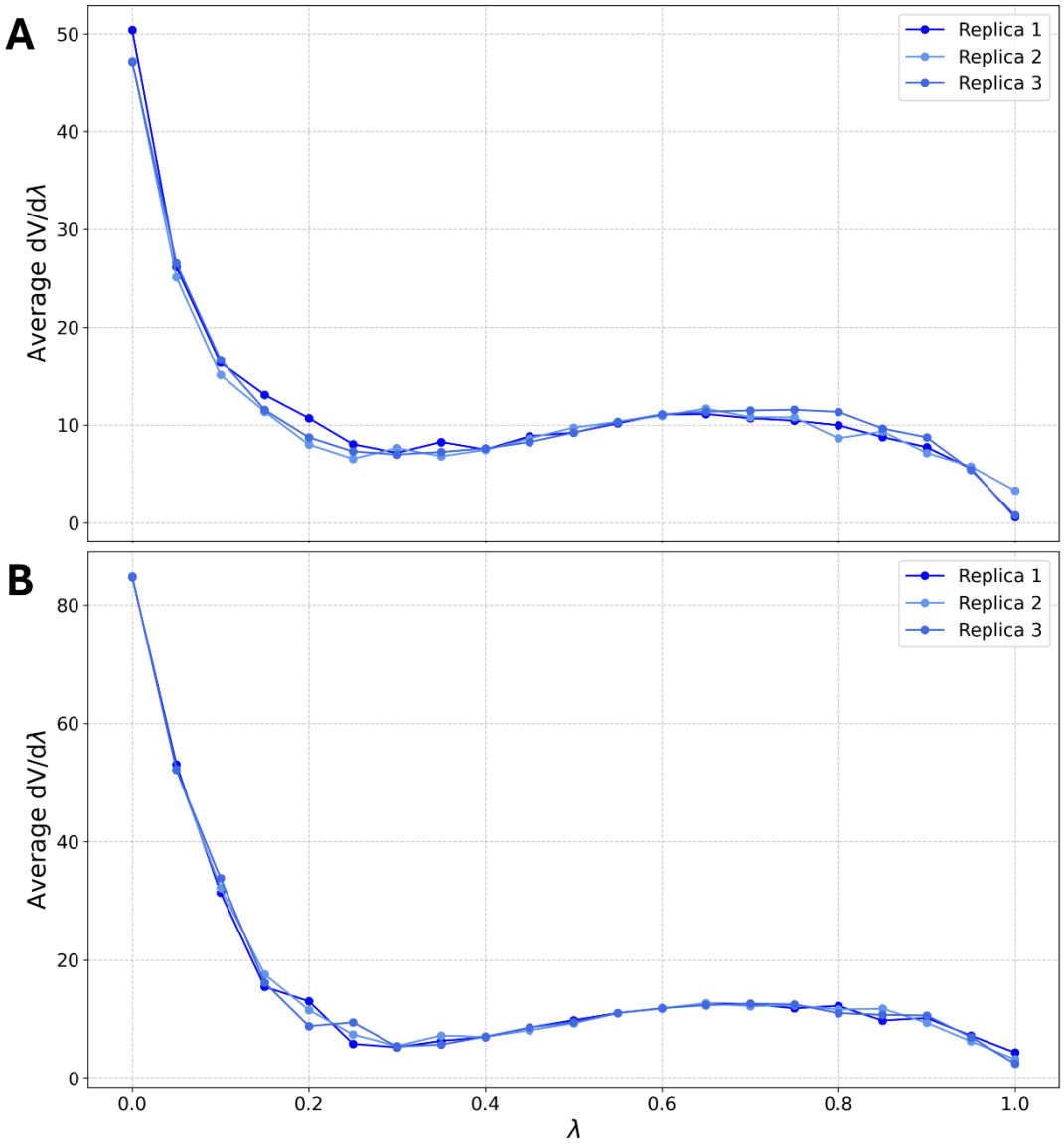

Figure S8 Potential energy curves for the TI simulation involving wild type BRC1 and S12C BRC1, using the default α value. A) Peptide in water B) RAD51-BRC complex.

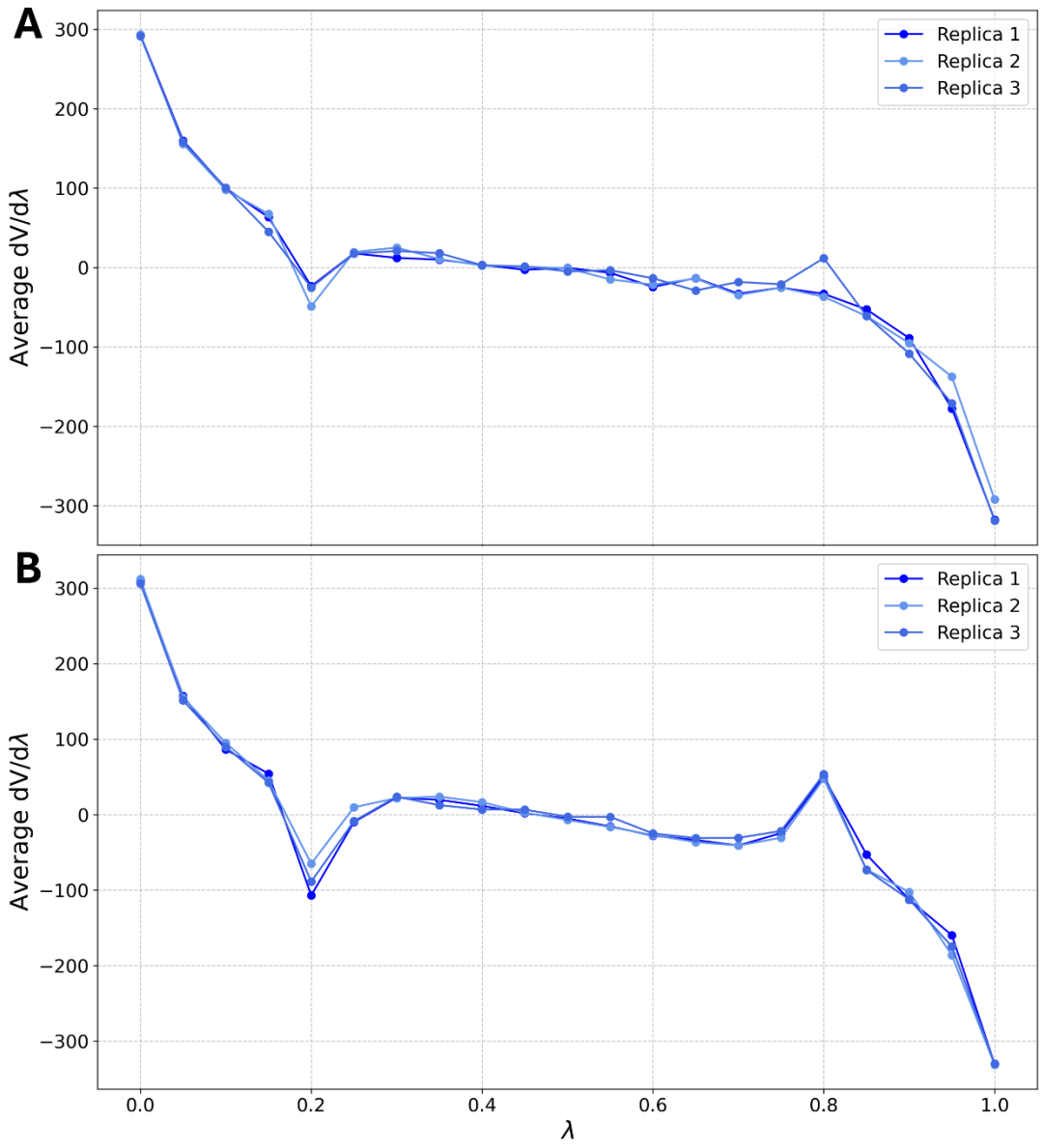

Figure S9 Potential energy curves for the TI simulation involving wild type BRC1 and D32E BRC1, using the default α value. A) Peptide in water B) RAD51-BRC complex.

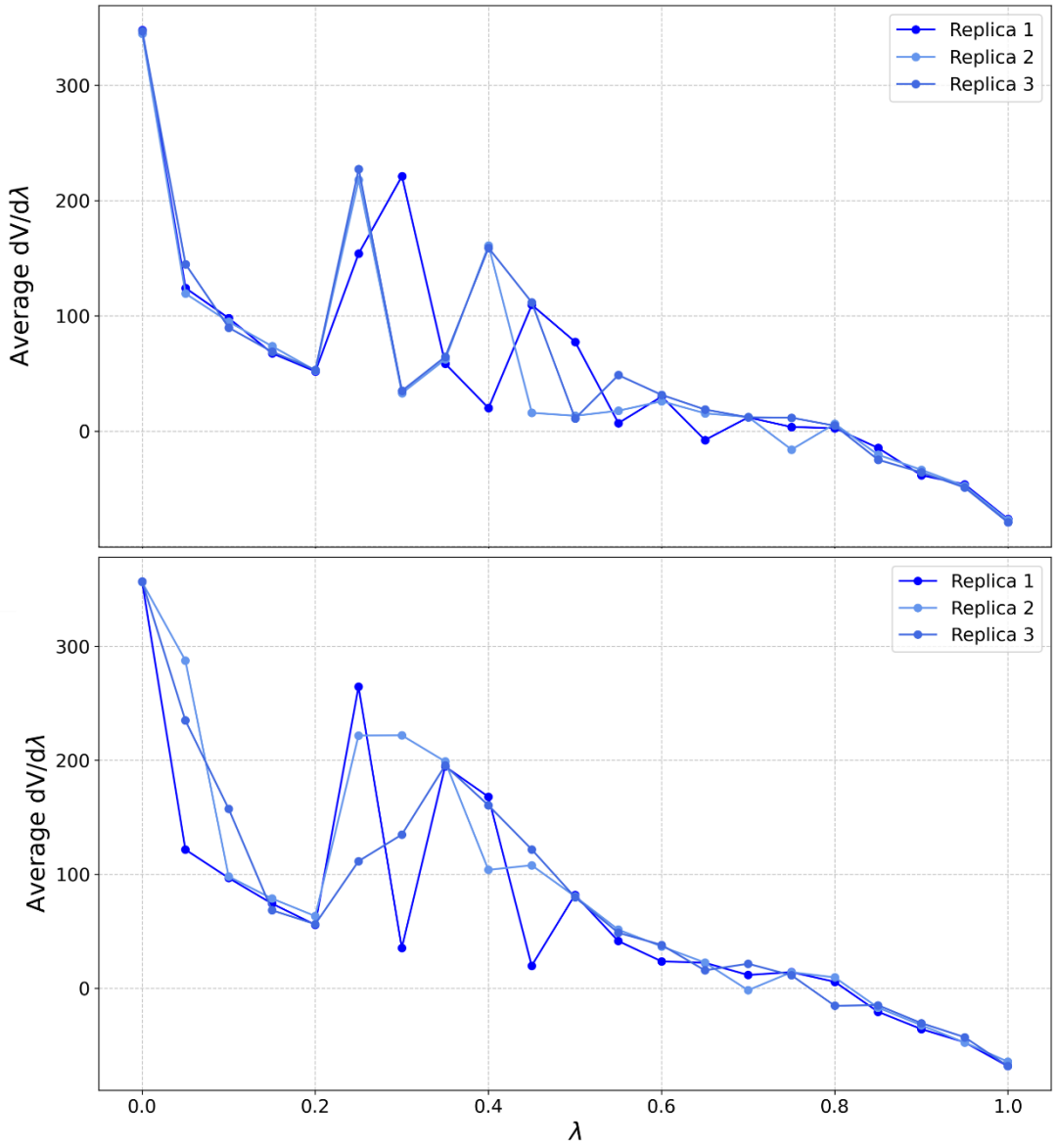

Figure S10 Potential energy curves for the TI simulation involving wild type BRC1 and K31T BRC1, using the default α value. A) Peptide in water B) RAD51-BRC complex

Acknowledgements

This work wouldn’t have been possible without the vital contribution of my superiors and colleagues, to whom I am deeply thankful; both the team in Bologna, Professor Matteo Masetti, Mattia Bernetti, Francesco Rinaldi, Stefano Bosio, and in Munich, Professor Martin Zacharias and Patrick Quoika. My gratitude also goes out to those that, albeit not directly involved in this project, always offered their support, Riccardo Ocello, Professor Federico Falchi, Riccardo Aguti, Luca Menestrina Dominga Evangelista, Simone Romer. Thanks to my partner, my family, and my friends, for always being there.

:wq