Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE BIOTECNOLOGICHE, BIOCOMPUTAZIONALI, FARMACEUTICHE E FARMACOLOGICHE

Ciclo XXXIV

Settore Concorsuale: 03/DI

Settore Scientifico Disciplinare: CHIM/08

Structural and biophysical characterization of RAD52 as novel pharmacological target for synthetic lethality

Presentata da: Beatrice Balboni

Coordinatore Dottorato
Prof. Maria Laura Bolognesi

Supervisore
Prof. Andrea Cavalli

Co-supervisore
Dott.ssa Stefania Girotto

Esame finale anno 2022

Ad maiora

Table of Contents

Abstract

RAD52 is a ring-shaped oligomeric protein, which binds both DNA and RNA and it is involved in various DNA repair mechanisms. In the last few years, RAD52 has been proposed as a novel pharmacological target for cancer synthetic lethality strategies. Hence, this work has the purpose to investigate RAD52 protein with biophysical and structural tools to shed light on proteins features and mechanistic details that are, up to now, poorly described, and to design and present novel strategies for its inhibition.

My PhD project had two main goals: the structural and functional characterization of RAD52 and the identification and characterization of novel RAD52 inhibitors.
Recombinant RAD52 Full-Length (FL) and RAD52 [1-212] forms were expressed in Rosetta pLysS *E.Coli* and purified. The oligomerization state of both proteins was characterized together with their propensity to form high MW superstructures. Moreover, RAD52-DNA interaction was also investigated and characterized through fluorescence-based biophysical techniques. Since no 3D structure is available for full-length RAD52, cryo-EM experiments were performed on this RAD52 form to gain additional structural information valuable both for understanding the mechanism of protein action and for drug discovery purposes.
The second part of my PhD project focused on the design and characterization of novel RAD52 inhibitors to be potentially used in combination therapies with PARPi to achieve cancer cells synthetic lethality. This approach would aim at increasing sensitivity to PARPi treatments while avoiding resistance occurrence side effects. With this aim we selected and characterized promising RAD52 inhibitors through three different approaches: ^{19}F NMR fragment-based screening; virtual screening campaign; aptamers computational design. Selected hits (fragments, molecules and aptamers) were investigated for their binding to RAD52 and for their mechanism of inhibition. Collected data highlighted the identification of promising hits worthy to be developed into more potent and selective RAD52 inhibitors.

In the last part of the thesis a second project carried out during my PhD is reported. GSK-3β protein, an already validated pharmacological target is here investigated using biophysical and structural biology tools.

In this work we proposed a drug discovery screening pipeline able to directly select compounds with binding affinities not higher than a reference binder compound (here, AMP-PNP). This approach, that can be generalized to the search of potent and selective inhibitors for any ATP dependent enzyme, has been here validated by the identification of promising fragments, inhibitors of GSK-3β, worthy of being developed into more potent compounds.

1. Introduction

1.1. DNA damage: double strand break (DSB) repair

Genome integrity is constantly affected by DNA damage and replication errors due to exogenous and endogenous sources. Significant numbers of cells in the human body are constantly subjected to DNA damage (10 000 to 1 000 000 cellular lesions per cell per day). Unrepaired or misrepaired DNA can lead to genome aberrations and mutations that affect cellular functions[1,2]. Among these types of lesions, DNA double-strand breaks (DSBs) are cytotoxic lesions that pose immediate threats to genomic integrity and are the result of chromosome breakage, dysfunctional replication fork processing or telomere deprotection. In many hereditary human diseases, germline mutations in DSB repair-associated genes, such as cancer development- and aging-associated genes, cause failure to repair DSB and consequent genomic instability. Misrepair of DSBs can lead to inappropriate end-joining events, which commonly underlies oncogenic transformation due to chromosomal translocations[3,4].

For this reason, eukaryotes have many different robust and redundant DNA DSB repair mechanisms, which can be exploited depending on cell conditions and health state; they can be both accurate and error-prone. Two main DNA DSB repair mechanisms in cells are commonly reported: Homologous Recombination (HR) and Canonical Non-Homologous end joining (C-NHEJ). Other secondary error-prone DNA repair mechanisms are alternative – end joining (ALT-EJ) and single strand annealing (SSA). The complete classification is illustrated in Fig. 1 and 2. The choice of DSB repair pathway depends on the cell cycle phase in which the DNA break occurs and the characteristics of the DNA break site end resections.

1.1.1. Homologous Recombination (HR)

Homologous recombination (HR) is an error-free high fidelity DSB repair process that uses a homologous or sister chromatid as a template to repair DNA, preventing loss of genetic information. This DNA repair mechanism is exploited mainly in S and G2 cell cycle phases when the sister chromatid is present and when the genetic material is more abundant. Interestingly, DNA ends resection, the first step of HR, is regulated by many proteins opposing each other to influence the choice of the DNA repair mechanism. For instance, 53BP1 and BRCA1 proteins counteract each other to inhibit or promote DNA end resection, respectively, during S/G2 phases[3,5].

After these regulatory early steps of the process, the DNA end resection is initiated by the trimeric complex MRN, which acts as a DNA damage sensor, and by the C-terminal binding protein (CtBP) interacting protein (CtIP)[6]. The MRN complex starts making a nick up to 300 bp upstream of the 5' strands that will be resected. MRN then exploits its 3'-5' nuclease activity to produce the 3' single-stranded overhangs at the DNA damage site[4,7]. Immediately, to avoid nuclease effects, strand degradations and DNA secondary structure formations, replication protein A (RPA) is recruited and bound on the 3'-overhangs[8]. At the same time, if required, further resection of the strands are performed by 5'-3' nuclease activity of exonuclease 1 (EXO1) or by the nuclease/helicase activity of Bloom Syndrome RecQ Like Helicase/DNA replication helicase nuclease 2 (BML/DNA2)[7]. The process then continues with the replacement of RPA with RAD51 recombinase. RAD51 loading on DNA strands overhangs is performed by BRCA2[9], which in turn is recruited to the DNA damage site by the BRCA1-PALB2 complex[10].

At this point, RAD51 protein, assisted by other mediators, such as RAD54 and PALB2, promotes 3'ssDNA strand invasion into a homologous sister chromatid to form a D-loop (displacement loop), a three-stranded-DNA bubble structure, in order to pair and anneal broken DNA strand with the identical sister chromatid template[11]. After D-loop formation and RAD51 filaments stabilization (through BCDX2 and CX3 complexes, for instance[12]), the DNA invading strand is extended by DNA polymerase σ (or other translesion polimerases[13–15]), proliferating cell nuclear antigen (PCNA) and replication factor C (RFC)[16].

The following step is the actual processing of the D-loop, which can be exploited in three different manners: synthesis dependent strand annealing (SDSA), double strand break repair (DSBR) and break-induced replication (BIR). SDSA takes place when RAD51 mediates the D-loop invasion of only one of the two resected ends of the DSB, whereas the second one remains passive and facilitates HR process annealing with the displaced strand. SDSA does not lead to Holliday Junction (HJ) formations and is defined as a non-crossover pathway[4]. By contrast, in the DSBR, the second end of DSB is not passive and is engaged to stabilize the D-loop structure forming a double HJ (dHJ), which can be solved both producing crossover and non-crossover products. This last resolution commonly occurs during the meiotic recombination. The third mechanism for HR is BIR and concerns real replicative responses during strand invasion. This pathway will be further discussed below.

1.1.2. Canonical Non-Homologous End Joining (C-NHEJ)

C-NHEJ occurs throughout the cell cycle but especially during G0/G1 and G2 phases[17]. In this mechanism, the DSB is repaired by blunt end ligation regardless of sequence homology. Despite its mutagenicity risk, it protects cell from gross chromosomal rearrangements and maintains

genome integrity[18]. The process begins with the KU protein (heterodimer KU70-KU80 in eukaryotes) initial recognition and binding to the DSB in a sequence independent manner. The KU-DNA complex nucleates and interacts with other factors, such as DNA- dependent protein kinase catalytic subunit (DNA- PKcs), DNA ligase IV (LIG4) and the associated scaffolding factors XRCC4, XRCC4-like factor (XLF) and paralogue of XRCC4 and XLF (PAXX)[19–22]. To complete the repair, a two-step process is required[23]: Ku70-80 and DNA-PKcs initially create a long synapses and then, through XRCC4-LIG4 and DNA-PKcs, the two ends flanking the break are aligned. Specifically, after making DNA ends eligible for ligation process through Artemis endonuclease or error-prone polymerase Polμ/Polλ, XRCC4 and XLF (XRCC4-like factor) form a sleeve-like structure around both ends of the DSB to stabilize the DNA and stimulate DNA sealing by LIG4. PAXX also interacts with Ku70/80, stabilizes the NHEJ-complex assembly and promotes the ligation of DNA overhangs[24]. Many are the accessory factors that support and regulate this mechanism and some of these are MRE11-RAD50-NBS1 (MRN) end recognition complex that assist end-bridging[25] and aprataxin and PNK- like factor (APLF), which interacts with Ku80 and with poly(ADP- ribose)-modified proteins near DSB[26].

1.1.3. Alternative End Joining (Alt-EJ)

ALT-EJ, also called microhomology-mediated end joining (MMEJ), is an error-prone DSB repair pathway, which leads to insertion and deletions (indel mutations) and chromosomal traslocations[24,27]. This mechanism exploits short sequences of microhomology (2-4 nts) to join DSBs even if evidence demonstrates that Alt-EJ does not totally rely on microhomology, since suitable Alt-EJ substrate are DSB ends even bearing only one homologous nucleotide or no homology at all[28]. This mechanism relies on different mediators, such as MRN-CtIP, Poly

(ADPribose) polymerase 1 (PARP1), DNA polymerase theta (Polθ) and ligase 1/3[29–31]. Specifically, MRN complex exonuclease activity is required for the 3'-overhang formation and, after that, PARP1 localizes at the DSB site and recruits Polθ through a poorly understood mechanism. After its recruitment, Polθ through its helicase activity displaces RPA from the 3'- single strand overhang and uses available microhomology sequences to form synapses, later stabilized by its DNA synthesis activity extending the 3'- minimal pairing. After synapsis formation and stabilization, DNA ligase 1/3 seal the DNA after removing non-annealed DNA flapping ends[32].

1.1.4. Single Strand Annealing (SSA)

Single Strand Annealing (SSA) is a RAD51-independent DSB repair pathway that joins two homologous 3' ssDNA ends in highly repeated regions, such as in tandem repeats, through annealing, deleting regions between sequences repeats. SSA is therefore an error- prone pathway[33,34]. Briefly, this mechanism starts with the DNA ends resections at the DSB site, generating 3'-ssDNA ends. After the end resection, the annealing of the flanking repeats and the following removal of the non-homologous 3'-ssDNA tails occur, mediated by RAD52 and ERCC1/XPF proteins, respectively. Notably, RAD52 is the main key factor of the whole mechanism, mediating not only the strand annealing, but also enhancing the nuclease activity of ERCC1/XPF complex[35]. Finally, the process is completed through the gap filling activity of DNA polymerases and ligase activity. This RAD52-mediated process will be further described in the following sections.

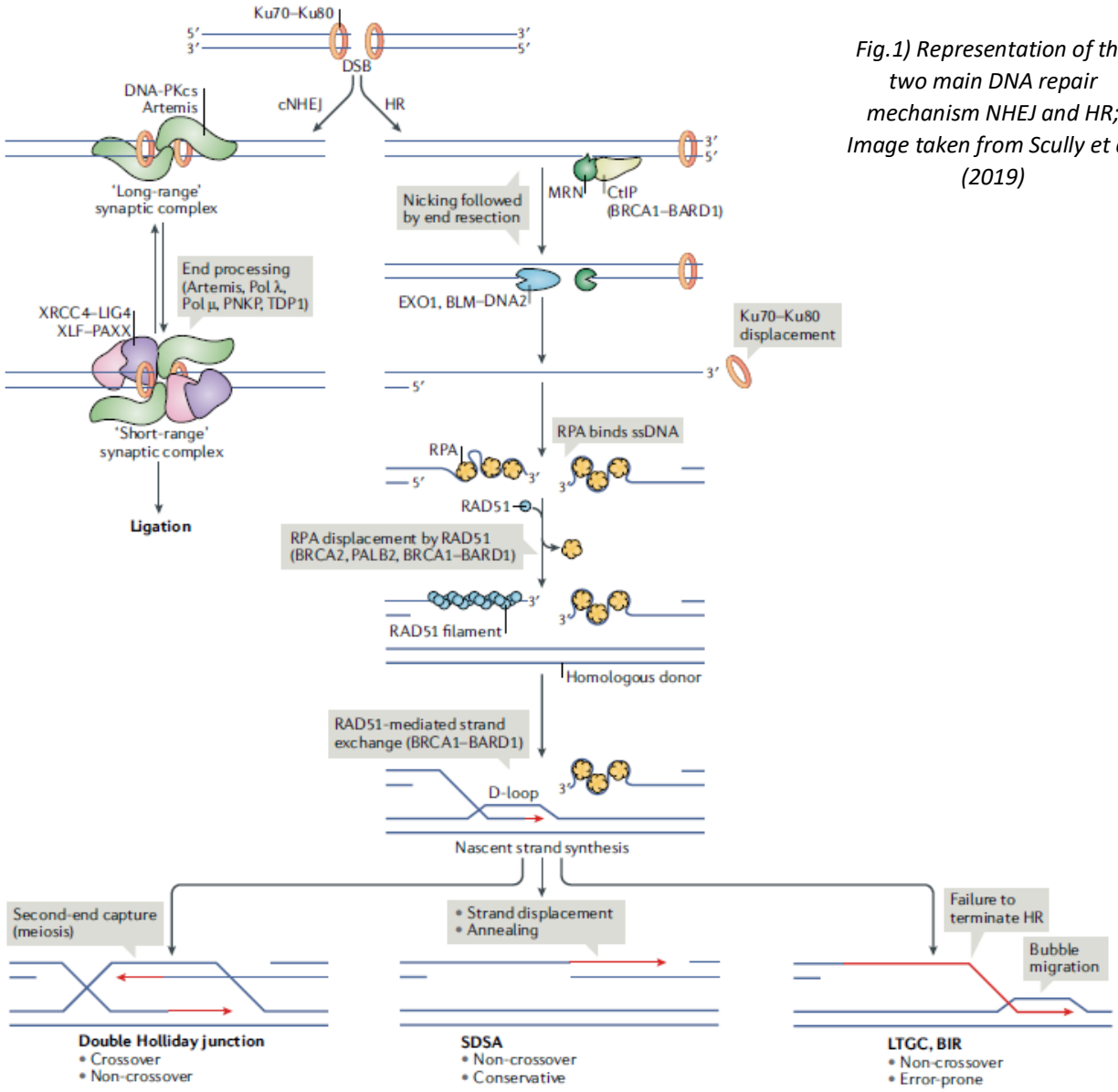


Fig.1) Representation of the two main DNA repair mechanism NHEJ and HR; Image taken from Scully et al. (2019)

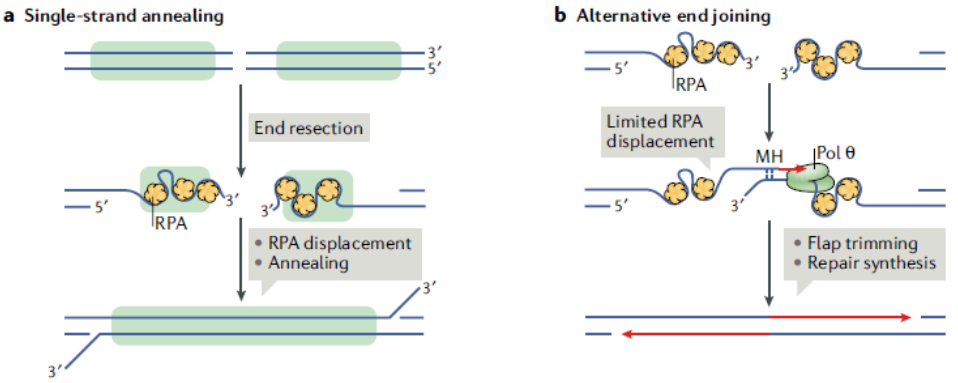


Fig.2) Representation of different DNA alternative repair mechanisms single stran annealing and alternative end joining; Image taken from Scully et al. (2019)

1.2. Synthetic Lethality

The synthetic lethality paradigm was first described in 1922 in *Drosophila Melanogaster* organisms by the geneticist Calvin Bridges and it was later further explained and named by Theodosius Dobzhansky in 1946[36,37]. This phenomenon arises when the simultaneous inactivation or depletion of two (or more) genes leads to cell death (synthetic lethality) or sickness (synthetic sickness), while a defect in either one of the genes has a little or no effect on cell viability[38]. This phenomenon has been studied for long time and after the groundbreaking works of Hartwell and Kaelin became also a fundamental tool for the development of innovative cancer-specific targeted therapies[39,40]. Indeed, they first proposed the synthetic lethality to be exploited in cancer therapies where a specific cancer-mutation related gene was already present. The pharmacological inhibition of a target protein, as a second gene product, would turn out in the simultaneous inactivation of two pathways with consequent lethality of tumor cells, leaving nonmalignant cells mainly unaffected[38,41].

Synthetic lethality is therefore a very innovative and potent tool to be used in therapies to selectively target malignant cells, potentially avoiding chemotherapeutic side effects and resistance mechanisms. Furthermore, synthetic lethality strategies may have other important advantages, such as the simple identification of patients who respond to treatments, due to its selectivity on specific cancer-related mutations and the increase in the efficacy of the chemotherapeutic drugs, leading to lower dosage use and avoiding adverse effects[42].

TARGETING "BRCAness": PARPi and beyond

The first FDA approved synthetic lethality therapy was the clinical use of poly (ADP-ribose) polymerase (PARP) inhibitors rucaparib and niraparib in patients with BRCA-mutated ovarian

cancer and olaparib for BRCA-mutated ovarian and breast cancer[43–45]. Briefly, tumors, such as ovarian or breast cancers, that can arise in individuals with BRCA germline mutations, are frequently associated with a loss-of-function aberration in the wild-type BRCA allele, leading to a defective homologous recombination (HR) mechanism for DNA double strand break repair. PARP is a mediator of the single strand break (SSB) repair pathway. Inhibition of PARP in these pathological conditions leads to unrepaired SSBs that are later converted to double strand breaks (DSB). In this situation, BRCA2-deficient cancers cells are not able to cope with DSB repair. PARPi inhibit also the Alt-EJ[46,47] and induce a collapse in the replication fork maybe due to PARP trapping on the DNA or to PARP inhibition itself. The accumulation of DSBs or replication fork blocks lead to cell death[48].

Unfortunately, resistance to PARPi-related synthetic lethality therapies is a reported common issue, often arising from the genetic reversion of BRCA1/2 mutation, leading to restoration of HR DNA repair mechanism[49–51]. Therefore, a huge effort has been made in finding novel synthetic lethality strategies to target "BRCA-ness" tumors, overtaking PARPi inhibitors resistance or boosting the potency of the therapies when used in combination with other chemotherapeutical agents. The most recent insights into novel synthetic lethal targets identification are nicely reported by Topatana and Li in their reviews[42,52]

One of the most recently studied synthetic lethal target in BRCA-deficient cancers is RAD52. This protein seems to play several roles in DNA alternative repair mechanisms even though it was proved not to be essential for cell viability in vertebrates[53]. However, in recent studies inhibition of RAD52 has emerged to be synthetically lethal in the presence of loss-of-function mutations in BRCA2, BRCA1 or PALB2[54–56]. RAD52 inhibition has been characterized in combination with PARPi

on BRCA-deficient tumors and it showed a significant enhancement of anticancer treatment effects[57]. This target and its roles in therapies will be further discussed in the next chapter.

1.3. RAD52

RAD52 is a DNA/RNA-binding protein that plays a multifaceted role in many intracellular pathways related to DNA-repair and maintenance of genomic stability[53,58]. RAD52 is a protein present in prokaryotes and eukaryotes species, either unicellular or multi-cellular[56,59,60]. However, it does not always play the same role among the various species. In this thesis the description and study of human RAD52 protein only will be discussed. This chapter will focus on RAD52 main structural and functional hallmarks, explaining the state-of-the-art on its main functions and its involvement in the development of various diseases.

1.3.1. RAD52 structure

Human RAD52 is a 47 kDa protein of 418 amino acids that forms multimeric-ring shaped functional units. Its N-terminal domain (1-177 AA) comprises the oligomerization domain and the DNA binding domain and shares a high homology with yeast Rad52 (>70% sequence homology[61,62]) whereas its C-terminal domain contains RAD51 binding site and Replication Protein A (RPA)-binding site and has no sequence homology with yeast Rad52 (Fig. 3)[61,63].

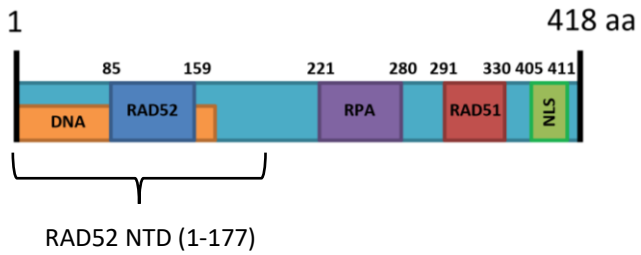


Fig. 3) Domain map of human RAD52: N-terminal domain contains the DNA binding region and a self-associating region; the C-terminal domain contains RPA and RAD51 interacting regions and a nuclear localization signal.

Although a high-resolution structure of full-length RAD52 has not yet been solved, low resolution evidences from electron microscopy studies[64–66] suggest that RAD52 forms heptameric rings and has a tendency to form even higher molecular weight (MW) superstructures that interact with other protein functional units in a stacked- or side-by-side- fashion. The tendency to form such high MW superstructures is increased in the presence of DNA[59,61,63,67].

Nevertheless, X-ray crystal structures available for truncated N-terminal portions of RAD52 (i.e. AA 1-212 or 1-209), (PDB 1H2I, 1KN0, 5JRB, 5XRZ, 5XS0), allowed the characterization of the DNA-binding and multimerization domain of the protein[61,68,69]. The crystal structure of RAD52 N-terminal domain shows a ring-shaped undecamer, resembling a mushroom, with a "stem" and a "domed cap" (Fig. 6). The stem part of each monomer has a β-β-β-α structure, in which the upper parts of the β-strands of all the monomers align side-by-side forming the inner part of the channel. The identified interaction among the sheets are hydrogen bondings, specifically connecting the CO of His86 with NH of Glu122, CO of Asp117 with NH of Phe26, NH of His86 with CO of Tyr120. The "domed cap" region, instead, is constituted by amino acids flanking the β-β-β-α structure both at the C- and N-terminal. α-helices 1 and 5 (Fig. 4; Fig. 7) have hydrophobic interaction with the upper part of the β-barrel-like stem, leading to the protrusion of a hairpin loop constituted by β1-L3- β2 fragments (named "lobe"). Instead, the C-terminal part, composed by L9-α4-L10-α5 is bound by hydrophilic interaction to the flanking monomer of RAD52 ring. Specifically, NH and CO of Lys190 in L10 loop interact with CO of Arg46 and NH of Gly48, respectively.

This structure leads to the formation of a negatively charged surface at the top of the flat domed cap and a positive charge at the bottom of the ring, between the stem and the hairpin loop,

representing the first DNA binding site of the protein. The DNA was reported to wrap alongside the outer part of RAD52 undecameric ring, fitting in this positive cleft[61,70]. In this positive region some amino acid residues, critical for DNA binding, are present: Arg55, Lys152 for ssDNA and Tyr65, Arg153, Arg156 for dsDNA/ssDNA[61,71].

Further studies allowed Kagawa and colleagues to identify a second DNA/RNA binding region located close to the entrance of the positive charged surface, comprising Lys102 and Lys133[71]. This new DNA binding site has provided a clearer overview of what the protein's mechanisms of action might be, i.e. mediating and promoting ssDNA annealing, homology search and D-loop formation[61,68,70,71].

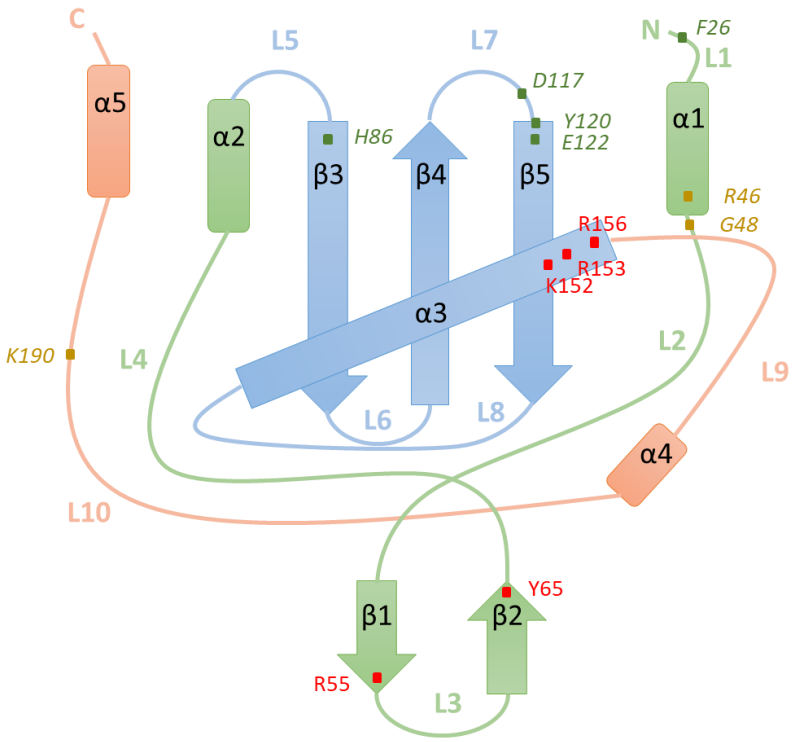


Fig. 4) Diagram showing the fold of the RAD52 [1-212] monomer. Rods and arrows indicate helices and strands, respectively. Amino acids involved in DNA binding are reported in red; amino acids involved in intramolecular interactions are reported in green; amino acids involved in protomer-protomer intermolecular interactions are reported in yellow.

Despite the significant difference in sequence length, both the truncated N-terminal form of RAD52 and full length RAD52 (FL) form ring-shaped oligomers with similar diameters (around 10 nm)[63–65,72]. Actually, the ring-model structure of the full length RAD52 was built by Kagawa and colleagues[61], merging information from the N-terminal structure of RAD52 and rough information obtained from Dynamic Light Scattering (DLS), Scanning Transmission Electron Microscopy (STEM) and analytical ultra-centrifugation (AUC) data on the full length protein[63,64]. The heptameric ring model of RAD52 FL was built on indirect experimental evidence and two main speculations: firstly, if RAD52 [1-212] monomers fit in a heptameric-ring, the distance among β- β- β-α folds would increase by 1 nm; secondly, it was assumed that whether the RAD52 FL monomer-monomer interfaces were composed of β-barrel structures, the difference in the distance among neighboring monomers could fit two β-sheets more compared to RAD52 [1-212] oligomeric structure. Notably, several structure-based prediction techniques allowed to identify possible β-sheets structures in the limited disordered region of the N-terminal portion of RAD52 [1-212] (Val23 to Phe26), Gln221 to Val343 and downstream the residue Ser346[61]. The suggested model is reported in Fig. 5.

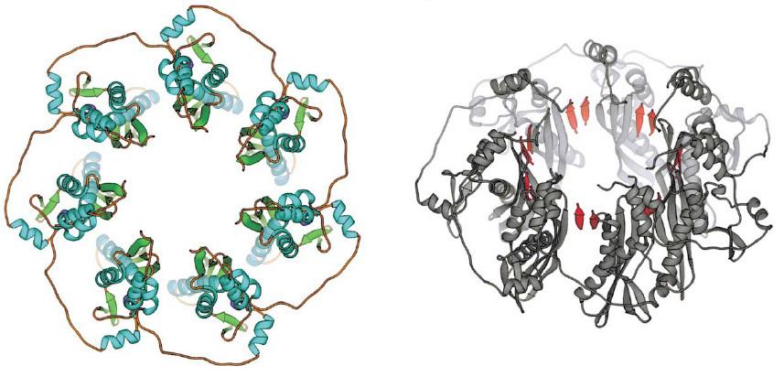


Fig. 5) Model of the RAD52 heptameric ring, viewed from the top and from an angle. The space between the top portions of the sheet in the stem could accommodate two strands (colored in red). Image taken from Kagawa et al. (2001).

Regarding the physiological forms of RAD52, in 1999 Kito and colleagues demonstrated the existence, along with RAD52 FL, of different RAD52 shorter isoforms, carrying the same DNA binding and homologous pairing activities of RAD52 FL[59]. These reported similar activities could be explained by the fact that RAD52 protein isoforms (AA 1-177 of the human RAD52) share 70% of homology at the N-terminal[61]. In agreement with these results, Kagawa and colleagues suggested that the 11-mer ring of truncated N-term RAD52 could represent one of the oligomerization states displayed by RAD52 homologues and shorter isoforms[59].

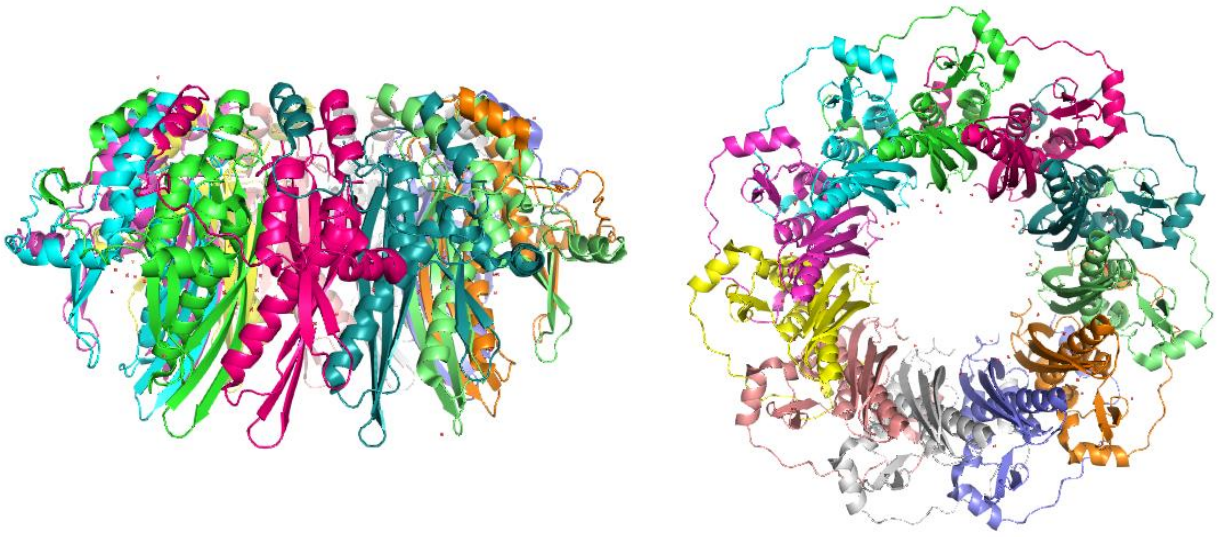


Fig. 6) Side and bottom views of the mushroom-like structure of the undecameric ring of RAD52 [1-212] (PDB 1KN0). Structures were prepared using pyMOL 2.4 software.

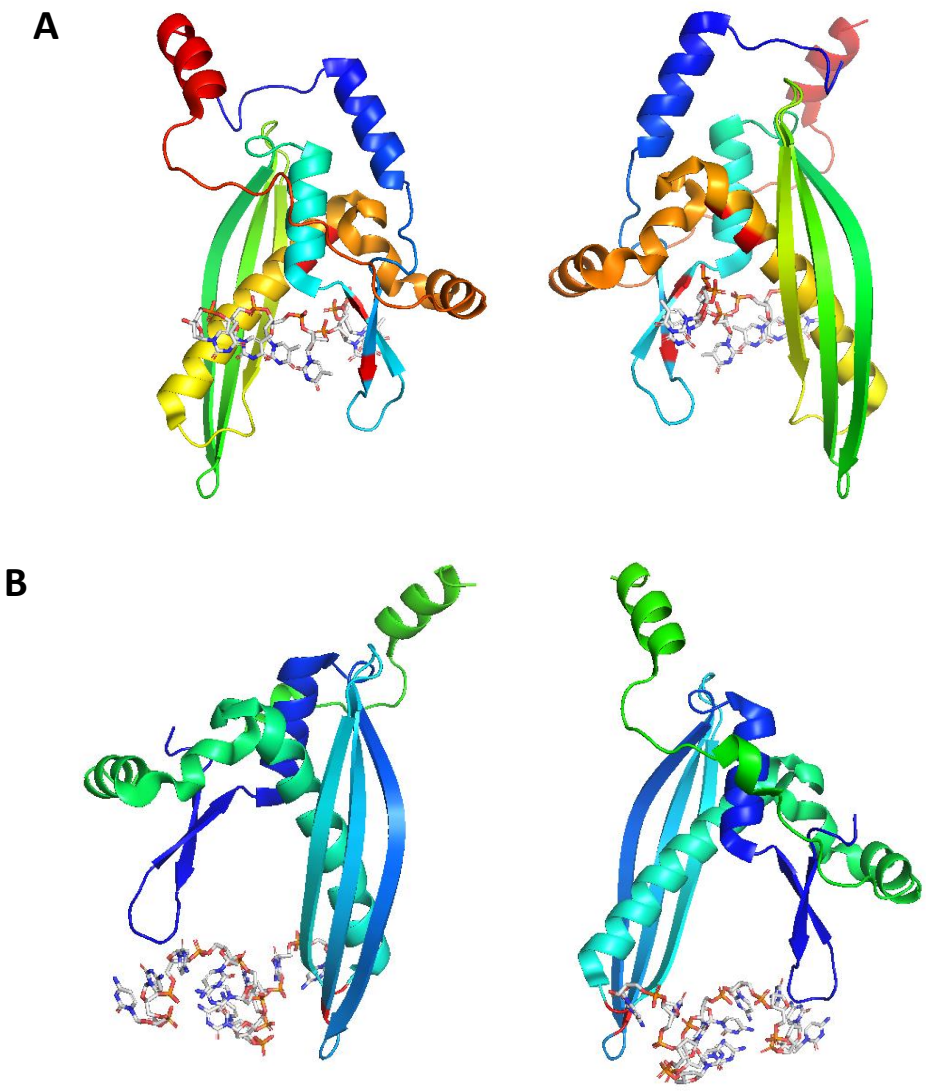


Fig. 7) A) two views of RAD52 [1-212] monomer in complex with a ssDNA molecule inside its inner binding cleft (PDB 5XRZ); B) two views RAD52 [1-212] monomer in complex with a ssDNA molecule inside its outer binding cleft (PDB 5XS0). Structures were prepared using pyMOL 2.4 software.

Both RAD52 FL and RAD52 N-terminal domain were reported to have an elevated thermal stability[67]. This feature is probably linked to their oligomeric state and to their propensity to form higher MW ring complexes. Notably, even though RAD52 N-terminal domain and RAD52 FL have always been reported to have a similar ring structure, the propensity to form high MW

superstructures was reported to be higher in RAD52 FL form compared to RAD52 N-terminal domain[63,65,67,73]. This may be due to the presence of the C-terminal domain portion that should favors intermolecular bonds and hydrophilic interactions[63].

1.3.1.1. RNA/DNA binding

RAD52 exerts its biological function through DNA and RNA binding, prompting DNA strands homology search and annealing and supporting genomic stability control. Many of the pathways in which RAD52 is involved and its full mechanism of action have yet to be elucidated. However, some features of DNA and RNA binding mechanisms of RAD52 have been investigated and several models have been proposed[33,69,74].

The first evidence of DNA binding was reported for yeast Rad52[75,76]. Afterwards, several studies, including structural EM investigations, have reported the characterization of DNA-RAD52 binding also in human RAD52[64–66,70].

Protein-DNA interaction is mediated by the N-terminal domain of RAD52, which is therefore critical for RAD52 DNA-related activity[63,69,77,78] and RAD52 intracellular mechanism of action, at variance with the C-terminal domain, which preserving RPA- and RAD51- binding sites, most likely have only a supporting role in this RAD52 activity.

The groundbreaking development for the characterization of the DNA-RAD52 interaction was made by Saotome and colleagues in 2018, who crystallized RAD52 N-terminal domain in presence of ssDNA[69] (Fig. 7). The crystal structures that was solved corroborated the existence, in RAD52, of two DNA binding sites as previously proposed[70,71,79]. Regarding the inner binding site, the single-stranded DNA wraps around RAD52, fitting inside a positively charged groove. Each protein

monomer should accommodate four nucleotides, with the bases of the base-pairing edges exposed to the solvent, most likely facilitating homology search and annealing to a second single-stranded DNA. Furthermore, protein conformation and oligomerization state appear to be unaffected by DNA binding, meaning that the inner DNA binding groove of RAD52 ring is in a "ready" state for DNA binding. The DNA inside the groove is stabilized both by stacked hydrophilic interactions between DNA bases and Arg55 and Val63 and by electrostatic interactions between the DNA stretched phosphate backbone and the basic amino acids in the DNA binding site (Fig. 8).

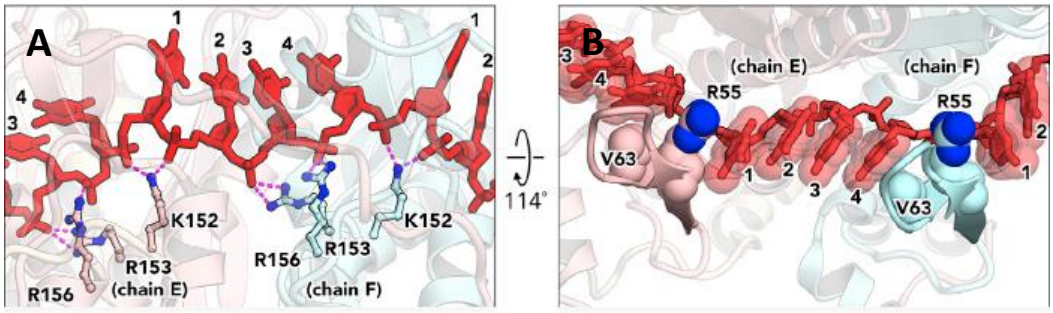


Fig. 8) A) Zoomed view of the electrostatic interactions of K152, R153, and R156 of RAD52 [1-212] monomer with the phosphate backbone of the ssDNA. Dashed lines (magenta) depict hydrogen bonds; B) Hydrophobic stacking interactions that sandwich the four-nucleotide repeats. The b-hairpin structure of RAD52 (amino acid residues 52–66) is located between the four-nucleotide repeats. The ssDNA bases are sandwiched between R55 and V63 of the b-hairpin. Images are taken from Saotome et al. (2018).

Intriguingly, a similar binding mode is also reported for bacterial RecA recombinase, even though this protein does not show sequence homology with RAD52 and it has a totally different oligomerization pattern with respect to RAD52[80]. Hence, this binding model must be a common interaction model among all the ssDNA binding proteins, which support DNA annealing and base pairing[69].

In the second outer DNA binding site, reported in the N-terminal RAD52 crystal structure (PDB 5XS0) [69], the DNA is buried between two different RAD52 ring structures, as a compact right-handed helix, suggesting that the role of the outer DNA binding site is promoting multiple RAD52 ring localizations on a DNA strand mediating DNA strands annealing[71]. The outer DNA binding site has a higher affinity for DNA binding then the inner binding site. Nevertheless, the two binding sites work cooperatively for DNA binding, since, the simultaneous binding of the DNA in the two sites, induces a reduction of the binding affinity constant (K_d) for both sites.

Based on these data, Saotome and colleagues suggested a putative mechanism of action for RAD52 protein: RAD52 may facilitate DNA strands annealing and homology search[69]. In particular, in order to anneal, DNA strands are supposed to first bind to the outer DNA binding site of each RAD52 ring before sliding to the inner DNA binding site. Once DNA strands are in this position, RAD52 rings can get closer, associate one to another and facilitate ssDNA annealing and homology search of the two "trapped" DNA strands. This DNA annealing mechanism was observed also in other single strand DNA binding proteins of lower complexity organisms, such as different types of bacteria[81].

A similar mechanism of action had been previously proposed by Rothenberg and colleagues[33]. Also in this work, DNA strand annealing was suggested to occur by interaction of two or more RAD52 rings, accommodating ssDNA with the bases presented outward; the association of complexes could facilitate pairing of bases and stabilization; at that point, if complementary was present, annealing initiates with 3-4 bases of nucleation length and the two or more nucleoprotein complexes could roll around each other with an energy-favorable duplex formation driven force[33,82].

Notably, also Grimme and colleagues in 2010 studied DNA-RAD52 interaction mechanisms and postulated two possible mechanisms for homology search, cis and trans, respectively reported in Fig. 9. In the cis mechanism, a portion of one DNA strand can come out from the deep inner binding groove and can be placed temporarily in the second DNA binding site of the second interacting nucleoprotein. In the trans mechanism, both DNA strands can be pulled out from the inner binding site and moved up to the secondary binding sites of their respective rings to start the homology search and annealing[79]. Notably, the most effective DNA strands annealing occurs between two RAD52-ssDNA nucleoprotein complexes and not between RAD52-ssDNA and protein free DNA[79].

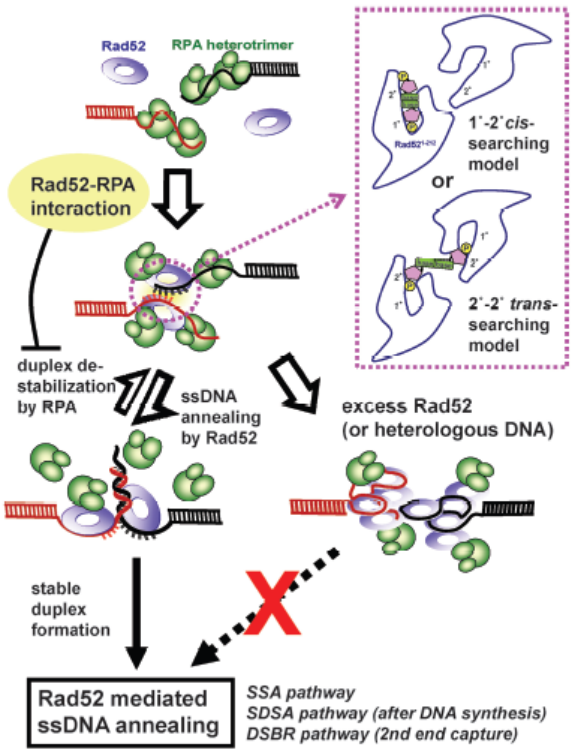


Fig. 9) Model for DNA binding and annealing by RAD52. RAD52 can bind the ssDNA–RPA complex and forms RAD52–RPA–ssDNA ternary complex in wrapped configuration. The two possible mechanisms for homology search between RAD52–RPA–ssDNA complexes (cis and trans) are reported in the purple box; figure taken from Grimme et al. (2010).

In the same work, Grimme and colleagues also suggested that RAD52 takes part in the DNA recombination activity facilitating RPA protein releasing from the DNA filament[79]. Specifically, thanks to its ability to form complexes with RPA, RAD52 can remove RPA from DNA and facilitate RAD51 loading on the DSB site[79]. This study pointed out the critical role played in the DNA strand annealing by the C-terminal domain of RAD52 in regulating the RAD52-RAD51-RPA interaction on DSB sites. RAD52 role in RPA turnover on DSB sites was confirmed also by Ma and colleagues in 2017[83].

As reported by Kagawa and colleagues in 2001 and 2008, RAD52 binds dsDNA, promoting D-loop formation. This evidence, later corroborated by other studies, highlights the importance of both RAD52 DNA binding sites for protein activity in homology search and strand invasion[70,71,84,85].

Interestingly, recent studies report RAD52 ability to bind not only DNA but also RNA[86,87]. ssRNA and ssDNA show the highest binding affinity for RAD52 whereas a significant lower affinity is observed for the double-stranded substrates[84,86]. However, among the double-stranded substrates, RNA-DNA hybrids were reported to have a more efficient binding to RAD52 compared to dsRNA and dsDNA. Finally, RAD52 has a higher affinity for R-loop structures compared to hybrid structures. These data are in line with the key role played by RAD52 in specific DNA repair mechanisms namely RNA-template recombination repair, which will be further discussed in the next chapter.

1.3.1.2. Post Translational modifications

In the most recent years, many have been the studies and hypotheses about the post-translational modifications that RAD52 would require to perform its function[53].

RAD52 acetylation is reported to be a critical modification of the protein that regulates its function[88,89]. Specifically, it has been shown that non-acetylated RAD52 can accumulate at DSB sites, where recruited, but it dissociates prematurely. In the absence of RAD52 acetylation, also RAD51, which plays a central role in HR, dissociates prematurely from DSB sites causing an HR impairment[88]. Moreover, SIRT1-SIRT2 deacetylases depletion induces effects equivalent to RAD52 depletion, but without affecting SSA and NHEJ repairs[89]. Also the recruitment of RAD51 to DSB sites is affected by SIRT2 or SIRT3 depletion, but not by RAD52 deacetylation[89]. These preliminary studies suggest that acetylation and deacetylation of RAD52 may be a regulatory mechanism controlling the protein–protein interactions between RAD52 and HR-related proteins in multiple HR steps. Nevertheless, further investigations are required to clarify these mechanisms.

Based on sequence homology with yeast RAD52, human RAD52 protein was suggested to undergo sumoylation modification which, however, should not affect protein-protein interactions and may only affect DNA-binding and strand annhealing[53,90]. Nevertheless, RAD52 sumoylation site was identified in its Nuclear Localization Domain (NTD) at the C-terminal of the protein, suggesting that sumoylation could play an important role in nuclear transport of RAD52[91].

Finally, phosphorylation of RAD52 at Tyr104 was reported to enhance ssDNA annealing activity while lowering dsDNA binding ability of RAD52. Additional studies, performed on constitutively active oncogenic BCR-ABL1 kinase, demonstrated that RAD52 phosphorylation facilitates its nuclear localization and stimulates SSA repair in leukemia cells[92–94]. However, phosphorylation of Tyr104 it is not strictly required for RAD52 to exert its DNA-binding activity[55].

1.3.2. RAD52 function

RAD52 protein in the last few years have been the focus of a plethora of investigations since many evidences report RAD52 as novel key mediator of DNA repair mechanisms for genomic stability and cell health maintenance. Nevertheless, detailed information on its cellular role and on its involvement in all the pathways in which it may be involved are still missing[84]. In this chapter a state-of-the-art description of RAD52 roles in DNA repair mechanisms is reported.

1.3.2.1. Homologous Recombination (HR)

In yeast, Rad52 is a key mediator of HR-based DNA repair mechanism. Rad52 has a key role in recruiting RAD51 on DSB sites and in promoting annealing of ssDNA complexed with RPA, thus facilitating RAD51 recombinase activity[95–97]. On the contrary, RAD52 in vertebrates appears to be only an auxiliary redundant factor within RAD51-dependent HR pathway, where its inactivation does not induce any significant cellular impairment. *In vivo* experiments reported that RAD52-null mice are viable with no evident phenotype[98], while experiments performed on DT40 chicken-B cells (DT40) showed that inactivation of RAD52 still do not affect viability and cell health, with a RAD52-depleted phenotype comparable to wild type phenotype. Specifically, RAD52 depletion in DT40 cells led only to a slight reduction in targeted chromosomal integration, without a detectable increase in radiation sensitivity, which would be expected whether RAD52 had an essential role in HR mechanism as in yeast[99]. Despite the apparent dispensability of RAD52 in higher organisms, the attention was drawn on RAD52 function in humans when Feng and Lok groups showed that BRCA2-deficient cells required RAD52 for survival. Indeed, they first highlighted that RAD52 may have an important role in replacing HR-related mediators role, acting as a backup in pathological conditions; therefore, RAD52 depletion in “BRCA-less” cells resulted

in severe phenotype effects, for instance in RAD51 impaired foci formation and genomic instability[54,56]. In a BRCA-deficient context, it has been shown that RAD51 loading can be carried out by RAD52, albeit at a lower efficiency, suggesting a backup BRCA2-independent RAD52 role as mediator for RAD51 activity in HR [54,56]. This HR-related RAD52 backup role in BRCA-deficient cells was also supported by a recent study by Mahajan and colleagues[100]. They reported, consistently with previous observations, that in BRCA2-deficient cells, RAD52 overexpression rescues the excessive origin firing and checkpoint control defects observed in BRCA2-deficient cells, compensating for BRCA2 loss. Moreover, RAD52 was proved to interact with pCHK1, acting as BRCA2 for HR regulation and maintaining checkpoint control during DNA damage response[100].

Nevertheless, RAD52 has also a role in the second end capture of the HR mechanism[101]. This HR step takes place right after the D-loop formation at the DSB site, upon completion of the DNA synthesis. The second end of the DSB, leading to the formation of HJs intermediates in the HR process, is captured via RAD52 annealing to the D-loop. Similar data, showing in involvement of RAD52 in the second end capture of the HR have been also reported in yeast[102].

Despite the clear involvement of RAD52 in the HR pathway, many are the missing information that would be necessary to achieve a comprehensive description of RAD52 role within this pathway.

1.3.2.2. Single Strand Annealing (SSA)

RAD52 is the key mediator of the SSA DNA repair mechanism. This is RAD52 most accredited and studied function. SSA is an error-prone mechanism used when DSBs occur in highly repeated DNA regions[34,103]. SSA relies on long homology tracts to anneal two 3'-ssDNA overhangs together. It

does therefore not require neither a template donor from a sister chromatid nor a strand invasion as for RAD51-dependent HR[58].

The process (reported in Fig. 10) starts with DNA ends resection and 3' overhangs creation by nuclease activity of MRN and CtIP complexes, as already mentioned for HR[6,103,104]. Additionally, other helicases and nucleases can be recruited for generating longer single-stretches (for instance DNA2, BLM, EXO1 and WRN RecQ Like Helicase (WRN))[3]. RAD52 protein is then recruited on the resected DNA ends to promote their annealing and to recognize the ssDNA region of homology (<30 bps). Following the annealing on the homology regions, ERCC1-XPF complex binds the N-terminal domain of RAD52 and cleaves non-homologous 3'-ssDNA flap ends[35,105]. RAD52 seems to stimulate ERCC1-XPF nuclease activity[35]. Any gap is then filled by DNA polymerases and DNA strands are blended by DNA ligase I to complete the process of SSA. Not all the polymerases and ligases involved in the SSA process have yet been identified and characterized [34].

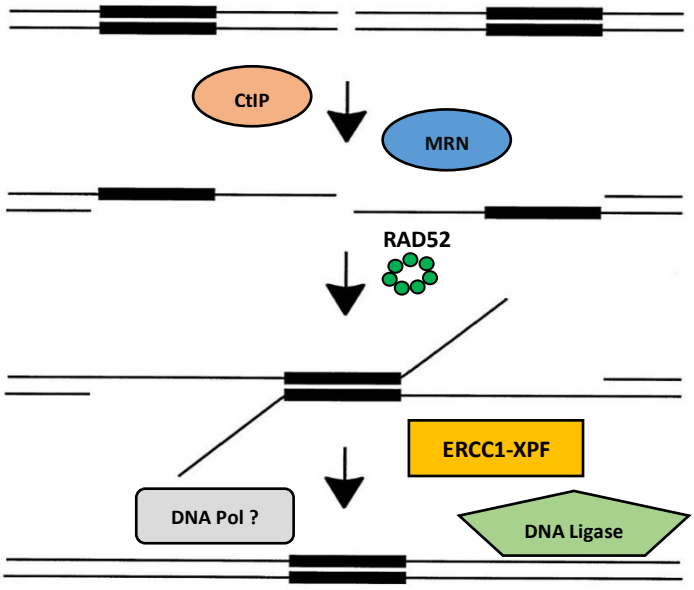


Fig. 10) Schematic representation of RAD52-mediated SSA DNA repair mechanism. RAD52 facilitates homology search and strands annealing. Figure adapted from Sugawara et al. (2020).

Interestingly, novel insights have demonstrated that RAD52 SSA activity is enhanced by its direct binding to the very acidic protein Deleted in Split hand/Split foot 1 (DSS1)[106]. Such interaction is believed to change the conformation of RAD52 and to modulate its binding to DNA, inducing a four-fold increase in the efficacy of the SSA reaction, mainly due to the higher annealing rate of RAD52-ssDNA. The DSS1-RAD52 complex formation is also important for Break Induced replication (BIR), described below.

1.3.2.3. Stalled Replication Forks: protection and processing

During DNA replications, many are the DNA damages that cells can encounter. Such damages can take place at replication fork site, leading to a stall and, if prolonged, to a fork collapse. Cells have therefore developed a number of mechanisms to recover DNA lesions that stall DNA replication forks[107,108]. The recovery mechanisms can be generally divided in damage bypass, fork reversal and fork breakage[108]. In this complex landscape, RAD52 seems to have important roles in solving these types of stress-replication structures both upstream and downstream the fork replication remodeling[108].

Fork reversal event allows the cell to bypass DNA damage by incorporating the correct nucleotides using the newly synthetized sister strand as template instead of the lesion containing strand (namely "chicken foot" formation). This process involves many players, such as ssDNA binding proteins and recombinases (RPA, RAD52, BRCA2, RAD51, RADX), traslocases (SMARCAL1, ZRANB3, HLTF, SHPRH, WRN, RECQ1, ATAD5) and exo/endonucleases (MRE11, EXO1, DNA2, MUS81)[58]. All these proteins are the characters of a highly regulated interplay and counterbalance their effects with each other. RAD52 here is involved in fork protection upstream of the actual reversal mechanism, acting as a "gatekeeper" for replicative fork state. This avoids

unscheduled MRE11-mediated degradation and above all, ensures that fork reversal enzymes load only when required[109]. RAD52 depletion or inhibition results in excessive loading of RAD51, SMARCAL1 and ZRANB3 at stalled replication forks, leading to unscheduled fork reversal and MRE11 dependent degradation causing genome instability. Intriguingly, RAD52 was also reported to recruit MRE11-MUS81 on unprotected reversed fork when in pathological conditions of BRCA2 deficiency[109,110]. Moreover, in CHK1-deficient cells where the G2/M cell cycle check point is lost, cell survival was shown to be dependent on RAD52 and MUS81 to face replication stress[111]. In addition to these kinds of mechanisms, RAD52 takes part also in the fork breakage mechanism for fork stall resolution. Here, one of the fork arms is detached leaving a one-ended DSB. Such structures can be than recovered by HR and RAD52-mediated Break Induced Replication (BIR) and/or Mitotic DNA Synthesis (MiDAS)[58,108].

BIR: At collapsed replication forks, RAD52 can activate BIR, a specialized pathway that repairs single-ended DSBs, well characterized in yeast systems[112]. Briefly, after fork cleavage, at one-ended of the DSB, the end is resected and Rad52 protein facilitates Rad51 filament formation on ssDNA. This nucleoprotein formation invades then the homologous region of the interacting sister strand forming a D-loop. The replisome then assembles with Pol32 (POLD3/POLD4 in human[113,114]) and the DNA synthesis starts. Studies in mammalian cells report that RAD52 is required to facilitate DNA strand invasion to form a D-loop and to anneal DNA strands after recruitment on collapsed fork of MUS81 and EEPD1 nucleases[58,115]. The BIR hallmark is the complementary movement of the D-loop along with the replication fork during the DNA synthesis[114].

MiDAS: MiDAS is a microhomology mediated BIR that is commonly reported when in the presence of Common Fragile Sites (CFS)[116] and other difficult-to-replicate regions, in order to complete DNA replication before cell division. These are usually AT-rich long coding regions in which transcribing RNA polymerases often collide with replicating DNA polimerase[117]. At colliding polymerases, the fork stalls and FANCD2/FANCI complex binds and tethers the sister chromatids. Here, RAD52 is thought to help a microhomology mediated annealing of DNA strands. The intermediate DNA structure is then processed by MUS81-EME1-SLX4 and other nucleases and the DNA synthesis is performed through POLD3 as for BIR[116]. Moreover, RAD52 is suggested to help the recruitment of MUS81 and POLD3 to CSFs in the early mitosis[116]. Interestingly, if MiDAS fails to repair the DNA damage before cell division, the daughter cells inherit damaged DNA that, during G1 phase is sequestered by 53BP1 bodies[118]. These formations, can, in the following S phase, be dissolved by RIF1 activity, triggering RAD52 recruitment and leading to a second DNA repair mechanism through a BIR-equivalent pathway[119].

As already mentioned, BIR activity of RAD52 can be stimulated by DSS1 protein[106].

BIR-associated RAD52 activity was also recently reported in Alternative Lengthening of Telomeres (ALT). Briefly, Zhang and colleagues, demonstrated that RAD52 can directly promote D-loop formation *in vitro* and maintains telomeres length in ALT-associated PML bodies (APBs) in ALT-proficient cells[120].

1.3.2.4. RNA-dependent DNA repair

Even though the HR mechanism for DNA repair was shown to be mostly active in G2/S phase, a HR sub-pathway that uses RNA transcripts as an alternative template for DSB repair was reported to be active during G0 and G1 phases, even though in absence of reverse transcriptases[53,84]. As this mechanism does not require a sister chromatid for DNA-template sequence, RNA-dependent DNA repair occurs also in non-dividing cells such as terminally differentiated neurons[86]. This mechanism has been describe both in yeast and human cells and was reported to be RAD52-dependent[121,122]. Specifically, two modes of actions in this pathway of RAD52 were presented (Fig. 11). First, RAD52 was suggested to allow homologous RNA transcripts "bridging" of DSB ends, leading to the formation of RNA:DNA heteroduplex complex upon DSB at highly transcribed loci[87,122]. Here, RAD52 utilizes RNA to tether both ends of a homologous DSB forming a DNA synapse for ligation and damage resolution. At the end of this process, RNA degradation by RNase H may also occur[123]. This first mode of action of RAD52 in RNA-dependent-DNA-repair was also studied by Mazina and colleagues in 2017: they suggested that RAD52 allows strand exchange between ssRNA and dsDNA through an unconventional "inverse strand exchange", forming a nucleoprotein complex with dsDNA and promoting strand exchange with ssDNA or ssRNA[87].

Secondly, RNA-mediated DNA repair mechanism is believed to require RAD52 also for promoting the annealing between complementary ssDNA and template RNA[58,123]. In this model, RAD52 forms a RNA-DNA hybrid along the 3' overhang of a DSB. The RNA is used as a template for DNA repair synthesis by reverse transcriptase (RT) activity. After that, the RNA is degraded by RNase H and RAD52 could promote SSA of the opposing homologous ssDNA overhangs. Final processing of the DSB involves gap filling and ligation[123].

In both of the mechanisms it is suggested that the RNA-dependent DNA synthesis for gap filling can be performed by specific DNA polymerases with specific reverse transcriptase activity (i.e. Polη and Polθ[124,125]).

RNA-dependent DNA synthesis was shown in eukaryotes to require only Rad52 and no other paralogs such as Rad59 or recombinase activity of Rad51[87,126].

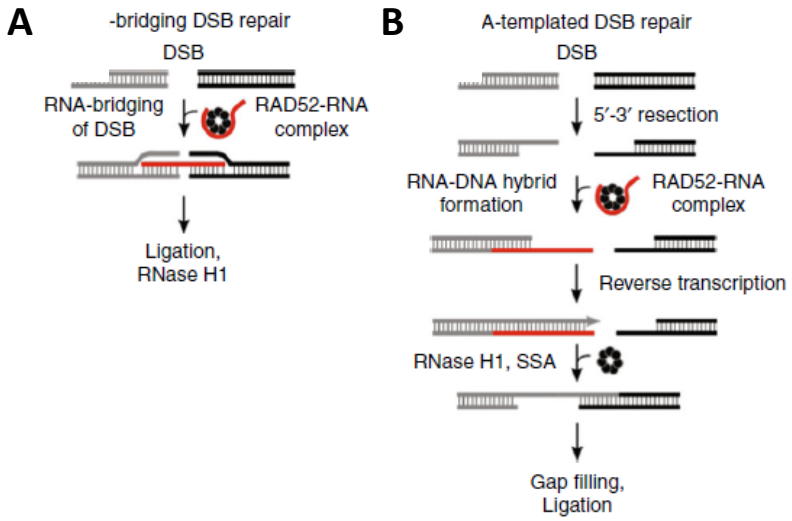


Fig. 11) RAD52-mediated RNA–DNA repair mechanisms; A) RNA-bridging DSB repair model. RAD52 utilizes RNA to tether both ends of a homologous DSB which forms a DNA synapse for ligation. B) RNA-templated DSB repair model. RAD52 forms an RNA–DNA hybrid along the 3'-overhang of a DSB. The RNA is then used as a template for DNA repair synthesis by RT. The RNA is then degraded by RNase H and RAD52 promotes SSA of the opposing homologous ssDNA overhangs. Image taken from McDevitt et al. (2018).

The formation and resolution of R-loops (three-stranded DNA:RNA hybrids) is reported to be commonly involved in specific intracellular signal transduction for efficient DSB repair. Moreover, transcriptionally active genes preferentially recruit HR mediators compared to untrascribed genes[127]. This make R-loop formation an important factor for downstream recruitment of repair proteins. Notably, Tseng and colleagues suggested that RAD52 has a role in the RNA-dependent DNA repair mechanism called Transcription Coupled – Homologous

Recombination (TC-HR)[128]. Here, RAD52 may be recruited at the R-loop through Cockaine Syndrome B (CSB) protein and help the loading of RAD51 on DSB site, in a BRCA1/BRCA2 independent manner[128].

A similar pathway called Transcription Associated – Homologous Recombination Repair (TA-HRR) was suggested by Yasuhara and colleagues[129]. Also in this case, the process of DNA repair in highly transcribed regions required R-loops and RAD52. RAD52 was demonstrated to recruit XPG endonuclease that processes R-loops into substrate with ssDNA overhangs, similar to resected ends that could undergo HR mechanism and DNA repair[129]. Differently from TC-HR, TA-HRR was shown to involve BRCA1 and to occur in S/G2 phase.

Interestingly, although RAD52 is reported to be able to directly bind to R-loop or other DNA:RNA hybrids in these mechanisms, its preferential substrate is ssDNA and not DNA:RNA structure[86,87]. However, recent studies showed an increased affinity of RAD52 for DNA:RNA hybrids containing m5C-modified RNA *in vitro*. Notably, this modification arises in mRNAs when in the presence of DNA damage sites[130].

Additional studies are required to further elucidate RAD52 role in these RNA-dependent DNA repair mechanisms.

1.3.2.5. Regulatory Role

A recent study is worth to be mentioned, even if additional evidences are required to further corroborate their hypothesis. Wang and colleagues reported a role of RAD52 in regulating the balance between single strand break repair (SSBR) mechanisms and double strand break repair (DSBR) mechanisms[131]. Specifically, they proved that RAD52 inhibits SSBR through strong ssDNA

and/or PARP1 binding affinity, reducing DNA-damage-promoted XRCC1/LIG3a co-localization. The inhibitory effects of RAD52 on SSBR neutralize the role of RAD52 in DSBR in specific cellular damage conditions, suggesting that RAD52 may maintain a balance between cell survival and genomic integrity. Moreover, they also reported that the disruption of RAD52 oligomerization affects RAD52's DSBR activity, while it has no effect on its ssDNA binding ability required for RAD52's inhibitory effects on SSBR. For this reason, Wang and colleagues suggest a novel RAD52-inhibition-based strategy to sensitize cells to different DNA-damaging agents[131].

1.3.3. RAD52 in cancer

Cancer cells usually show a high level of genomic instability and DNA damage. More specifically, BRCA2 deficient cell lines, which cannot use canonical DNA repair mechanism and are much more sensitive to DNA damaging agents, rely more on alternative error-prone DNA repair mechanism such as SSA, alt-EJ and BIR. RAD52 plays a central role in regulating these pathways and indeed, it is a significant mediator of the highly mutagenic propensity of some type of cancers[58]. For instance, the BIR-like mechanism, named alternative lengthening of telomeres (ALT), is one of RAD52-dependent processes, which may be linked to cancer development, since many types of cancer rely on this route to efficiently maintain telomere length and to rapidly divide[132].

RAD52, is important to enhance cancer cells viability , or in the dysregulation of cancer cells DNA repair mechanisms[133]. For instance, RAD52 overexpression was reported to correlate with hepatocarcinogenesis in TGF-α/c-myc mice[134] and its depletion or inhibition was shown to decrease cancer incidence and have antileukemic effect in ATM deficient mice[135] and in acute

myelogenous leukemia (AML), B-cell acute lymphoblastic leukemia (B-ALL), T-cell acute lymphoblastic leukemia xenografts with low BRCA1/2 expression[55].

Notably, RAD52 genetic variants were shown to play a role in cancer progression and prognosis in patients, with RAD52 acting either as oncogene or tumor-suppressor.

Expression studies performed on lung squamous cell carcinoma (LSCC) demonstrated that RAD52 protein expression variations correlate with Non-Small Cells Lung Cancer (NSCLC) risk and that RAD52 depletion increases the death of cells undergoing carcinogenic transformations and increases the antitumoral activity *in vivo*[136].

RECQL4-deficient breast, colon and lung cancer cells, presenting a significant RAD52 upregulation were synthetized to ionic radiation[137], whereas cervical and rectal cancer cells, presenting low RAD52 expression levels, were associated with poor response to platinum-based chemotherapies and increased resistance[138,139].

RAD52 aberrant high expression was reported to correlate with the poor life span of patients in rectal cancer, whether in urothelial cancer patients, poor overall survival was reported to correlate with RAD52 low expression[139,140].

Recently, it was showed that RAD52 S346X variant significantly reduced breast and ovarian cancer risk among germline BRCA2 mutation carriers, maybe due to DSB repair reduction by SSA, confirming the idea that RAD52 defects in BRCA-mutated carriers could correlate with a lower tumor development risk[141].

Finally, other studies were performed in these last years to understand the correlation between specific RAD52 single nucleotide polymorphisms (SNP) and tumorigenic risk, for instance in hepatitis B Virus (HBV) – hepatocellular carcinoma (HCC)[142] and in colorectal cancer[143].

1.3.4. RAD52 inhibition in Synthetic Lethality Therapies

Cancer cells, due to their intrinsic genome mutations and subsequent inactivation of many of the canonical DNA repair mechanisms, must rely on alternative DNA repair pathways. Indeed, tumor cells accumulate high levels of spontaneous and drug-induced DNA damage, but they survive because of enhanced or altered DNA repair activities alternative to the canonical ones[57]. RAD52 is, as already mentioned, a mediator of many of these alternative pathways that may act as backups[58,84].

In the last few years, RAD52 has emerged to have an important oncogenic role, as a mediator of many DDR pathways that cancer cells rely on when canonical ones are disrupted. For example, RAD52 induces synthetic lethality when inhibited in many cancer cells defective of DNA repair-related proteins, such as BRCA1, BRCA2, PALB2, XIAB2 and RAD51 paralogues (i.e. RAD51B, RAD51C,RAD51D, XRCC2, XRCC3)[12,54,56,144]. Among these DSB repair mediator proteins, the synthetic lethality relationship between RAD52 and BRCA1/2 has, in recent years, been thoroughly investigated in cells[145].

Nonetheless, RAD52 is normally non-essential for cell survival, but its role becomes fundamental only in cancer conditions where other DSB repair-related proteins are mutated[54,56]. This makes, RAD52 an attractive target for synthetic lethality-based anticancer therapy, to pursue selectivity and targeted cancer therapies.

Indeed, a great effort in these last few years was put in the search of RAD52 inhibitors to pursue the synthetic lethality paradigm in BRCA1/2-depleted cancers and in drug-induced-“BRCA1/2-ness” mimicking cells. Many are the approaches that have been used to discover and characterize these molecules.

In the next paragraphs the most relevant RAD52 inhibitors will be presented with their hallmarks and possible application.

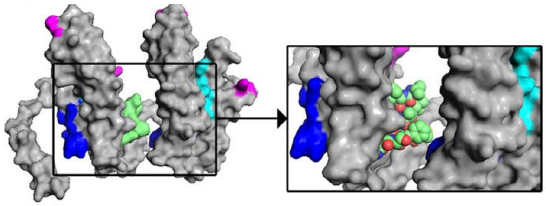


Fig. 12) F79 aptamer surface (light green) is shown between 2 RAD52 monomers and it better fits to the binding groove than the other RAD52 monomer. The zoomed box focuses on the area that the aptamer occupies between the 2 RAD52 monomers. Image taken from Cramer-Morales et al. (2013).

F79 is a peptide aptamer developed in 2013 from Cramer-Morales and colleagues and it is the first RAD52 inhibitor ever described. This peptide was designed in order to inhibit RAD52 and exert synthetic lethality in BRCA-disrupted and/or HR mutated tumor cells[55]. The peptide aptamer was designed after mutagenic assays reporting the fundamental role played by residue Phe79 in RAD52 DNA binding and in RAD52 protomer-protomer hydrophobic interactions (Fig. 12). Using computational methods, Cramer-Morales group designed a peptide aptamer containing the 13-aminoacid sequence surrounding Phe79 (F79) in the α2 helix of RAD52. What they reported was a significant abrogation of RAD52-DNA binding activity after treatment with F79 aptamer.

In cellulo studies reported that F79 aptamer was able to selectively kill BRCA-deficient leukemia cells, with low risk for normal cells counterparts (EC_{50} < 5 μM). Indeed, synthetic lethality was observed in CML cells carrying BCR-ABL1 mutation (BRCA1 downregulated), acute promyelocytic leukemia (APL) cells with PML-RAR mutation (RAD51C downregulated) and in other patient-derived leukemia cells with epigenetic modifications resembling "BRCA-ness" phenotype[55]. *In vivo* tests showed that F79 aptamer treatment significantly extents life spans of Severe Combined Immunodeficiency (SCID) mice with BCR-ABL1—positive leukemia. Additionally, F79 aptamer treatment resulted in synthetic lethality in BRCA1/2-mutated breast, pancreatic, and ovarian

cancer cells and displayed synergistic effect with imatinib (approved for BCR-ABL1-positive leukemia) and ATRA (for PML-RAR-positive leukemia)[55].

Later on, in 2019, F79 aptamer was observed to inhibit acute myeloid leukemia (AML) cells proliferation and to promote cell apoptosis in AML cells treated with etoposide. Moreover, RAD52 aptamer treatment affected also the expression and activation of the apoptotic signal protein STAT3[146]. For this reason, F79 was proposed as a novel therapy against STAT3 continuous activation in myeloid leukemias[147,148].

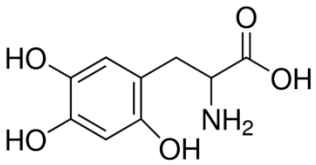


Fig. 13) Chemical structure of 6-OH-DOPA compound.

6-OH-DOPA (Fig. 13) was the first small molecule inhibitor discovered for RAD52 in 2015[73]. The group of Chandramouly and colleagues set up a high through put screening (HTS) using the Fluorescence Polarization method to detect in a drug-like compounds library (*Sigma Lopac*) molecules that could affect DNA-RAD52 interaction. 6-OH-DOPA was identified and further characterized with other biophysical studies. Indeed, it was reported that 6-OH-DOPA was not only a DNA-RAD52 interaction disruptor with an IC_{50} of 1.1 µM, but also a disruptor of RAD52 heptamer structures. 6-OH-DOPA was able to inhibit *in vitro* SSA (RAD52-dependent DNA repair mechanism) with no or little effect on other mechanisms such as HR, D-NHEJ in BRCA-proficient cells[149]. Moreover, they showed also that in BRCA1-depleted triple negative breast cancer (TNBC) cells and in BRCA-deficient AML and CML patient cells, 6-OH-DOPA selectively inhibited cell proliferation.

While these results may appear promising for the development of a new RAD52 inhibitor, 6-OH-DOPA is known to be a dopaminergic toxin contributing to Parkinson disease and degeneration of mitral neurons[150] and is therefore unsuitable for anticancer therapy.

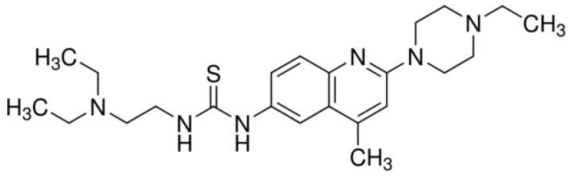


Fig. 15) Chemical structure of D-I03.

D-I03 is the compound that we selected as reference inhibitor of RAD52. It was identified in 2016 by Huang and colleagues, through a HTS screening using a fluorescence quenching assay[151]. The assay was set up to test the ability of compounds to inhibit RAD52 activity on complementary strand annealing and invasion of ssDNA into homologous duplex of DNA. Among all the compounds, D-I03 (Fig. 15) came out as a promising RAD52 inhibitor showing the highest inhibitory effect on D-loop formation *in vitro,* on proliferation of BRCA1/2-deficient cell lines (Capan-1, MDA-MB-436, UWB1.289) and on proliferation of BRCA1-deficient BCR-ABL positive CML cells from patients. No D-I03 effect was observed in BRCA-proficient cells. Additionally, D-I03 led to inhibition of RAD52-mediated SSA foci formation after cisplatin treatment. Finally, the binding affinity K_d of D-I03 for RAD52 was 26 µM as determined by SPR, supporting the idea that the effects observed *in vitro* and *in cellulo* could be correlated to D-I03 direct binding to RAD52 and to the consequent abrogation of its DNA-binding activity[151].

Later on, many other groups put a great effort in the search of alternative and more efficient RAD52 inhibitors. Two were the main approaches which were exploited to screen molecules: 1) HTS fluorescence-based screenings of new small molecules and drug-like molecules were used to identify potential inhibitors of RAD52-DNA[152] or RAD52-RPA interactions[153]; 2) virtual screening campaigns were also performed to identify molecules able to inhibit RAD52 activity, targeting different RAD52 pockets, either the DNA binding pocket[154,155] or the oligomerization/self-associating domain[156] of RAD52 (using the PDB of RAD52 N-terminal domain).

Both these drug discovery approaches allowed to obtain promising hit compounds, with corroborated effects on RAD52 activity *in vitro* and with a significant effect on different cancer cell lines (i.e. colony forming units inhibition, growth inhibition, mortality increase, X-rays sensitivity increase), also bearing mutations on BRCA2-mediated homologous recombination DNA repair mechanism. Notably, some studies have been also performed in tumor xenografts[157].

All these evidences allowed to identify and characterize RAD52 inhibitors to be eventually exploited in synthetic lethality therapies, with an IC_{50} / EC_{50} in the low µM range (ranging from 5 µM to 50 µM). Notably, Epigallocatechin-3-monogallate, one of the RAD52 inhibitor compound described from Hengel and colleagues in 2016, is the only RAD52 inhibitor reported with an *in vitro* IC_{50} in the mid-nanomolar range (RAD52-DNA interaction inhibition)[152]. Nevertheless, this potent compound did not undergo further characterization, maybe due to its fast hydrolysis in aqueous buffer.

However, although the list of novel RAD52 inhibitors is growing alongside the interest in this therapeutic "hot target", none of these new inhibitors have been further developed or have yet reached clinical phase. Therefore, the need for new detailed information on RAD52 mechanism of action is as urgent as the medical need to identify novel molecules capable of effectively inhibiting RAD52 in order to improve synthetic lethality anticancer strategies.

2. Aim of the Project

Synthetic lethality therapies have several limitations mainly due to resistance mechanisms that frequently occurs: in particular, cells have many redundant metabolic pathways and DNA repair mechanisms, which they rely on to survive. Indeed, it is important to continue to develop new strategies to enhance the anticancer effects of in-use-chemotherapies and to prevent the onset of resistance to treatment.

We focused our investigation on RAD52, which is a promising anticancer target to be inhibited in BRCA-deficient cell lines[54,56] and in cells with drug-induced "BRCAness" status[158,159], in order to pursue synthetic lethality therapies with high selectivity and efficacy limiting potential side-effects. Hence, we pursued 3-different drug discovery pipelines focusing on RAD52 and its inhibition using multidisciplinary approaches (i.e. Virtual Screening Campaign; Fragment-Based-Approach 1D ^{19}F NMR screening of fluorinated fragment library; Aptamers).

Nevertheless, RAD52 mechanism of action is still poorly understood. Therefore, RAD52 biophysical characterization is indeed one of the main aims of this project that runs in parallel with, but it does also support, the drug discovery approach to explain RAD52 role and to clarify its interactions with DNA.

Finally, since the complete structure of RAD52 is still elusive, another important milestone of this project is the characterization of RAD52 quaternary structure. These structural information are critical to support the description of its function and interaction with other proteins as well as to provide a new twist to the drug discovery approach built around this novel cancer-related target.

In this PhD project, the initial description of the expression and purification protocols of two recombinant RAD52 protein constructs are followed by their biophysical and structural characterization. The last part of the thesis is instead entirely devoted to the description of three drug discovery approaches applied on the full-length recombinant protein and to the characterization of the hit compounds identified by each of these approaches.

3. Materials and Methods

3.1. Biophysics and structural biology

3.1.1. Cells transformation

Two aliquots of *E. Coli* Rosetta (DE3)pLysS - Novagen competent cells (50 µL) (*Sigma Aldrich* – Merck) were thawed and transformed with 0.3 µL of pET16b-His-hRAD52 (1227 ng/µL) vector or 0.6 µL of vector pET16b-His-hRAD52 [1-212] N-Term (617 ng/µL) vector, respectively. These vectors were realized starting from full length RAD52 sequence available on Protein Data Bank (PDB), using the custom-made preparation service of *NeoBiotech*. Vectors maps are reported in figure 16. After 30 minutes of incubation on ice, the mixture containing cells and plasmids were heat shocked by placing each transformation tube at 42 °C for 30 seconds; tubes were then immediately put on ice for 2 minutes and, 200 µL of Super Optimal Broth (SOC) medium were added to the bacterial solutions. Cells were grown at 37 °C for 1 hour with shaking at 400 rpm and then plated on Luria-Bertani broth (LB) agar plates supplemented with Amp 100 µg/mL for the selection of single colonies of transformed cells. Plates were kept at 37 °C overnight and stored at 4 °C for no more than 1 week.

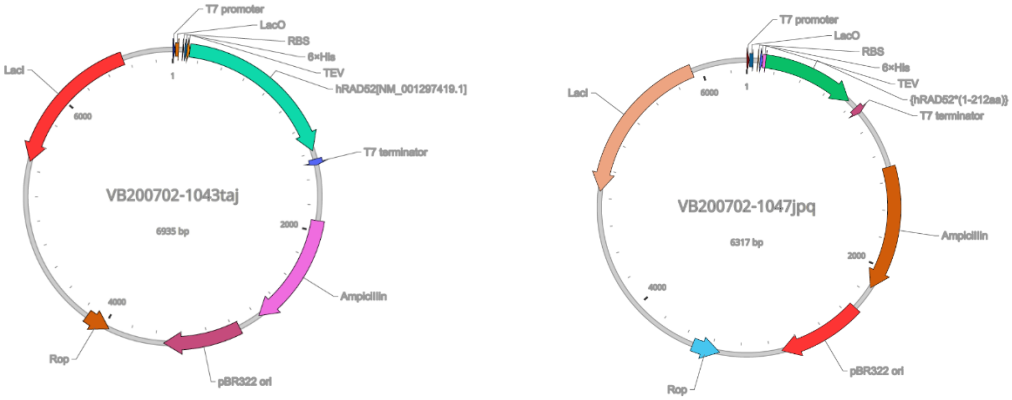


Fig. 16) pET16b vectors of 6x-His-RAD52 FL and 6x-His-RAD52 [1-212]. These vectors encode for a 6x-His-TEV tag, T7 promoter and T7 terminator sequences, replication origin pBR322 ori and Ampicillin resistance cassette; maps taken from NeoBiotech vectors datasheet.

3.1.2. Protein expression

To express recombinant RAD52 FL and RAD52 [1-212], one single colony of transformed *E. Coli* Rosetta (DE3)pLysS was picked from the plate and added to a starter culture of 10 ml of LB buffer supplemented with Ampicillin 100 µg/µL . Bacteria were grown overnight at 37 °C in a shaking incubator (180 rpm). The next day, 10 mL of the saturated overnight culture were inoculated into a 1 L LB culture. The bacterial culture was grown at 37 °C with shaking at 180 rpm. Once the mid-exponential growth phase was reached, at optical density at 600 nm ($0D_{600}$) = 0.8, the culture was induced using Isopropil-β-D-1-tiogalattopiranoside (IPTG). After induction, the bacterial culture was grown at 25 °C for other 4 hours. The expression protocol has been initially optimized testing different growth times, temperatures and IPTG concentrations to increase the yield of protein expression. The selected conditions for both proteins were: induction with 0.8 mM IPTG for 4 hour at 25 °C, in line also with other already published protocols[73].

Bacterial cells were harvested by centrifugation at 6000g for 45 minutes at 4 °C. Pellets for both RAD52 FL and [1-212] were stored at -20°C up to 15 days.

3.1.3. Protein purification

3.1.3.1. RAD52 FL

Bacterial pellet was resuspeded in a 50 mL volume of lysis buffer (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 10 mM Imidazole, Tween20 0.01%, 2mM DTT, Protease Inhibitors EDTA-free 1x (*Roche*)). Cell suspension was then lysed on ice by sonication (5 minutes with rounds of 30 seconds each, 65% amplitude, KE 76 Tip, Bandelin Sonoplus HD2070 sonicator). After centrifugation of 1 hour at 30000g the surnatant was collected and filtered with a 0.2 µm MiniSart

Syringe Filter and the, loaded on a 5 mL HisTrap HP chromatography column (*Cytiva*) equilibrated with Buffer A (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5%glycerol, 10 mM Imidazole, Tween20 0.01%, 2mM DTT). Washing steps were performed with 4% and 8% of buffer B (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 M Imidazole, Tween20 0.01%, 2mM DTT). Fractions eluted at 40% buffer B were pooled together and loaded on a 5 mL HiTrap Desalting chromatography column (*Cytiva*) equilibrated with buffer C (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20). The eluted protein was finally loaded on HiTrap Heparin HP chromatography column (*Cytiva*) equilibrated with buffer C. The protein was then eluted from this last column using a linear gradient up to 100% buffer D (25 mM Tris-HCl pH 7.5, 1.5 M NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20) in 5 CVs. Eluted protein fractions were collected pooled together and quantified by Nanodrop, using the extinction coefficient of the monomeric RAD52 protein ε_{280} = 41370. Purity and degradation were checked through SDS-Page gels. Protein storage buffer was 25 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20. The purified protein was flask frozen in liquid nitrogen and stored at -80°C.

3.1.3.2. RAD52 [1-212]

Bacterial pellet was resuspeded in a 50 mL volume of lysis buffer (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 10 mM Imidazole, Tween20 0.01%, 2mM DTT, Protease Inhibitors EDTA-free 1x (*Roche*). Cell suspension was then lysed on ice by sonication (5 minutes with rounds of 30 seconds each, 65% amplitude, KE 76 Tip, Bandelin Sonoplus HD2070 sonicator). Lysed cells were centrifuged for 1 hour at 30000g. The collected surnatant was filtered with a 0.2 µm MiniSart Syringe Filter and then loaded on a 5 mL HisTrap HP chromatography column (*Cytiva*) equilibrated with Buffer A (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5%glycerol, 10 mM Imidazole, Tween20

0.01%, 2mM DTT). Washing steps were performed with a 4%, 8% and 40% of buffer B (25 mM TRIS-HCl pH 7.5, 500 mM NaCl, 5%glycerol, 0.5 M Imidazole, Tween20 0.01%, 2mM DTT). Fractions eluted at 100% buffer B (0.5 M Imidazole) were pooled together and quantified by Nanodrop instrument using the extinction coefficient ε_{280} = 21890 calculated from the protein sequence. Purity and degradation were checked through SDS-Page gels. Protein storage buffer was 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20. The purified protein was flask frozen in liquid nitrogen and stored at -80°C.

3.1.4. Protein characterization

3.1.4.1. SDS-Page

SDS-PAGE was performed using precast polyacrylamide gels (NuPAGE 4-12% BisTris Gel, *Invitrogen*). Different concentrations of protein samples were mixed with 4X Loading buffer (0.25 M Tris-HCl pH 6.8, 8% SDS, 0.3M DTT, 30% Glycerol, 0.4% Bromphenol Blue) prior to denaturation at 95°C for 5 minutes. After samples loading, precast gels were run in XCell SureLock Mini-Cell Electrophoresis System (*Invitrogen*) in MOPS SDS running buffer (*Invitrogen*) with a constant voltage of 120 mA for about 90 minutes. Gel were then stained with Comassie Blue staining buffer (40% EtOH, 10% Acetic Acid, 0.05% w/v comassie blue G-250) for 15 – 30 minutes and destained with a destaining Buffer (8% acetic acid, 25% EtOH). Protein bands images were visualized and quantified using ChemiDoc Imaging System (*BioRad*).

3.1.4.2. Western Blot

SDS-PAGE gels were run as described in the previous section, without Comassie Staining. Gels were blotted onto a 0.22 µm nitrocellulose membrane (Protran BA83, GE Healthcare) applying a

constant voltage of 30 V for 2 hours in NuPage MOPS running buffer (*Invitrogen*) with 10% isopropanol. Protein transfer to the membrane was checked through a fast incubation in a Red Ponceau Solution. The membrane was then incubated for 1 hour in blocking solution (5% no-fat milk-TBST) at room temperature with agitation. The following incubation with primary antibody (anti-HisTag, *Thermofisher*) diluted 1:1000 in 5% no-fat milk-TBST was performed overnight at 4°C with agitation. The following day, after three washes (10 minutes each) with TBST, an incubation with a HRP-conjugated anti-rabbit secondary antibody (Abcam) diluted 1:10000 in 5% no-fat milk-TBST was performed for 1 hour at room temperature with agitation. After three additional washes, the membrane was finally incubated with Amersham ECL Star substrate (*Euroclone*) and exposed to ChemiDoc chemiluminescence for detection (*BioRad*).

3.1.4.3. Liquid Chromatography-Mass Spectrometry (LC-MS)

For RAD52 FL and [1-212] quality checks, Liquid Chromatography-Mass Spectrometry experiments were performed by dr. N. Liessi and dr. A. Armirotti (IIT Genova). 1 mL of cold acetone was added to protein samples (100 µL, 0.45 µg/mL) and the sample was incubated at -20°C, o/n. Samples were then centrifuged at 20000*g* at 4 °C for 30 min. The supernatant was the removed and the pellet was dissolved in 200 µL H_2O:CH_3CN=1:1 + 0.1% formic acid for LC-MS analysis.

Samples were acquired in positive (ESI+) ion mode. The column used was an ACQUITY UPLC® BEH300 C4, 1.7 µm internal diameter, 1.0x100 mm (Waters).

Eluents were water with formic acid 0.1% (A) and acetonitrile with formic acid 0.1% (B). Injection volume was 1 µL. The flow rate was 0.100 mL/min, the column was kept at 60 °C, samples were

eluted with a gradient program: 0.0 -1.0 min 10 % B; 1.0 – 6.0 min 5 to 100 % B; 6.0 – 7.5 min 100 % B; 7.5 – 7.6 min 100 to 10% B. The column was then reconditioned for 2.4 min. The total run time was 10 min. Scan range was set from 500 to 2000 m/z (resolution mode). Cone voltage was 35 V, source temperature was set to 90 °C, desolvation gas was 800 L/h and cone gas flows was 30 L/h. Desolvation temperature was set to 450 °C. The scan time was 0.3 sec, low collision energy was set to 4 eV. The real-time spectra recalibration was performed using Leucine enkephalin 2 ng/mL as Lock Mass.

For data acquisition, MassLynx software (Waters) was used. The MaxEnt 1 algorithm available in the MassLynx software (Waters) was used for deconvolution of ESI mass spectra.

3.1.4.4. Native Gel Electrophoresis

Native Gel electrophoresis was performed using precast native gels (NativePAGE 3-12% BisTris Gel, *Invitrogen*). Protein samples were mixed with 4X Native loading buffer (*invitrogen*). After samples loading, precast gels were run in XCell SureLock Mini-Cell Electrophoresis System (*Invitrogen*) using Cationic and anionic running buffers (*Invitrogen*), with a constant voltage of 150 mA for about 90 minutes. Gel were than stained with Comassie Blue staining buffer (40% EtOH, 10% Acetic Acid, 0.05% w/v comassie blue G-250) for 15 – 30 minutes and destained with Destaining Buffer (8% acetic acid, 25% EtOH). Protein bands images were visualized and quantified using ChemiDoc Imaging System (*BioRad*).

Native gels-based experiments for small molecules inhibitory activity determination are schematically represented in Fig. 17. The assays were set up by incubating RAD52 FL protein (0.8 mg/mL) with different compounds at the maximum concentration which ensured the solubility

of the compounds in assay buffer the presence of 2% DMSO (20 µL reaction volume). After 1 hour of incubation, samples were loaded on the gel as descripted above. Whether the compounds had an effect on RAD52 high MW superstructures arrangement, the bands intensity corresponding to the high MW structures should decrease and the bands intensity corresponding to low MW should increase.

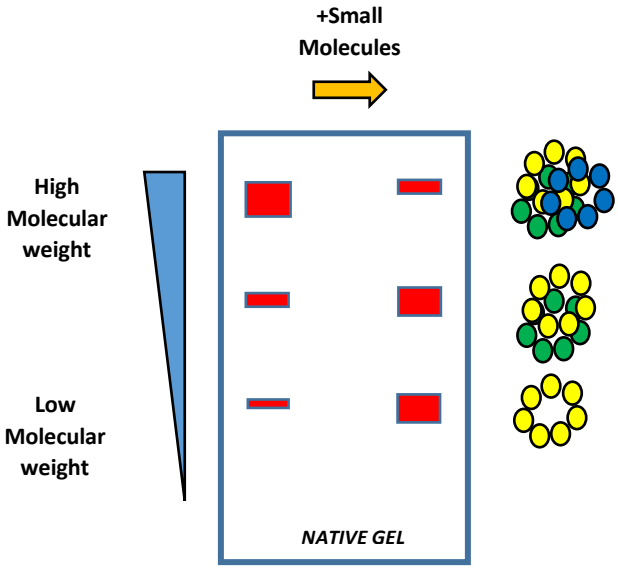


Fig. 17) Schematic representation of Native Gel assay used to detect inhibitors disruptive effect on RAD52 high MW superstructures.

3.1.4.5. Circular Dichroism (CD)

Far-UV CD spectra were recorded on a Jasco J-1100 spectropolarimeter (Jasco, Essex, United Kingdom), equipped with a temperature control system, using a 1 mm quartz cell. The spectra were recorded in the far-UV range 190–260 nm, using RAD52 FL and RAD52 [1-212] protein at 5 µM and 10 µM concentrations, respectively. Assay buffers were optimized in order to remove CD interference signals, avoiding the use of chlorine ions and glycerol and maintaining, however, an ionic strength comparable with the one of the storage buffers. RAD52 FL assay buffer was 25 mM phosphate buffer NaPi, pH 7.5 and 250 mM NaF; RAD52 [1-212] assay buffer was NaPi, pH 7.5

and 500 mM NaF. Constant N_2 flush at 4.0 L/min was applied. Raw spectra were corrected for buffer contributions and the detected signal was expressed as mean residue molar ellipticity [θ] ($deg \times cm^2 \times dmol^{-1}$).

For protein secondary structure analysis, the scanning speed was set to 100 nm/min, digital integration time to 1 s, and the temperature set to 20 °C for all experiments. Each spectrum was obtained as an average of 10 scans. No shaking was applied on the samples during measurements. Data analysis were performed using *Dychroweb* (Institute of Structural and Molecular Biology Birkbeck College – University of London).

For protein thermal stability analysis, CD experiments were performed using the *variable temperature method* with a temperature scanning from 20°C to 95°C. No shaking was applied during data collection. Thermal stability was measured monitoring the CD signal at 222 nm wavelength during temperature scan.

For aptamers secondary structure analysis, far-UV CD spectra were recorded using a 1 mm quartz cell in the far-UV range 190–320 nm. The concentration of the selected aptamers was 100 ng/µL and the assay buffer was 12.5mM NaPi, 250 mM NaF. The scanning speed was 100 nm/min, digital integration time 1 s, and the temperature 20°C for all experiments. Each spectrum was obtained as an average of 10 scans. No shaking was applied on the samples during data collection.

3.1.4.6. Size Exclusion Chromatography (SEC)

For oligomerization state analysis, the recombinant proteins RAD52 FL and RAD52 [1-212] (500 µL) were loaded on a size exclusion chromatographic column (Superdex200 Increase 10/300 GL, *Cytiva*) previously equilibrated with protein storage buffers: 25 mM Tris-HCl pH 7.5, 250 mM

NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20 for RAD52 FL; 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20 for RAD52 [1-212].

For small molecules and aptamers activity characterization, RAD52 FL protein was loaded on the size exclusion chromatographic column after 1 hour 30 minutes incubation with E5 compound and Apt1 aptamer, respectively.

3.1.4.7. Negative Staining-Transmission Electron Microscopy (TEM)

Recombinant pure RAD52 FL and RAD52 [1-212] proteins were diluted up to 0.01 mg/mL in the final buffer containing 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2% glycerol, 1 mM DTT, 0.005% Tween20. For negative staining 5 µL of protein samples were placed onto 400- mesh copper carbon film grids and, after a single wash in the same buffer they were negatively stained with 1% uranyl acetate in distilled water. Once dried, the samples were observed with a JEM-1011 (JEOL) transmission electron microscope (TEM) with thermionic source (W filament) and maximum acceleration voltage of 100 kV equipped with a Gatan Orius SC1000 charge-coupled device (CCD) camera (Gatan AMETEK). Experiments were performed in collaboration with dr. R. Marotta (IIT Genova).

3.1.4.8. Cryo- Electron Microscopy (Cryo-EM)

3.1.4.8.1. Grid optimization and data collection

Three µL aliquots of purified RAD52 FL protein (0.16 mg/mL nominal concentration) were applied to Quantifoil holey carbon grids (Cu, 300 mesh, 1.2/1.3 µm) at 100% humidity and 4.5 °C, blotted with filter paper for 8 s and vitrified in liquid ethane using a semiautomated cryo plunger (Vitrobot Mark IV ThermoFisher Scientific).

Grid optimization has been performed *in house* on a Tecnai F20 transmission electron microscope (ThermoFisher Scientific) operated at 200 kV equipped with a Gatan Ultrascan CCD camera (Gatan AMETEK). The final data collection has been performed at the European Molecular Biology Laboratory (EMBL) Heidelberg under the iNEXT project titled "*Full-length RAD52: an innovative potential target for cancer-related drug discovery*" (PID: 15983). The final data set was acquired in energy filtered (EF) TEM mode on a Thermo Fisher Scientific Glacios SelectrisX TEM equipped with a Falcon4 EC direct electron detector (Thermo Fisher Scientific).

3.1.4.8.2. Image Analysis

The final data set consisted in 831 gain corrected counting mode movies acquired from the vitrified grid with a defocus ranging from -1.2 µm to -2.5 µm. Each movie, with a pixel size of 1.154 Å/pix, was composed of 34 frames collected with a total dose of 40.9 $e^-/Å^2$ (1.2 $e^-/Å^2$/movie frame).

Motion correction was performed on all movies (frames 2-31) with dose-weighting using the Relion3.1 MotionCor implementation[160]. CTF correction was performed with CTFFIND 4.1 on all the 831 dose weighted motion corrected micrographs[161].

About 3900 representative particles were picked from 10 micrographs (391 particles/micrograph) using the Laplacian of Gaussian filter as implemented in Relion 3.1. The obtained preliminary low pass filtered 2D class averages have then been used for automated particle picking on a total of 831 micrographs. This resulted in 580952 particles which were extracted and down-sampled for several iterative rounds of 2D classification and selection. A total of 258839 particles from 20 selected 2D classes were then subjected to unsupervised 3D

classifications (number of classes K = 4) using a RAD52 low resolution initial model obtained from 2D averages and using the 3D initial model algorithm implemented in Relion 3.1. Similar 3D classes have been obtained using a sphere as unbiased initial model. The subsets of particles corresponding to the two more represented classes, after being re-extracted at full resolution, were used for the final refinement. The final RAD52 full length electron density map was resolved at 3.4 Å by the 0.143 FSC criterion after post-processing (including CTF refinement and polishing). Cryo-EM grids preparation and data elaboration were performed by dr. R. Marotta (IIT Genova), in collaboration with dr. S. Fromm (EMBL Heidelberg).

3.1.4.9. Atomic Force Microscopy (AFM)

RAD52 FL samples were diluted in AFM buffer (30 mM MOPS pH 7.3, 3 mM $MgCl_2$, 20 mM NaCl) to reach low nM protein concentration (35 nM, 70 nM, 350 nM) in the presence or absence of linearized plasmid 62 nM pBR322 (*Thermofisher*)[162]. 20 µL of each sample were loaded onto freshly cleaved mica and left at room temperature for 5 min to allow samples to deposit on the mica slide. After 5 minutes, samples were washed with ultrapure water (MilliQ, *Millipore*) and dried using nitrogen gas.

Morphological images of protein and protein-DNA complexes were acquired using a jpk Nanowizard AFM (*Bruker*) operated in a tapping mode using NSG10 Tips (*NT-MDT Company*). Images 256x256 pixel size were captured at scan size of 10 µm, 3 µm and 1 µm and processed by first-order-flattening to remove sample tilt (cantilever resonance excitation frequency of 1.5 Hz for 3 µm-scan-size images and 2 Hz for 1 µm-scan-size images). Measurements of particles diameters and heights were performed on 1 µm-scan-size-images. Reported values are the average of >20 measurements on 3 independent protein samples.

3.1.4.10. Electrophoretic Mobility Shift Assay (EMSA)

EMSA assay was optimized to detect compounds effects on DNA-protein interaction. Similar assays have been performed also by Chandramouli and colleagues in 2015[73]. A schematic representation of the experiment is reported in Fig. 18. Briefly, this technique is based on the detection of a fluorescently labelled probe (i.e. 30 nucleotides ssDNA labelled with cyanine5) that migrates faster in the agarose electrophoresis gel when alone compared to ssDNA-RAD52 complex, which migrates slowly and remains in the upper part of the gel.

For assay optimization 10 nM dT_{30}–cy5 labelled ssDNA was incubated with RAD52 FL increasing concentrations (20 nM - 300 nM) in assay buffer 25mM Tris-HCl pH 7.5, 0.1mM DTT, 3 mM NaCl, 0.001% Tween20, 0.1mg/mL BSA, 0.5mM $MgCl_2$, 10% glycerol in a final volume of 20 µL. After 30 minutes of incubation at room temperature, reactions were fixed with 0.2% glutaraldehyde for 10 minutes at room temperature and were run in a 0.8% agarose gel containing 10% glycerol in TAE buffer for 60 minutes at 5.3 V/cm. Finally, labelled ssDNA and ssDNA-protein complexes were visualized using ChemiDoc Imaging System (*BioRad*).

To assess the effect induced by E5 fragment on DNA-RAD52 interaction, 200 nM RAD52 and 10 nM dT_{30}–cy5 labelled ssDNA were incubated with different concentrations of E5 (20 µM - 5 mM). After 60 minutes of incubation, reactions were fixed with 0.2% glutaraldehyde for 10 minutes at room temperature and were run in a 0.8% agarose gel containing 10% glycerol in TAE buffer for 60 minutes at 5.3 V/cm. Finally, labelled ssDNA and ssDNA-protein complexes were imaged using ChemiDoc Imaging System (*BioRad*).

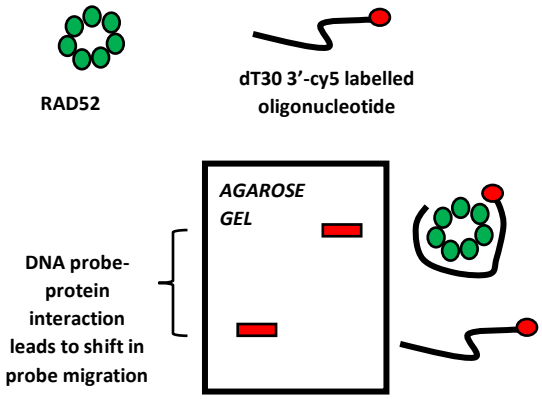


Fig. 18) Schematic representation of EMSA assay used to detect and measure RAD52-DNA interaction.

3.1.4.11. Fluorescence Resonance Energy Transfer (FRET)

FRET-based analyses were carried out to characterize DNA-RAD52 interaction and to determine the effect of small molecules on DNA-RAD52 complex. A similar assay have been proposed by Grimme and colleagues[74,79]. A schematic representation of the experiment is reported in Fig. 19. Briefly, this technique is based on the use of a 30 nucleotides dually labelled-ssDNA with cyanine3 (donor) at 5' end and cyanine5 (acceptor) at 3'. When the donor fluorophore is excited by light, it can transfer its energy to the acceptor fluorophore, only if it is nearby. This technique allows to determine if DNA wraps around RAD52 and if the two fluorophores come close enough to allow FRET phenomenon to occur, upon light irradiation.

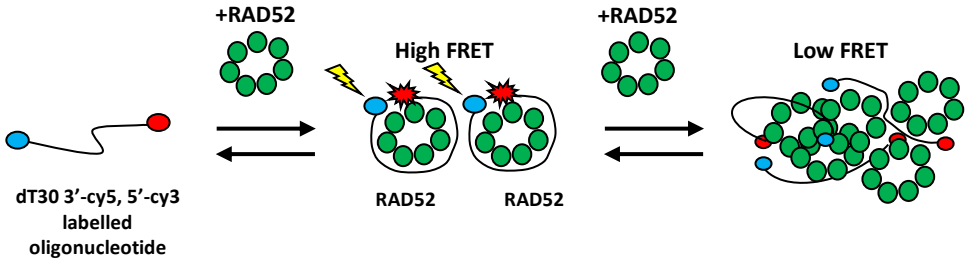


Fig. 19) Schematic representation of FRET assay used to detect and measure RAD52-DNA interaction. When cy5-cy3 labelled ssDNA wraps around RAD52 oligomeric rings, the two fluorescent dyes are close and FRET effect occurs; when RAD52 concentration is higher (>> RAD52:ssDNA = 1:1) FRET effect decreases due to steric hindrance induced by RAD52 high molecular weight arrangements.

Measurements were carried out using a Spark Microplate multimodal reader instrument (*Tecan*) equipped with a fluorescence filters system with the following set up: excitation wavelength 535 nm, excitation bandwidth 25 nm, emission wavelength 595, emission bandwidth 35 nm, gain 66, 40 µs integration time (for cy3 fluorescence detection); excitation wavelength 535 nm, excitation bandwidth 25 nm, emission wavelength 680 nm, emission bandwidth 30 nm, gain 66, 40 µs integration time (for cy5 fluorescence detection). Experiments were set up in a flat black 96-well-plate (*Corning*).

All the experiments were carried out in 25 mM Tris-Acetate, 1 mM DTT.

To optimize this assay, cy3- dT_{30} –cy5 double labelled ssDNA were incubated with different concentrations of RAD52 FL in order to determine the best ssDNA-RAD52 ratio able to provide the highest FRET signal.

Briefly, RAD52 FL protein at 20 different concentrations, ranging from 2.5 nM to 150 nM, was added to plate wells, each containing 1 nM labelled ssDNA in assay buffer, to determine a dose-response curve. After 5 hours for FRET signal stabilization, the fluorescence of donor and acceptor Cy3 and Cy5 were measured, monitoring, after the excitation of Cy3 at 530 nm, the

emission of the acceptor Cy5 at 660 and of the donor Cy3 at 535, respectively. FRET efficiency (E_{FRET}) was calculated as $I_{Cy5}/(I_{Cy5} + I_{Cy3})$ after blank subtraction. The data were collected as the average of three independent experiments.

FRET-based experiments for small molecules activity determination were set up incubating 1 nM Cy5-Cy3 ssDNA and 5 nM RAD52 FL with different compounds at 500 µM concentration (where indicated, concentrations were reduced according to compounds solubility limitations in the assay buffer). After 5 hours, fluorescence values were collected and E_{FRET} was calculated as described above. The effect of each compound was determined by comparing the E_{FRET} value of RAD52 FL – labelled ssDNA in the absence and presence of the compounds. The presence of active small molecules should induce a decrease of RAD52 FL – labelled ssDNA FRET. A schematic representation of the assay is reported in Fig. 20.

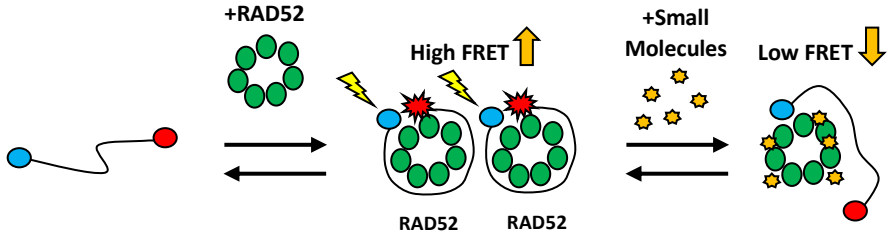


Fig. 20) Schematic representation of FRET assay to detect small molecules effect on DNA-RAD52 interaction.

3.1.4.12. Fluorescence Polarization (FP)

Fluorescence Polarization experiments were performed to investigate DNA – RAD52 FL interactions using a fluorescently labelled DNA probe. Upon light excitation, a small molecule (i.e. probe ssDNA only) rotates fast and emits a low polarized light, whether a big molecule (i.e. probe DNA in complex with protein) rotates much slowly and emits a highly polarized light. Differences

in light polarization detected upon addition of protein to the fluorescently labelled DNA allow to determine the binding state of the DNA to the target protein. A schematic representation of the experiment is reported in Fig. 21. 10 nM 6FAM (fluorescein)-conjugated dT_{30} ssDNA (*Merck*) was added to 16 different concentrations of RAD52 FL ranging between 6.25 µM and 0.2 nM in assay buffer (25 mM Tris-HCl pH 7.5, 62.5 mM NaCl, 5% glycerol, 1 mM DTT). 100 µL of each reaction was incubated in a flat black 96-plate (*Corning*) at room temperature for 15 minutes. FP of samples were then measured using a Spark Microplate multimodal reader instrument (*Tecan*) with the following set up: Excitation wavelength 485 nm; Emission wavelength 535 nm; Emission Bandwidth 20 nm; Integration time 20 µs; reference mP value was set on 20 mP.

To determine the affinity (K_d) of DNA binding to RAD52, FP values were plotted as a function of protein concentrations and fitted using a sigmoidal dose-response curve with GraphPad Prism 7.0. Data are the average of multiple (>3) independent experiments.

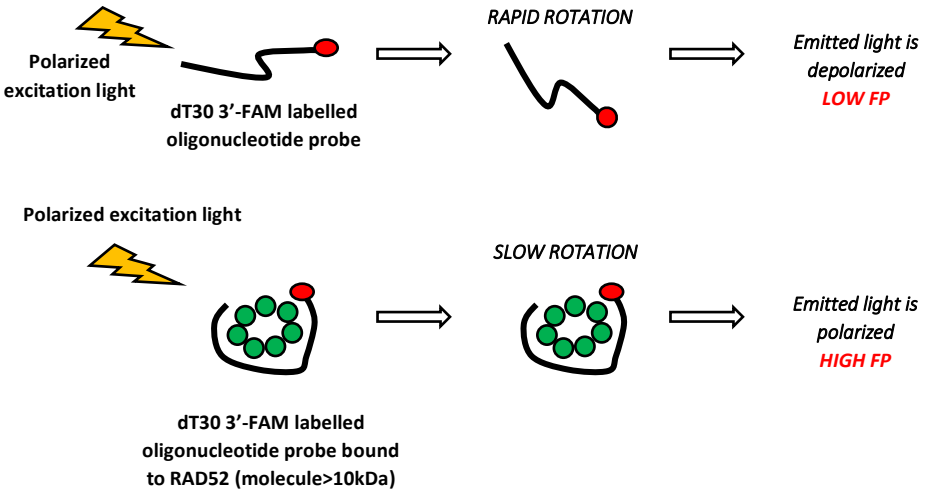


Fig. 21) Schematic representation of FP assay.

3.1.4.13. Static Light Scattering (SLS)

RAD52 FL and RAD52 [1-212] protein samples were analyzed for their oligomerization tendency. Static Light Scattering (SLS) analysis was performed on a Viscotek GPCmax (Malvern, UK) instrument. In our experimental set ups, the detectors were connected with two TSKgel G3000PWxl size-exclusion chromatographic columns (Tosoh bioscience, King of Prussia, PA) in series. The system was equilibrated in 25 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20 and 25 mM Tris-HCl, 500 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20, for RAD52 FL and RAD52 [1-212], respectively. For both the SLS experiments, calibration with BSA was performed. RAD52 FL and RAD52 [1-212] were loaded at 0.8 mg/mL and 0.7 mg/mL, respectively, and isocratically eluted.

3.1.4.14. Dynamic Light Scattering (DLS)

RAD52 FL and RAD52 [1-212] samples were analyzed through DLS to test oligomerization and superstructures formation tendencies. Specifically, RAD52 FL and RAD52 [1-212] proteins were tested in their storage buffer (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT, 0.005% Tween20 and 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT, 5 mM Imidazole, respectively) at 0.8 mg/mL and 0.7 mg/mL, respectively, immediately after protein purification. Sizes of the different samples were analyzed using Zetasizer Nanoparticles Analyzer Software (*Malvern*) at 25°C using the standard operating procedures for size measurements, repeating the measurements scans 13 times for each sample. Collected data were expressed in terms of mass distribution. Reported data are the average of >3 independent experiments.

3.1.4.15. Nuclear Magnetic Resonance (NMR) spectroscopy

NMR experiments were performed in collaboration with dr. M. Veronesi. NMR spectroscopy was used to asses and characterize the fragments and small molecules binding to RAD52 FL. All NMR experiments were recorded at 25 °C with a Bruker FT NMR Avance III and AvanceNeo 600-MHz spectrometer equipped with a 5-mm CryoProbe QCI $^{1}H/^{19}F-^{13}C/^{15}N$-D-Z quadruple resonance with shielded z-gradient coil, and the automatic sample changer SampleJet™ with temperature control. For all sample, a 1D ^{1}H spectrum with water suppression was obtained using the standard NOESY (nuclear Overhauser effect spectroscopy) preset Bruker pulse sequence, with 32 k data points, a spectral width of 20 ppm, 64 scans, an acquisition time of 2.7 s, a relaxation delay (d1) of 4 s and a mixing time of 100 µs. ^{1}H and ^{19}F binding experiments were conducted testing the compounds, in single or in mixture, in the absence and in the presence of RAD52 FL in 25 mM Tris pH 7.5, 250 mM NaCl, 2 mM DTT, 10% D_2O (for the lock signal), 200 µM 3-propionic-2,2,3,3-*d*$_4$ acid (TSP, for the ^{1}H chemical shift reference), 1% DMSO*d*$_6$ and 0.5% Glycerol.

For the ^{19}F NMR fragment screening the compounds were tested in mixtures of 20-25, at the concentration of 20 µM (CF_3 labeled fragments) and 40 µM (CF labeled fragments). The binding of fluorinated compounds, resulting from Virtual Screening (see below), was determined by testing the molecules as single compound in presence of a non-binder (internal negative control) at the concentration of 20 µM (CF_3 labeled compound) and 40 µM (CF labeled compound) in the absence and in the presence of 2 µM RAD52. For each sample a 1D ^{19}F and a T_2 filter experiments were recorded for a total of 32 scans (for CF3 fragment) or 64 scans (CF fragments) with a repetition time of 5 s and proton decoupling during the acquisition period. For the R_2 filter experiments, spin-echo scheme (cpmg) with total τ = 0.2 and 0.4 s were used[163]. The data were

multiplied by an exponential function of 1 Hz prior to Fourier transformation. The reference standard in the ^{19}F spectra was $CFCl_3$.

The binding of compounds resulting from the Virtual Screening without ^{19}F, was assessed by WaterLOGSY and ^{1}H T_2-filter experiments, testing the compounds as single molecule at 100 µM in the absence and in the presence of 2 µM RAD52. The water suppression in both experiments was achieved with the excitation sculpting sequence: the two-water selective 180° Sinc1.1000 pulses and the four PFGs were 2 and 1 ms in duration, respectively. Selective inversion in WaterLOGSY experiments was achieved with a 7.5 ms long 180° Gaussian-shaped pulse[164]. 384 scans with a mixing time of 1.7 s. ^{1}H T_2-filter experiments were achieved using the cpmg sequence with total τ = 1s and 2s, 16 scans and 2s of relaxation delay.

3.1.4.16. MicroScale Thermophoresis (MST)

MST experiments were performed in order to screen hit compounds binding RAD52 FL and to determine, when possible, compounds binding affinities. MST was also used to assess RAD52 FL aggregation tendency.

MST measurements were performed using Monolith NT.115pico instrument (*NanoTemper Technologies*, Munich, Germany). Assays were conducted at 5%-10% (RED dye) LED excitation power and MST power of 40%. Premium capillaries from NanoTemper Technologies were used. Measurements were carried out at 25 °C in the following buffer: 25 mM Hepes pH 7.5, 5% glycerol, 250 mM NaCl, 0.05% tween20.

The recombinant RAD52 FL protein was labeled with the Monolith labeling kit RED-NHS (ammine dye NT-647-NHS) according to manufacturer indications (*NanoTemper Technologies*). Before

MST experiments, to remove aggregates, the labelled protein stocks were centrifuged at 20000 for 10 minutes.

For binding check experiments, proteins were diluted to 10 nM in assay buffer and compounds were tested at different concentrations corresponding to their maximum solubility in buffer and maintaining a final concentration of 1%DMSO in assay buffer.

For binding affinity experiments of RAD52 FL with E5 compound or unlabeled RAD52 FL, the same assay set-up procedure was followed. In this case, the affinity parameter K_d was determined by performing the experiment in parallel on 16 capillaries, each containing a constant concentration of the labelled target (RAD52 FL, 10 nM) and increasing concentrations of unlabeled ligand (E5 compound 30 nM- 1 mM, unlabeled RAD52 FL 0.7 nM – 25 µM). The recorded MST data were then plotted as ΔF_{norm} against the ligand concentration to yield dose-response curves. Experiments were analyzed with MO.Control and MO.Affinity analysis software (NanoTemper Technologies).

3.1.4.17. Bio-Layer Interferometry (BLI)

BLI assays were performed to test aptamers affinities (K_d) for RAD52 protein. Experiments were performed on an Octet K2 (*Sartorious - Pall FortèBio* Corp., Menlo Park, CA, United States) operating at 25°C. Selected assay buffer was RAD52 storage buffer (25 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM DTT) supplemented with 0.05% Tween20. DNA molecules (2 µL/mL) (*atdbio)* were loaded onto streptavidin coated biosensors (Super Streptavidin (SSA)) (*Sartorious - Pall FortèBio*) for 300 s to saturate the sensors. RAD52 protein was used in a wide concentration range (2 µM, 1 µM, 500 nM, 250 nM, 125 nM, 62.5 nM, 31.2 nM). For each protein concentration, after a baseline recorded for 180 s, the association step was recorded for 600 s. Finally,

dissociation step was maintained for 600 s. Dissociation constant (K_d) values were determined by fitting the response intensity (nm) as a function of protein concentration (μM) with the Octet Data Analysis Software.

3.2. Molecular and Cellular Biology

3.2.1. Cell Cultures

BxPC-3 and Capan-1 cells (pancreatic human adenocarcinoma, kindly provided by prof. G. di Stefano (University of Bologna)) were grown in RPMI 1640 (*Sigma-Aldrich*) supplemented with 10% FBS (*Sigma-Aldrich*), 100 U/mL penicillin/streptomycin (*Sigma-Aldrich*), 2 mM glutamine (*Sigma-Aldrich*). All cultures were routinely tested for Mycoplasma contamination. Treatments with olaparib (*Selleckchem*) and E5 compounds were administered in culture medium supplemented with 0.6% DMSO. The same amount of DMSO was added to the control, untreated cultures.

3.2.2. Cell Viability Assay

Cell viability was assessed with the CellTiter-Glo luminescent cell viability assay (*Promega*). For this experiment, 5000 or 10000 cells in 200 μL of culture medium were seeded into each well of a 96-multiwell white body plate and allowed to adhere overnight. After 144 hours-incubation with different concentrations of E5, alone or in the presence of olaparib (10 μM), the plate was allowed to equilibrate at room temperature for 30 min and the CellTiter-Glo reactive was directly added to each well in a 1:1 ratio, following manufacturer protocol. Plate was then kept on a shaker for 10 minutes to induce cell lysis. Samples luminescence was measured after 30 minutes

at room temperature (signal stabilization) with Spark Microplate multimodal reader instrument (*Tecan*).

To evaluate the effect of combination treatments of our compound with olaparib and to assess their interaction index, the following formula was applied:

Combination Index (CI) = (Surviving cells treated with combination) / [(Surviving cells treated with olaparib) X (Surviving cells treated with RAD52 inhibitor)]

As described by Dos Santos Ferreira and Fisher, an additive effect is suggested by a combination index ranging from 0.8 to 1.2. A synergic effect is instead indicated by a CI lowed than 0.8; an antagonist effect by a CI higher than 1.2[165,166].

3.2.3. Immunofluorescence

Immunofluorescence was performed to investigate cellular RAD52 nuclear localization and to evaluate DNA damage through the detection of γH2AX nuclear foci. To visualize RAD52 or γH2AX in cell nuclei, BxPC-3 cells were seeded in an 8-wells glass slide (Merck *Millipore*) (10 000 cells/well) and allowed to adhere overnight. The next day, cultures were preincubated with 400 µM E5 for 1 hour and subsequently exposed to 50 µM cisplatin for an additional 1.5 hour. Medium was then removed, and cells were maintained in the presence of 400 µM E5 and/or olaparib for 6 hours. DMSO was always 0.6% in all treatments. Cells were then fixed in PBS containing 4% formaldehyde for 15 minutes and washed 3 times with PBS. The following day, cells were washed twice with PBS for 5 minutes, permeabilized with 0.1% Triton-X for 15 minutes and washed three more times with PBS for 5 minutes under agitation. Samples were incubated in 3% bovine serum albumin (BSA) in PBS for 1 hour and subsequently exposed to anti- γH2AX

rabbit monoclonal antibody (*Cell Signaling,* 1:500 in 3% BSA/PBS) overnight at 4 °C. The following day, after washing three times, the, samples were incubated with an anti-mouse or anti-rabbit FITC-conjugated secondary antibody (1:1000 in 3% BSA/PBS) for 1 hour at room temperature, washed and stained for 3 minutes with Hoechst 1:1000 in PBS and washed three more times in PBS. For detections, images were acquired using a Leica 6000 fluorescent microscope equipped with filters for FITC, TRITC, and DAPI.

3.2.4. Flow Citofluorimetry

Flow Cytofluorimetry experiments were performed to verify the internalization levels of selected fluorescent aptamer in cells. For this experiment, 250 000 cells in 1000 µL of culture medium were seeded into each well of a 12-multiwell plate and allowed to adhere overnight (reaching 80% confluency). The following day, the transfection mixtures were prepared. Briefly, DNA-Lipofectamine2000 reagent (*Thermofisher*) was mixed with Opti-MEM medium in 1:25 ratio, incubating for 5 minutes (Master Mix). Fluorescent Labelled 5'-Texas Red- Apt1 250 µg/mL (*atdbio)* was 1:50 diluted in Opti-MEM Medium to a final concentration of 5 µg/mL (Aptamer solution). Master Mix solution and Aptamer solution / Negative control were mixed 1:1 and incubated for 20 minutes at room temperature reaching 2.5 µg/mL aptamer concentration. When the transfection solution was ready, old cell medium was removed from all the plate wells and was then replaced with 800 µL RPMI 1640 supplemented with glutamine 2mM and 10% FBS. Finally, 200 µL of aptamer or negative control solution was added to each well drop by drop (0.5 µg/mL final aptamer concentration in each well). The plate was incubated overnight at 37°C. The following day, after three washing steps in PBS, each cellular solution was collected in a final

volume of 500 µL of cold PBS, quickly vortexed and filtered. Analyses were performed with cytofluorimeter SH800S (*Sony Biotechnology*).

3.3. Computational Methods

3.3.1. Small Molecules RAD52 Virtual Screening Campaign

3.3.1.1. Site identification

With the aim of identifying small-molecules able to interact and potentially hinder RAD52 protomer-protomer interaction, our computational group (dr. D. Gioia, IIT Genova) analyzed the protein interface to identify residues critical for the oligomerization ("hotspots") and small druggable pockets suitable for targeting[167].

Two different computational methods were exploited (SiteMap and Alanine Scanning Mutagenesis) that allowed to explore protomer-protomer interface and to converge towards a rational choice. The computational model consisted of the N-terminal domain of 2 adjacent protomers, obtained selecting chain A and chain B of the PDB 1KN0. The structure was treated with the Schrödinger *Suite* 2018-4 Protein Preparation Wizard tool. All water molecules were removed and, after sampling residues protonation state with the PROPKA utility, the optimization of the hydrogen-bonding network was performed. The resulting structure was then subjected to a restrained minimization step using the OPLS3 force field and a convergence threshold of 0.3 Å in order to relief any potential steric clash.

SiteMap is a tool of the Schrödinger Suite that analyzes the protein surface in the search of druggable sites, i.e. pockets and regions with a good balance between hydrophobic groups and hydrogen-bond acceptors and donors[168]. The analysis was performed on both chains and

evidenced three potential binding sites ranked according to a proprietary scoring function (SiteScore). Among them, only one was found lying at the interface between monomers (Fig. 22). Specifically, it relates to a region where a small loop, formed by residues Phe79, Gly80, Tyr81, Asn82 of one monomer, lies in a cavity of the adjacent monomer (shaped by the Tyr31, Tyr36, Ile39, Leu43, Phe79, Leu115, Asp117, Ser119, Phe158 residues) with a score of 0.74. It is indeed worth underlying that this site has already been identified from Cramer-Morales and colleagues in 2013 for F79 inhibitor design and development[55].

This initial suggestion was confirmed by the subsequent virtual Alanine Scanning Mutagenesis, as implemented in Schrödinger *Maestro*, which allowed to identify residues whose mutation to Alanine had a large impact on protein stability and/or subunits affinity[169]. Systematically, each residue of one of the two monomers was mutated to Alanine and the inherent free energy difference was evaluated using the Prime MM-GBSA approach, where standard Molecular Mechanics (MM) energy terms are coupled to the polar and non-polar contributions to the solvation free energies[170]. Here, the polar contribution was evaluated using a Generalized Born (GB) model and the non-polar one was estimated from a linear relation to the solvent accessible surface area (SASA). The mutation with the largest variation in the monomer-monomer affinity was the Tyr81Ala, with a value of 26.37 kcal/mol.

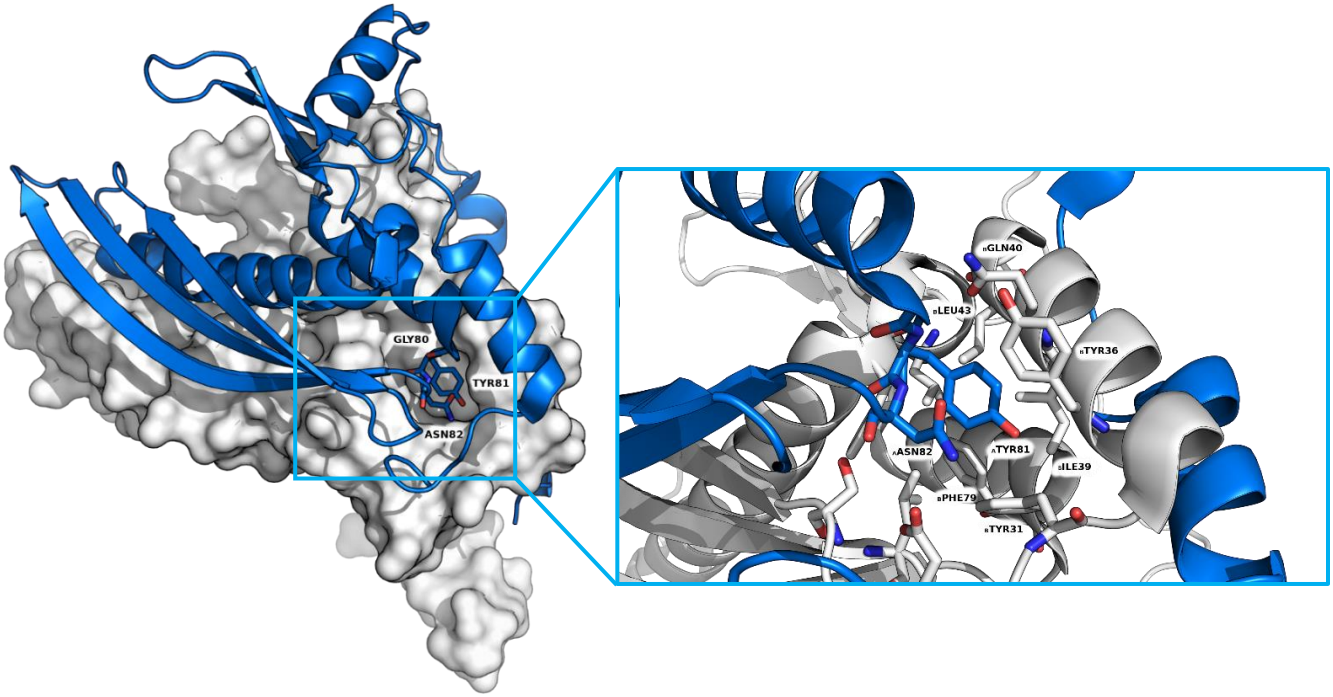


Fig. 22) Schematic representation of Tyr81 pocket of RAD52 identified using the computational tool SiteMap.

3.3.1.2. Virtual Screening

All the virtual screening experiments were performed by dr. D. Gioia (IIT Genova).

Database Preparation. An internal library of commercially available compounds was prepared with the LigPrep utility of the Schrödinger Suite. Starting from the 2D information, the 3D structures and all the possible stereoisomers were generated. The resulting molecules were then submitted to Epik and all the tautomers and ionization states at pH 7.0 ± 1.0 were evaluated. At last, duplicates, compounds with more than two chiral centers, and Pan-Assay Interference Compounds (PAINS) were discarded obtaining a final set of 100k of compounds.

High-Throughput Docking. The docking was conducted using Glide taking advantage of the Standard Precision (SP) scoring function[171]. The binding site was defined centering the box grid

on the center-of-mass of Tyr81 of chain A. Grid generation was performed using default parameters with a box dimensions set to 30 Å for all the axes. Regarding docking calculation, the sampling for ligand conformer generation was enhanced by 2 times and the enhanced sampling for the selection of initial poses was selected. 1000 top-scored molecules were subjected to visual inspection to discard scaffolds not adequately fitting the protein site; molecules with hydrophobic regions at solvent interface and with either excessively complex or relatively small structures were also discarded. This led to the selection of 15 molecules (an example is reported in Fig. 23).

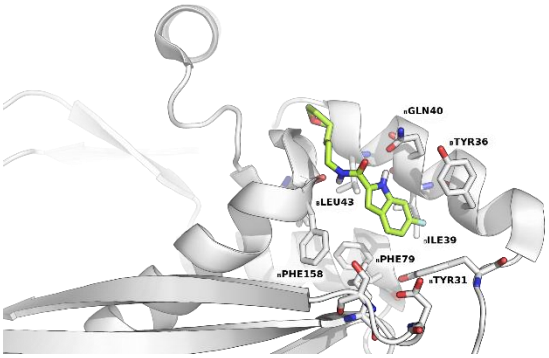


Fig. 23) Schematic representation of one of the virtual screening molecules inside pocket Tyr81 of RAD52.

3.3.2. In silico aptamers design

DNA aptamers targeting RAD52 were designed by prof. G. Tartaglia and dr. A. Armaos (IIT Genova) using catRAPID® software. Briefly, crystal structure of RAD52 [1-212] (1KN0) was used to identify the amino acidic residues required for DNA-binding. When these amino acids were identified, the protein fragments of 51 amino acids from RAD52 sequence centered on these positions were extracted. As control, protein fragments of Actin B, which is a protein highly expressed in most cell lines, were used. These fragments were selected to have the same length as the RAD52 fragments (51aa).

Protein-DNA Interactions were computed using the catRAPID algorithm, which is a sequence-based predictor that takes into account physicochemical properties as well as secondary structure features to estimate the interaction scores between a protein and an RNA sequence (here, catRAPID algorithm was used for DNA sequences design)[172].

Oligonucleotides sequences were identified calculating interaction propensities of the custom RAD52 fragments against RNA fragments from ~20K human mRNAs and ~20K human lncRNAs. RNA fragments with a catRAPID score > 25 were selected as top interacting and RNA fragments with a catRAPID score < -25 as low interacting. The aim was to find RNA fragments with high interaction propensity against RAD52 and use these RNA fragments to extract RNA motifs that would guide us to design the aptamers.

DREME software was then used to extract motifs highly enriched in the top interacting oligonucleotides fragment list compared to the less interacting ones (*Configuration of DREME: minimum motif length: 4, maximum motif length: 16)*. Based on these enriched motifs found using DREME, a list of potential DNA aptamers was built: for short motifs of 4-8 nucleotides, aptamers sequences were built repeating the motif up to ~ 16 nucleotides; for longer motifs, the sequence was used as it was.

Interaction scores of the aptamer list were than calculated against the custom RAD52 fragments (each aptamer against each one of RAD52 fragment). For internal control, similarly, the interaction scores of the aptamers against the 13 ACTIN B control fragments were calculated.

Finally, sequences that obtained catRAPID interaction scores against RAD52 fragments higher than the mean + 0.5*SD of the control scores were considered potential aptamers.

4. Results

4.1. Structural Studies

4.1.1. Optimization for protein expression and purification

In order to investigate the structural and biophysical properties of RAD52, both RAD52 full-length protein (FL) and RAD52 [1-212] were expressed and purified. Notably, a part from RAD52 FL form, we decided to express and study RAD52 [1-212] to have a better understanding of the protein N-terminal domain, involved in DNA interaction and protomer-protomer interaction. Many are the examples, already reported in the literature, in which this portion of the N-terminal domain is used, given its adaptability for crystallization experiments and, above all, given its DNA binding mode comparable to the one of the full-length protein[61,69,71]. Indeed, in this thesis the N-terminal RAD52 portion was also taken into consideration as a reference tool for some experiments.

The pET16b-RAD52 FL and pET16b-RAD52 [1-212] vectors were transformed in *E. Coli* Rosetta (DE3)pLysS cells as already mentioned in the *material and methods* section. Similar vectors and E.coli strains have been already reported in the literature for successful expressions of recombinant RAD52[61,69,73]. Both proteins were expressed with a 6x Histidine Tag at the N-terminus to support RAD52 purification procedures. The expression protocols for both 6xHis-RAD52 FL and 6xHis-RAD52 [1-212] (from now on named RAD52 FL and RAD52 [1-212], respectively) were optimized setting up a series of expression trials to identify the conditions able to provide the highest yield of soluble proteins. Expression trials are reported In Fig. 24 and 25. The selected conditions for both proteins envisaged protein induction at 0.8 of bacterial optical density (OD_{600}) (late-log phase), with 0.8 mM IPTG at 25 °C for 4 hours.

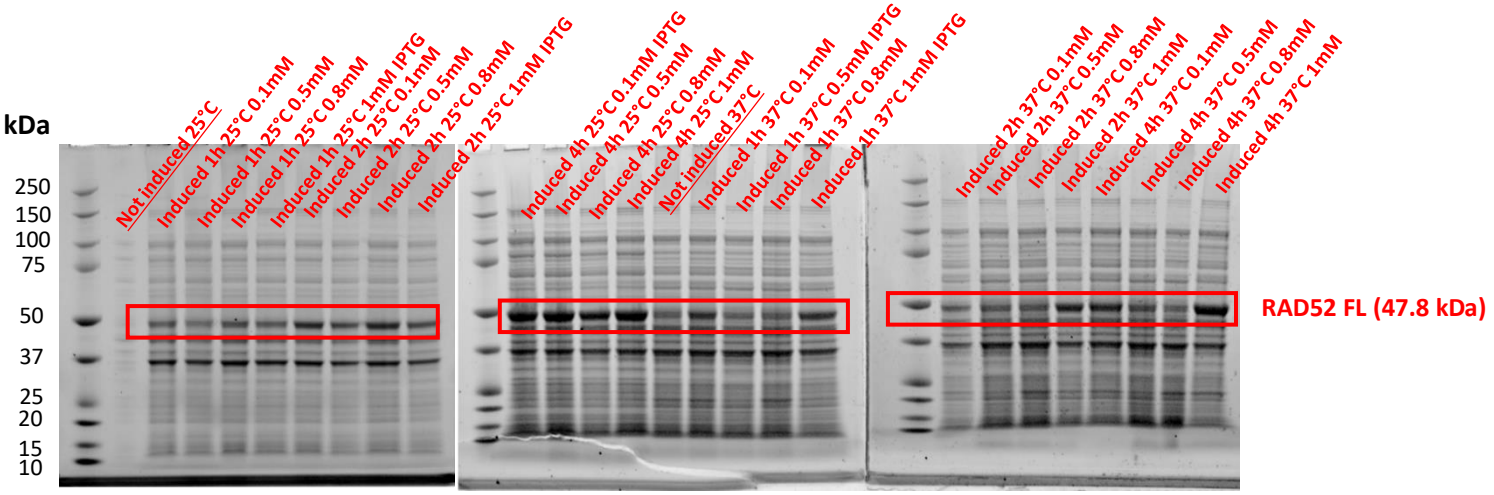


*Fig. 24) SDS Page analysis of RAD52 FL expression in different conditions. Experiments were performed at 25°C or 37°C; IPTG concentration tested were 0.1 mM, 0.5 mM, 0.8 mM, 1 mM; the expression was measured after 1h, 2h, 4h; the volume of bacterial culture to load on the gel was determined using the following equation: 1/OD*0.5.*

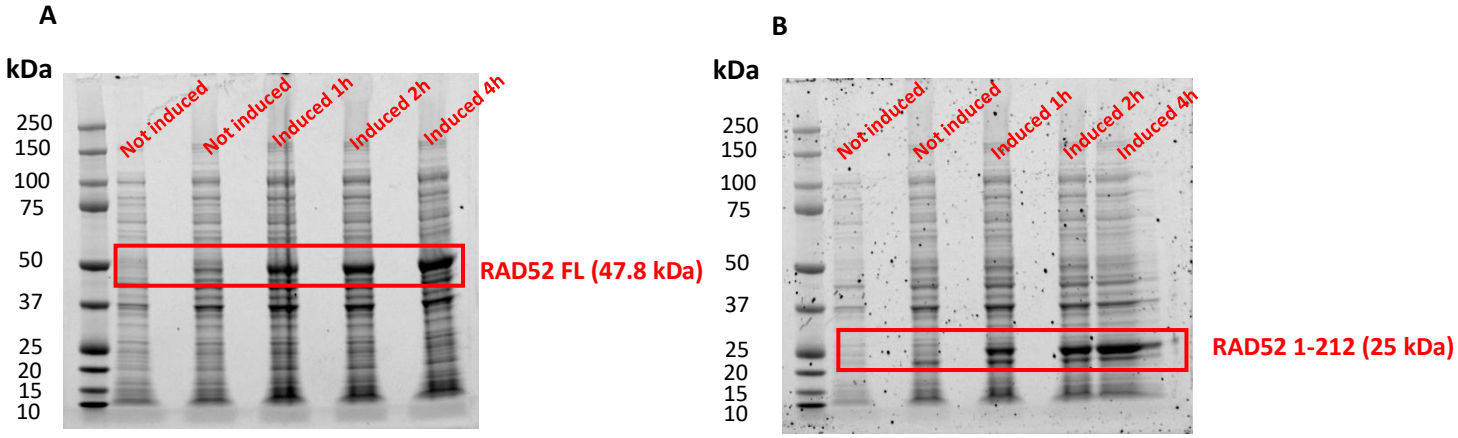


Fig. 25) SDS Page analyses of RAD52 FL (A) and RAD52 [1-212] (B) expression in the selected conditions.

The protocol for RAD52 FL purification was optimized using three chromatographic steps: a His-Trap column, a desalting column and a HiTrap Heparin HP. Examples of elution chromatographic profiles are reported in Fig. 26. Samples collected from each purification step were analyzed

through SDS-Page gel and Western Blot as reported in Fig. 27. Final protein yield was about 10 mg for 500 mL of bacterial culture.

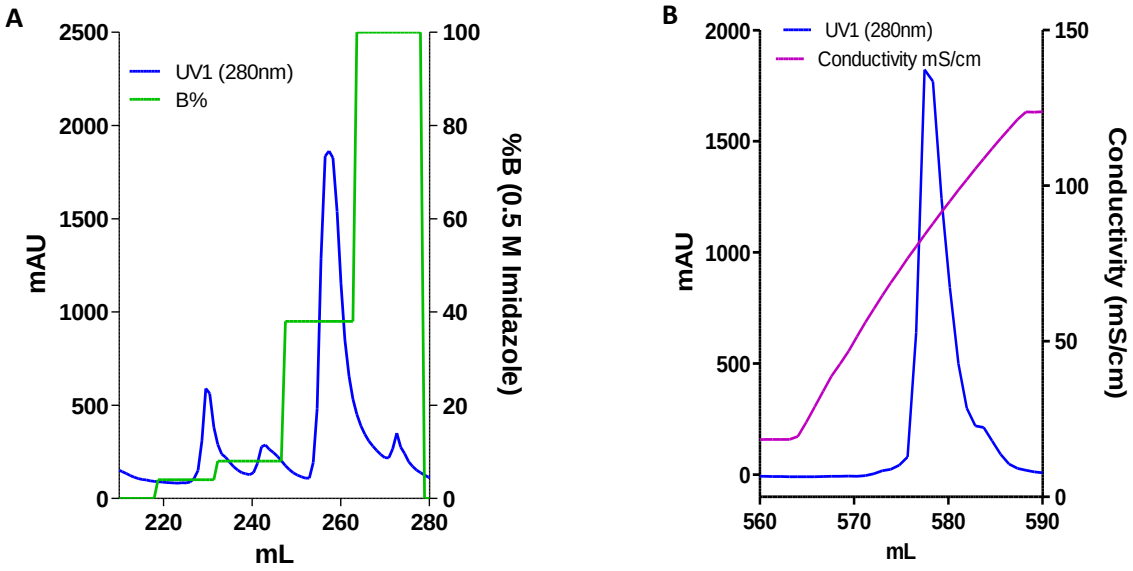


Fig. 26) Chromatogram profiles of RAD52 FL purification steps HisTrap (A) and Heparin affinity column (B); UV_{280} are reported in blue, B% buffer in green and conductivity in purple

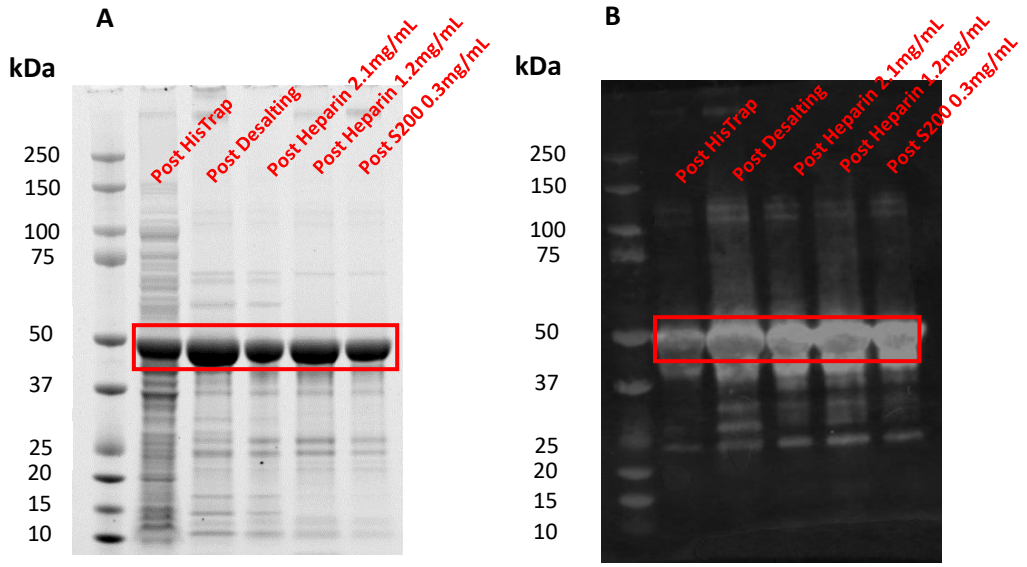


Fig. 27) SDS Page (A) and Western Blot (B) (1:1000 Abl antiHis Tag) analysis of purification protein pools.

The protocol for RAD52 [1-212] was optimized using one single His-Trap column. The relative chromatographic profile is reported in Fig. 28. Samples from each purification step were analyzed with SDS-Page gel and Western Blot as reported in Fig. 29. Final protein yield was about 17 mg for 500 mL of bacterial culture.

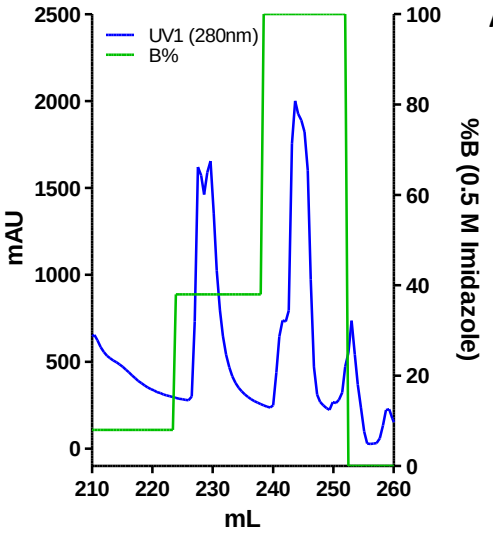


Fig. 28) Chromatogram corresponding to RAD52 [1-212] HisTrap elution profile.

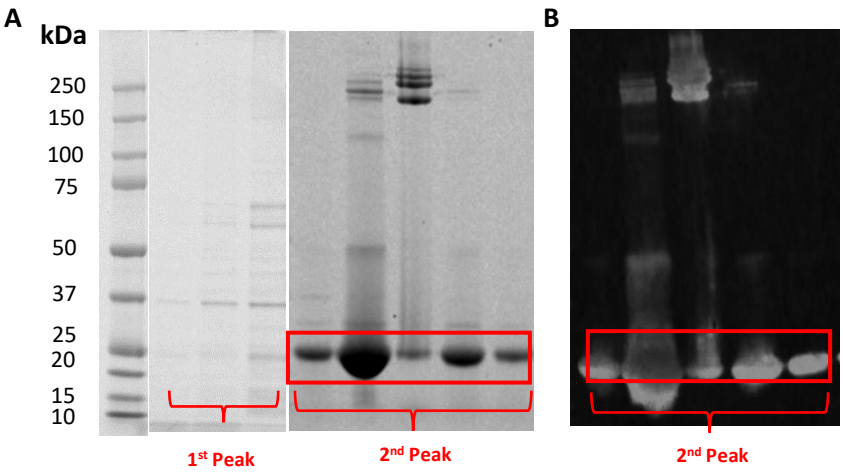


Fig. 29) SDS Page (A) and Western Blot (B) (1:1000 AbI antiHis Tag) analysis of purification protein pools.

4.1.2. Protein Characterization

4.1.2.1. LC-MS

LC-MS analyses (performed by dr. N. Liessi (IIT Genova)) were performed to evaluate the identity and purity of the purified recombinant proteins (Fig. 30). Both RAD52 FL and RAD52 [1-212] showed a >90% purity. LC-MS deconvoluted proteins spectra showed observed masses (47975 Da and 25207 Da for RAD52 FL and RAD52 [1-212], respectively) consistent with RAD52 FL and RAD52 [1-212] primary sequences, plus an initial methionine residue, a 6xHis Tag and a TEV protease cleavage site (ENLYFQG/S).

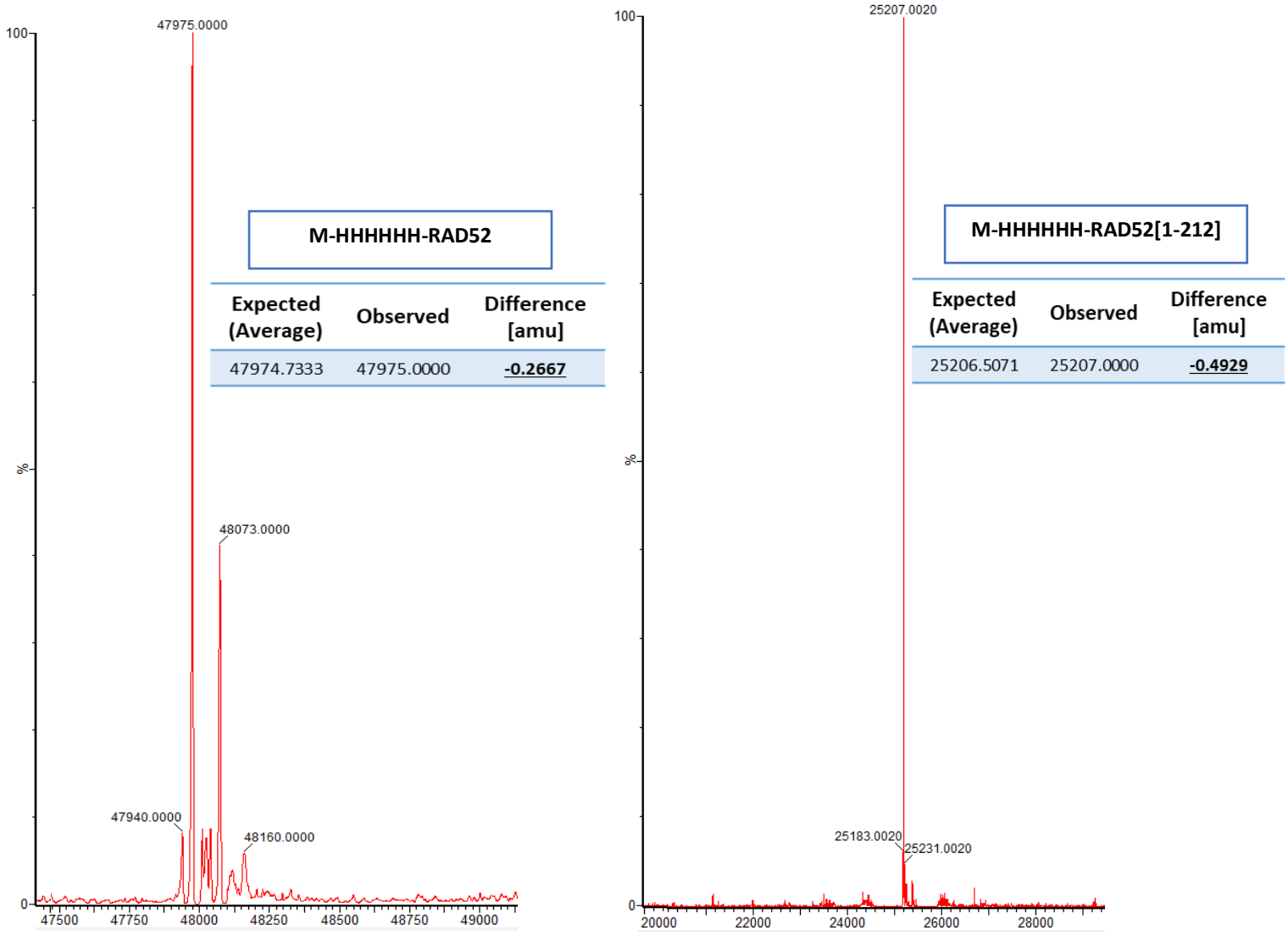


Fig. 30) LC-MS profile corresponding to RAD52 FL purified sample and RAD52 [1-212] purified sample.

4.1.2.2. Size Exclusion Chromatography (SEC)

RAD52 FL and RAD52 [1-212] recombinant proteins are expected to exist in ring-shaped oligomeric forms, formed by 7 and 11 monomers, respectively as reported in the literature[61,64] . To evaluate and compare the oligomerization states of the proteins, purified RAD52 FL and RAD52 [1-212] were concentrated and loaded on a Superdex200 column. The RAD52 FL elution profile of the size exclusion column showed an elution peak at 9.5 mL, whereas the elution profile of RAD52 [1-212] showed two peaks, one at 10.2 mL and the other one at 11.8 mL (Fig. 31A). Samples of the eluates were afterwards loaded on a SDS-Page (Fig. 31B).

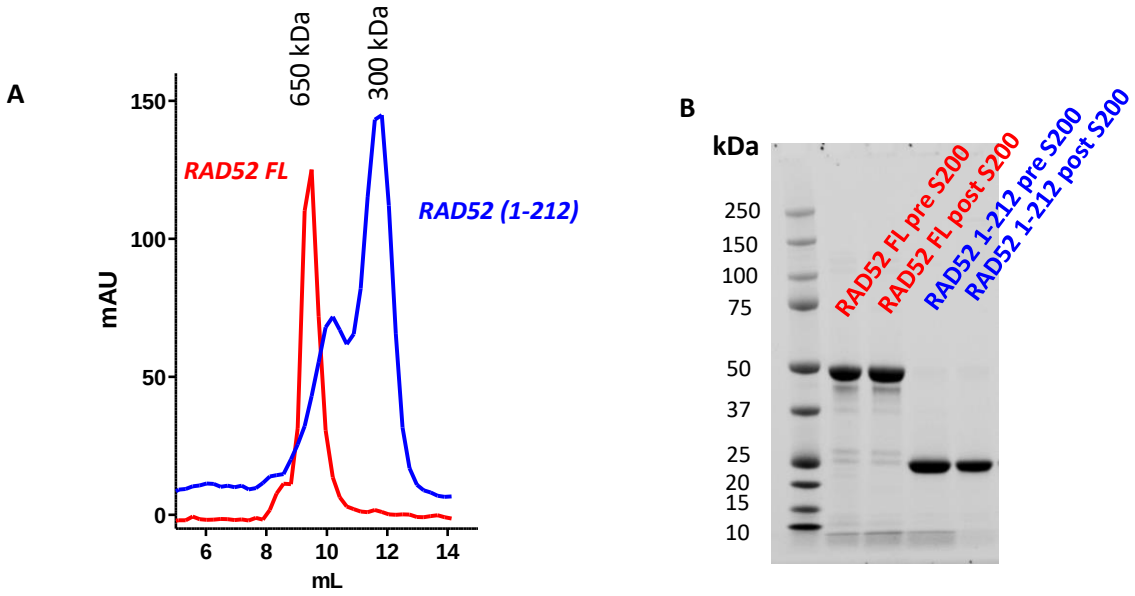


Fig. 31) A) Chromatogram corresponding to RAD52 FL (red) and RAD52 [1-212] (blue) S200 increase elution profile (size exclusion chromatography); B) SDS-Page analysis of the samples before and after gel filtration.

According to a rough estimation provided by the calibration of the Superdex 200, size exclusion column and by the markers of the SDS-page gel, RAD52 FL eluted at a volume corresponding to about 660 kDa, whereas RAD52 [1-212] eluted at a volume corresponding to about 300 kDa.

Considering that the two proteins should have a ring-shaped conformation, from the calculated molecular weights we can suggest that while RAD52 [1-212] is stable in a 1 undecameric ring structure, RAD52 FL oligomeric ring structures should be organized in superstructures formed by ~2 heptameric rings.

4.1.2.3. Secondary structure

A rough estimation of the secondary structures of the two proteins RAD52 FL and [1-212] was obtained through Circular Dichroism (Appendix 7.1). The two CD spectra are reported in Fig. 32[173]. Even though at first glance the CD profile of the two proteins looked similar, data analyses through *Dichroweb* software allowed to assess that RAD52 FL, compared to RAD52 [1-212], showed a loss of secondary structure composition to the advantage of an increase in disordered regions.

From data analyses RAD52 FL and [1-212] have similar absolute composition in α-helices (9% and 15%, respectively) while RAD52 FL has more disordered regions (51% and 32%, respectively), a reduction in β-sheets composition (7% and 32%, respectively) and an increase in turns composition, with respect to RAD52 [1-212] (31% and 22%, respectively). Data percentage should be normalized to the absolute numbers of amino acids present in the primary sequences of the two proteins (431 AA and 225 AA, respectively).

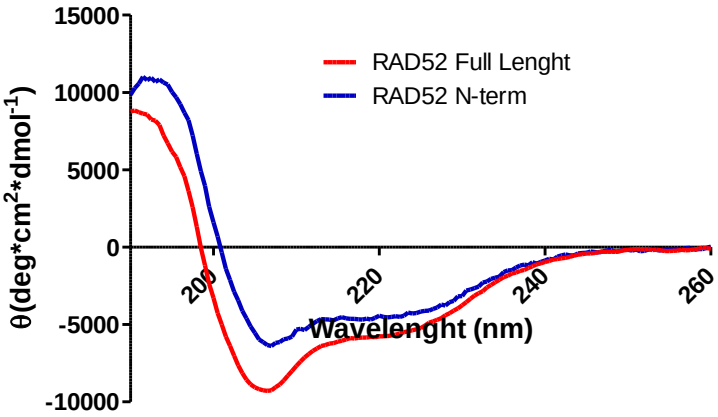


Fig. 32) Secondary structure determination: comparison of CD spectra of RAD52 FL (red) and RAD52 [1-212] (blue); data analyses were performed after normalization; Dichroweb software (Birkberg college; University of London) was used to predict the secondary structure information.

4.1.2.4. Thermal Stability

Proteins thermal stability was investigated through Circular Dichroism experiments. Protein α-helices have a recognizable CD spectrum with a minimum at 222 nm. The effects induced by temperature increase on protein structure (i.e. thermal stability) are detected by recording the changes in CD values at 222 nm (i.e. changes in the structural state of the protein) as a function of the increasing temperature. From these experiments, RAD52 FL and RAD52 [1-212] showed similar melting temperatures T_m of 50.0 ± 0.6 °C and 49.9 ± 0.3 °C, respectively. This similarity can be ascribed to similar α-helices composition of the two proteins, since the N-terminal domain of RAD52 is the most structured portion of the protein, rich in α-helices and β-sheets (Fig. 33).

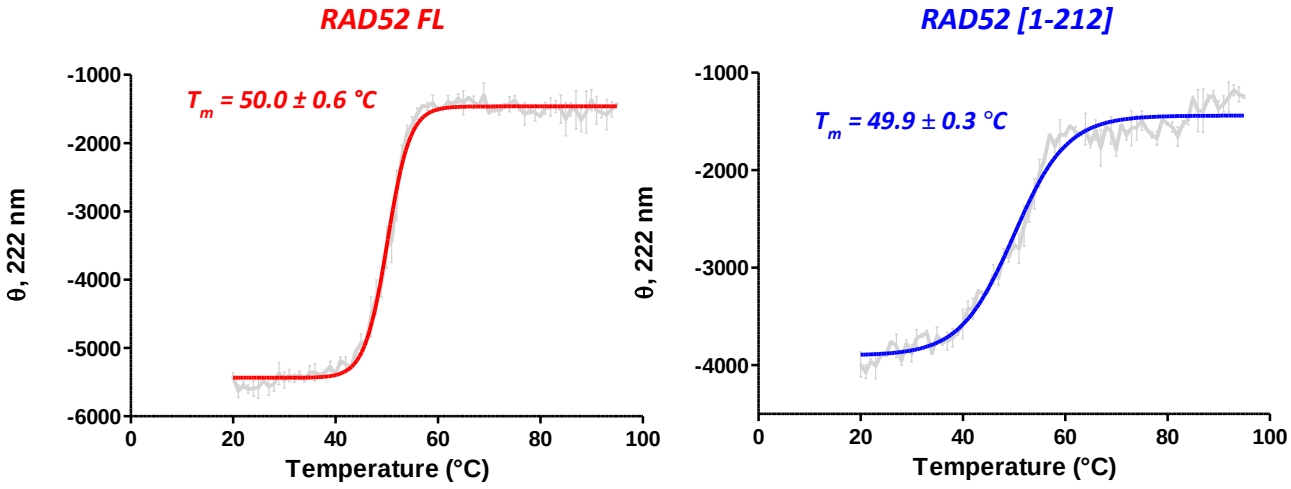


Fig. 33) Thermal stability analysis: comparison of ϑ at 222 nm of RAD52 FL (red) and RAD52 [1-212] (blue) for melting temperature (T_m) determination; data analyses were performed after normalization.

4.1.3. RAD52 oligomerization and superstructures formation

In order to further characterize the purified proteins in terms of samples heterogeneity and propensity to form high MW superstructures, Native Gel electrophoresis, SLS, DLS, and MST experiments were performed.

4.1.3.1. Native Gel Electrophoresis

Native Gels electrophoresis of RAD52 FL and RAD52 [1-212] were initially performed to have a rough estimation of the oligomerization and superstructures pattern of the two samples. Native gel of RAD52 FL samples, reported in Fig. 34A showed the coexistence of several bands at high MW (850kDa, ~1300kDa, >1600kDa), therefore suggesting that the heptameric ring formations may assemble in higher molecular weight superstructures constituted of several rings.

On the contrary, for RAD52 [1-212] native gels showed a one-band migration pattern at roughly 480 kDa (Fig. 34B).

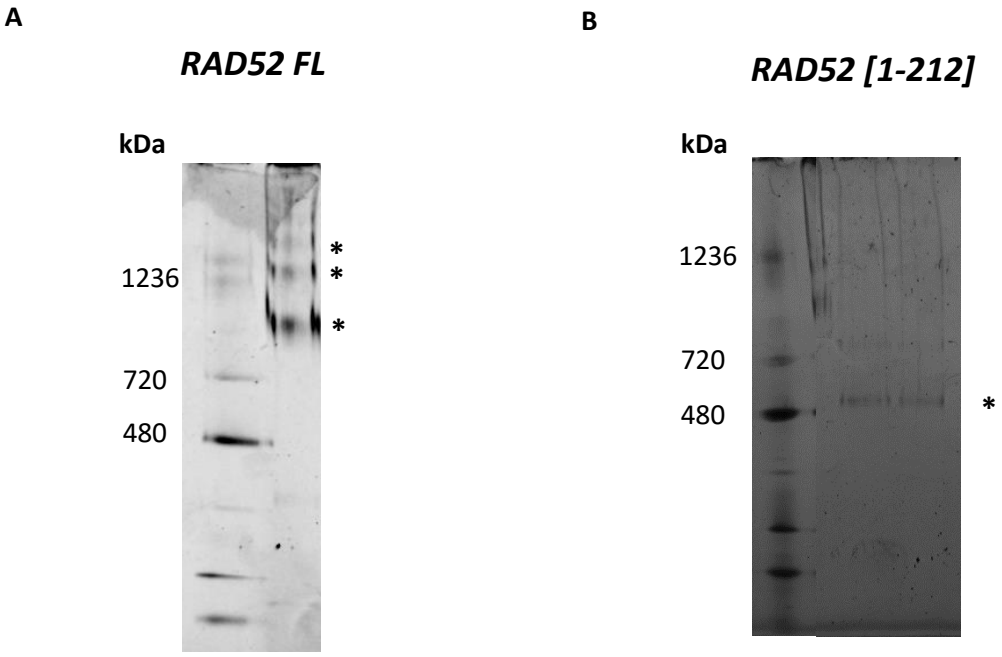


Fig. 34) Native gel electrophoresis showing oligomeric forms and high MW superstructures of RAD52 FL (A) and RAD52 [1-212] (B); RAD52 FL shows different bands corresponding to 850 kDa, ~1300 kDa, >1600 kDa and RAD52 [1-212] shows a discrete band corresponding at 480 kDa.

4.1.3.2. Static Light Scattering (SLS)

Oligomerization and superstructure formation tendency of RAD52 FL and RAD52 [1-212] were further analyzed by Static light scattering (Fig. 35). In SLS, the average scattered intensity of a population of particles in solution is measured by integrating the scattered signal over a period of time. SLS allows to obtain information about size and MW of the particles in solution. SLS data of RAD52 FL samples showed a heterogeneous population with an estimated average MW of 700 kDa (roughly 2 heptameric ring units). However, due to the high heterogeneity of these samples, even though the average data are in agreement with the above reported analyses (size exclusion

chromatography and SDS-Page gel/native gel), they are only a rough estimation of the molecular weights, given also the complexity of proteins superstructure organization.

RAD52 [1-212] showed, on the other end, a more homogeneous SLS profile, with an average MW of 350 kDa about 1 undecameric unit, in agreement with the above reported data.

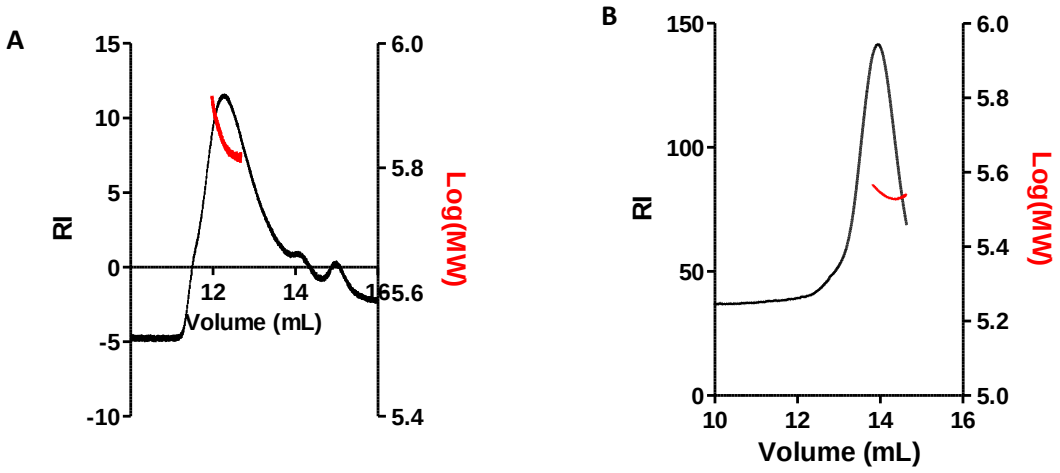


Fig. 35) SLS protein dimension determination: SLS chromatogram representing the dimensions of the different protein species in the purified protein sample of RAD52 FL (12.23 retention volume) (A) and RAD52 [1-212] average (13.94 mL retention volume) (B) corresponding to 700 kDa and 350 kDa on, respectively.

4.1.3.3. Dynamic Light Scattering (DLS)

Further information were obtained from the Dynamic Light Scattering analysis of RAD52 FL and [1-212] samples. In DLS the scattering fluctuations allows to calculate the polidispersivity (Pd) of samples and the hydrodynamic radius (HyR) of each sample species in terms of mass distribution. Moreover, assuming a general globular conformation of the proteins, it is also possible to extrapolate an indicative MW. From our analyses, both RAD52 FL and [1-212] (0.8 mg/mL and 0.7 mg/mL, respectively) showed a polidisperse profile of the sample (Fig 36A and Fig. 36B). However, also in this case, RAD52 FL showed higher MW superstructures in comparison with RAD52 [1-212], suggesting a higher propensity of the full-length protein to form superstructures

in comparison with the truncated form. The Pd of RAD52 FL and [1-212] was 57.2 ± 9.7 % and 71.5 ± 2.5 %, respectively. The most abundant species of RAD52 FL had the HyR of 10.52 nm corresponding to 2.5 heptameric ring units (828.9 kDa). The most abundant species of RAD52 [1-212] had the HyR of 6.77 nm corresponding to ~ 1 ring unit (295.5 kDa).

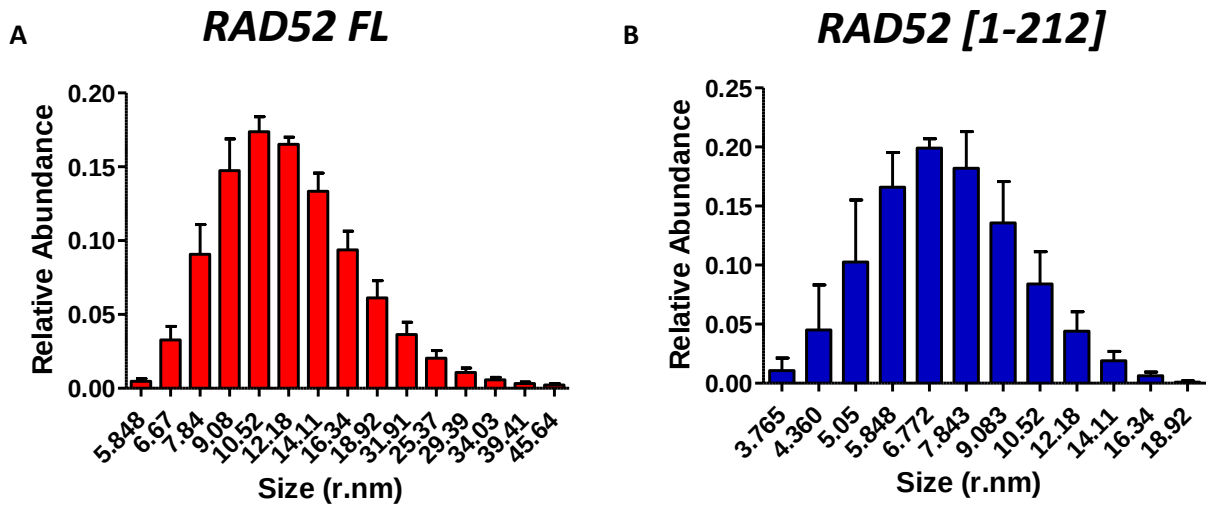


Fig. 36) DLS graph representing the hydrodynamic radius of the different protein superstructures in the sample, with the main species at 10.5 nm of hydrodynamic radius, corresponding to ~830kDa for RAD52 FL (A) and at 6.77 nm, corresponding to ~295kDa for the N-terminal [1-212] RAD52 (B).

4.1.3.4. MicroScale Thermophoresis (MST)

MST experiments (Appendix 7.3) were then used to investigate the kinetic of RAD52 FL high MW superstructures formation. Fluorescently-labelled RAD52 FL was titrated with increasing concentrations of non-labelled RAD52 FL to identify a binding curve able to describe the propensity of RAD52 to form high MW structures. The results showed a propensity of RAD52 FL, already in a heterogeneous high MW assembly, to form superstructures in a concentration-dependent fashion, with an apparent binding affinity (K_d) of 14.2 ± 4.0 nm (Fig. 37).

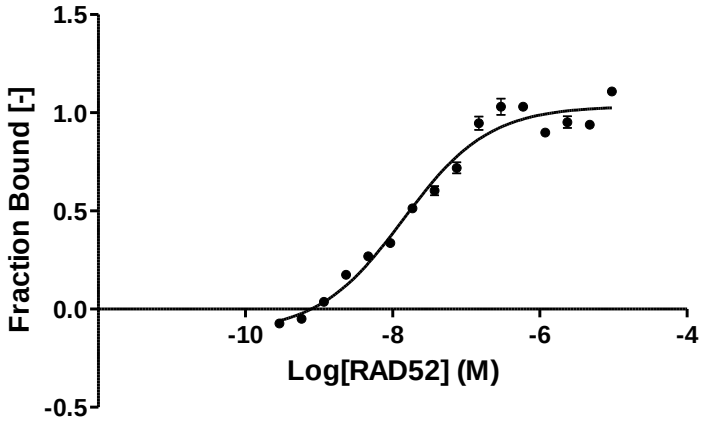


Fig. 37) MST graph of protein-protein titration reporting RAD52 full length tendency to form high MW superstructures in a dose-dependent fashion.

Aggregation studies were all in agreement in assessing RAD52 FL and RAD52 [1-212] propensity to form high MW superstructures. According to literature data RAD52 FL and RAD52 [1-212] exist as heptameric and undecameric ring units, respectively. RAD52 FL tends to form higher MW superstructures, formed by multiple rings units, in a dose-dependent manner (MST experiments). Notably, even though also RAD52 [1-212] tends to form high MW superstructures, this propensity seems to be lower compared to RAD52 FL. These data suggest that the C-terminal domain of RAD52 has a critical role in the formation of superstructure arrangements.

4.1.4. DNA-Protein Interaction

RAD52 is a DNA/RNA binding protein, which, among several other roles, assists annealing and homology search of DNA strands[58,84,174,175]. In order to further understand RAD52 and its activity we characterized the DNA-RAD52 interaction. Fluorescence Resonance Energy Transfer, Electrophoretic Mobility Shift Assay and Fluorescence Polarization experiments were set up to pursue this goal.

4.1.4.1. Fluorescence Resonance Energy Transfer (FRET)

Using a dually labelled 30 bp ssDNA with cyanine5 and cyanine3 fluorophores located at the 3' and 5', the DNA wrapping around RAD52 ring was detectable exploiting the FRET phenomenon (*materials and methods*). Here, RAD52 FL was incubated with dually labelled ssDNA to measure FRET intensity that changed depending on the proximity of donor and acceptor fluorophores. The fluctuations of E_{FRET} values were analyzed as a function of protein concentration. From this experiment, RAD52 was confirmed to bind ssDNA (Fig. 38). Moreover, higher RAD52 concentrations were reported to limit E_{FRET} as already described in literature[74,79]. The highest FRET signal intensity detected in the presence of 1 nM labelled-ssDNA was at 5 nM RAD52 concentration. This condition was later used for further drug discovery inhibition assays.

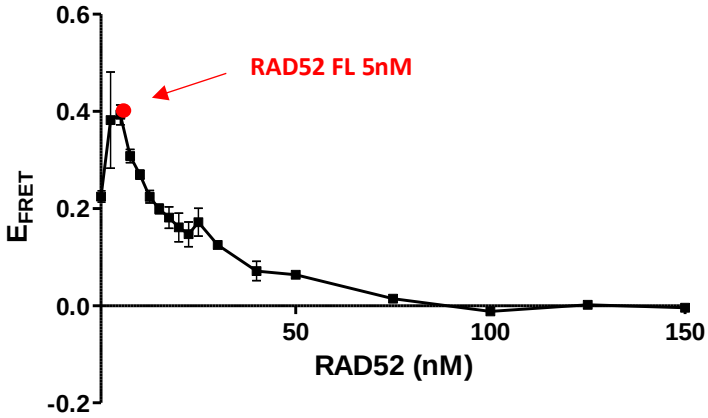


Fig. 38) FRET ssDNA-RAD52 binding assay; experiments were performed titrating increasing concentrations of RAD52 FL in a solution containing 1 nM cy3-cy5 labelled ssDNA; FRET signal decrease when RAD52 concentration is >> ssDNA concentration.

4.1.4.2. Electrophoretic Mobility Shift Assay (EMSA)

EMSA experiments were set up to corroborate DNA-RAD52 binding[73]. RAD52 FL at different concentrations was incubated with fluorescent labelled ssDNA (10 nM dT_{30}-cy5) and the obtained

samples were run on an agarose gel (as described in the *material and methods* section). Upon increasing RAD52 FL concentrations, the fluorescent band of the unbound ssDNA showed a decrease in intensity while a higher molecular weight band progressively appeared, as reported in Fig. 39, suggesting that RAD52- ssDNA binding took place. 200 nM RAD52 concentration was selected as on optimized condition to perform drug discovery inhibition EMSA assays as described in the following paragraphs.

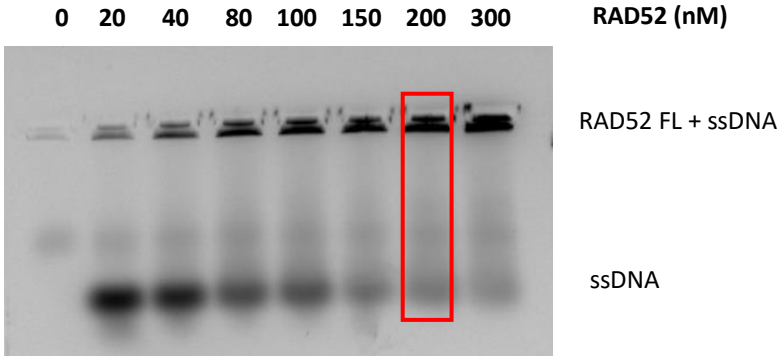


Fig. 39) DNA-RAD52 interaction measurement through EMSA assay; fluorescent labelled ssDNA bands signals decrease upon addition of increasing concentration of RAD52. Complex ssDNA-RAD52 bands are visible in the upper part of the gel and their signals increase upon addition of RAD52 protein in a concentration-dependent fashion.

4.1.4.3. Fluorescence Polarization (FP)

While FRET and EMSA assays qualitatively assessed the interaction of RAD52 with DNA, Fluorescence Polarization (FP) was used to quantitatively characterize this interaction. As described in the *materials and methods* section, variations in fluorescence polarization signals of a fluorescent labelled ssDNA (10 nM 6FAM-dT_{30}) were detected in the presence of increasing RAD52 FL concentrations. The detected FP variations, proportional to the protein –DNA interactions, were used for quantitative analyses. FP data elaborations allowed to determine a

binding affinity parameter (K_d) of 37.2 ± 0.4 nM for DNA binding to RAD52 (Fig. 40). Notably, different incubation times of protein and labelled DNA did not affect the calculated K_d value.

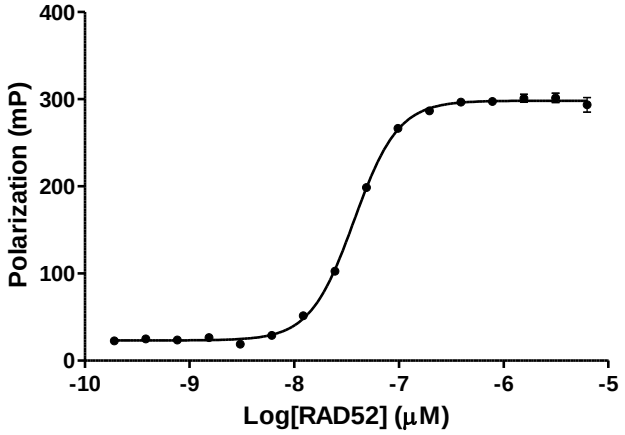


Fig. 40) FP experiment for measuring ssDNA-RAD52 interaction affinity. The plot shows polarization values (mP) in function of Log[RAD52]. Experiments were performed titrating increasing RAD52 concentrations in a solution containing 10 nM 5' -FAM labelled ssDNA.

4.1.5. Atomic Force Microscopy (AFM)

AFM is a scanning probe microscopy which is based on the use of a mechanical probe that touch and scan a sample surface. This technique is able to follow the surface composition and build a three-dimensional image of the analyzed sample measuring the interaction between its tip and the sample profile surface. AFM experiments were used to validate our data on RAD52 superstructures formation and to have further information on the RAD52 superstructures state through a rough 3D reconstruction. Results are reported in table 1 and the relative sample images are in Fig. 41.

Table 1

	RAD52 (70 nM)	RAD52 (350 nM)	RAD52 70 nM + DNA (62 nM)
Height	**1.55 ± 0.53 nm**	**3.18 ± 0.73 nm**	**27.62 ± 6.62 nm**
Diameter 1	43.77 ± 12.38 nm	38.30 ± 6.09 nm	130.84 ± 14.54 nm
Diameter 2	30.85 ± 6.70 nm	28.45 ± 6.20 nm	106.08 ± 16.95 nm

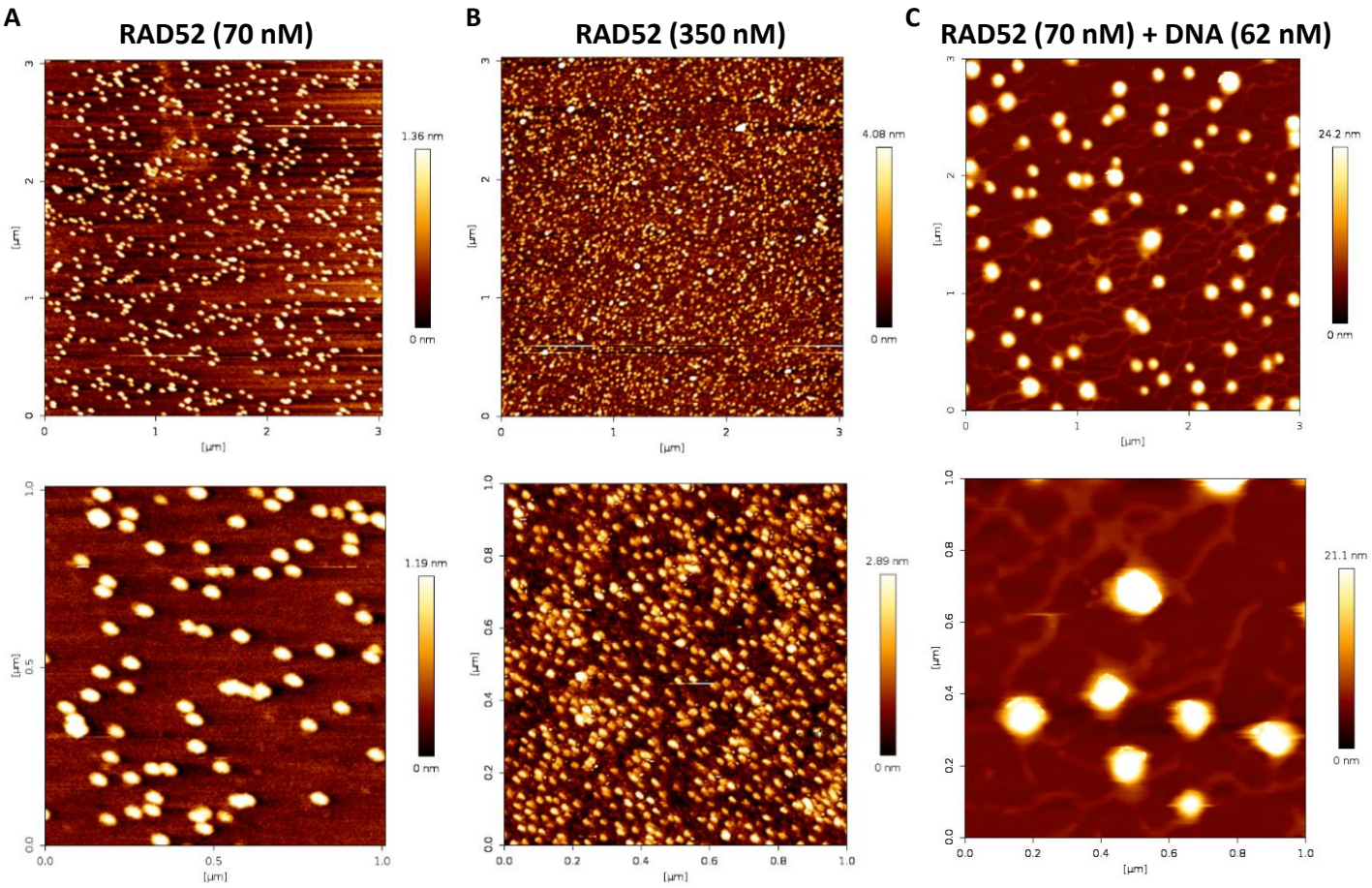


Fig. 41) Afm analysis of (A) RAD52 FL 70nM, (B) RAD52 FL 350 nM, (C) RAD52 FL 70nM and DNA 62nM.

The micrographs of RAD52 FL at 70 nM showed rounded species with a height of 1.55 ± 0.53 nm and a diameter of 43.77 ± 12.38 nm, 30.85 ± 6.70 nm (x and y dimension, respectively). When RAD52 FL concentration was increased, reaching 350 nM, the discrete protein units maintained the same dot shape. However, species height increased in comparison with lower protein concentration conditions (Table 1). Interestingly, significant changes in diameter sizes were not observed in comparison with lower protein concentration condition (Table 1). When RAD52 FL 70 nM was incubated with DNA at 62 nM concentration, no change in particles morphology was observed; however, an increase in both height and diameter was observed in comparison with the protein alone (Table 1).

These data show RAD52 propensity to form high MW superstructures in a concentration-dependent fashion and this propensity is even more evident when in the presence of DNA. Moreover, the key role played by the C-terminal portion of the protein in favoring high MW superstructures formation is further corroborated. This observed RAD52 behavior can be ascribed to its physiological role, which may require RAD52 oligomers interaction and stacking, in order to facilitate DNA strand annealing and homology search.

4.1.6. Electron Microscopy (EM)

Electron Microscopy is a transmission microscopy technique, which allows a high-resolution image construction through an electron beam interaction with the sample. EM is often associated with negative staining sample preparation in order to increase image contrast. This technique was selected both to have a clear overview of RAD52 FL protein oligomerization and to assess samples quality for further Cryo-EM experiments. In Fig. 42, negative staining representative

micrographs are reported. They were acquired at a pixel size of 6.9 Å (A) and 3.5 Å (B) with a total electron dose of ~ 20 $e^-/Å^2$. Particles observed putatively in front and partially side views are marked in white and black, respectively. From these protein images, the oligomerization state of the protein and the ring shape were confirmed. EM experiments were performed by dr. R. Marotta (IIT Genova).

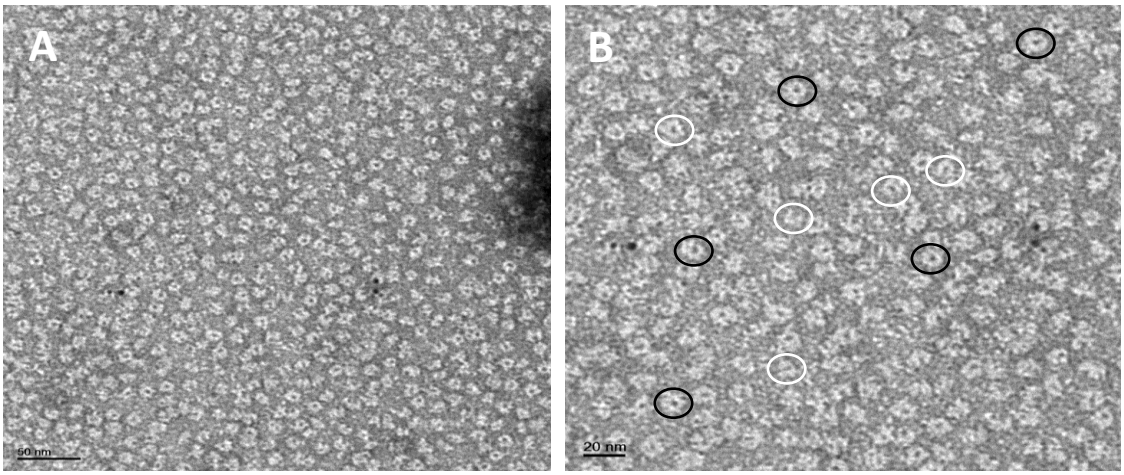


Fig. 42) Negative staining representative micrographs acquired at a pixel size of 6.9 Å (A) and 3.5 Å (B) with a total electron dose of ~ 20 $e^-/Å^2$. Particles observed in front and partially side views are marked in white and black, respectively.

4.1.7. Cryo-EM

By single particle analysis we obtained the first electron density map of the full length human RAD52 oligomer. We initially confirmed that the protein present in the samples used for cryo-EM sample preparation was in the full-length form even after prolonged storage through SDS-page gel and native gel (data not shown). The cryo electron density map was determined at 3.4 Å imposing C11 symmetry (Fig. 43). Indeed, top views class averages with a clear C11 symmetry were present in the 2D classification (Fig. 44). Our results pointed out for the first time that

RAD52 FL organizes in a undecameric ring complex, and not in a heptameric ring as previously suggested[61,64]. Intriguingly, RAD52 FL was reported to form the same complex as its N-terminal truncated form ($RAD52_{21-209}$, PDB 1KN0), consisting of a mushroom-like closed ring complex[61]. Moreover, the model of the human $RAD52_{21-209}$ crystal structure fits very well (0.785 cross correlation coefficient) inside the obtained cryo electron density map (Fig. 43). The RAD52 FL ring complex is formed by eleven monomers rotated each other by approximatively 33° (Fig. 43). The closed ring is composed by a stem region, formed by the β-β-β-α fold of each monomer (residues 79-156), and by a domed cap region that ends with a flat top (Fig. 43). The highly conserved stem region is largely rigid and well resolved as corroborated by ResMap results (Fig. 45). The domed cap instead contains several flexible regions, including part of the hairpin loop (part of β-sheet β1, the loop L3 and part of β-sheet β2), a large portion of the L10 loop and the α4 helix and the region at the top of the domed cup, corresponding to the N- and C-terminal portions of RAD52 N-terminal part (Fig. 45). Notably, the large RAD52 FL C-terminal region, roughly corresponding to residues 209-418, is highly unstructured and flexible, as also corroborated by predictor of natural disordered regions (PONDR) analysis (Fig. 63A). It is visible in some 2D class averages as a large and undefined electron density cloud close to the top of the ring, where the RAD52 C-terminal is supposed to be (Fig. 46). The RAD52 C-terminal is also detectable, at lower density threshold, in the more resolved 3D class averages as a large unstructured region close to the top of the ring (Fig. 47). Sample preparation, data collection and data elaboration were performed by dr. R. Marotta (IIT Genova).

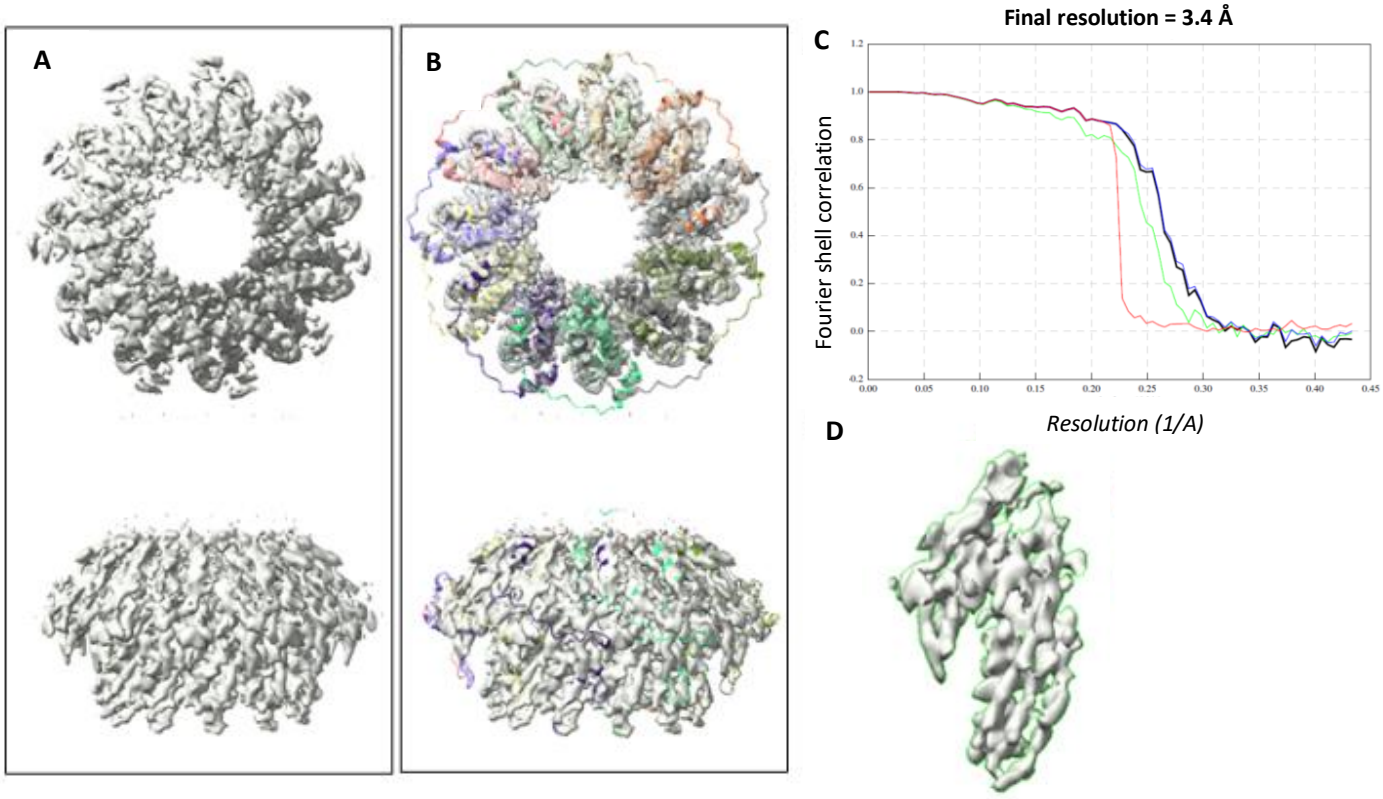


Fig. 43) A) human full length RAD52 cryo electron density map resolved at 3.4 Å in top (up) and side (bottom) view; B) the same map shown in A fitted with the crystal model of the human RAD52$_{21-209}$ N terminal truncated protein (PDB ID 1KN0); C) Fourier shell correlation (FSC) curves (red, FSC phase randomized masked curve; black, FSC corrected curve; blue, FSC masked map; green, FSC unmasked map); D) detail of the map showing the 3 β sheets forming the characteristic human RAD52 β-β-β-α fold.

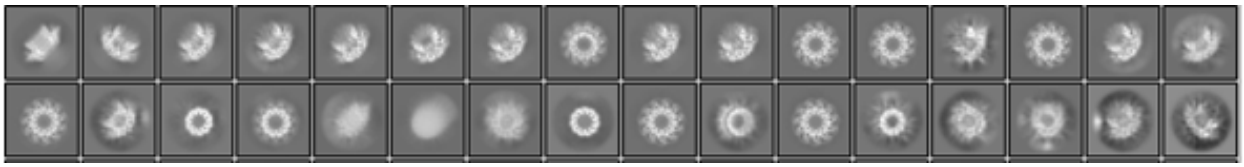


Fig. 44) Human RAD52 full length unsupervised 2D classification of 318638 particles.

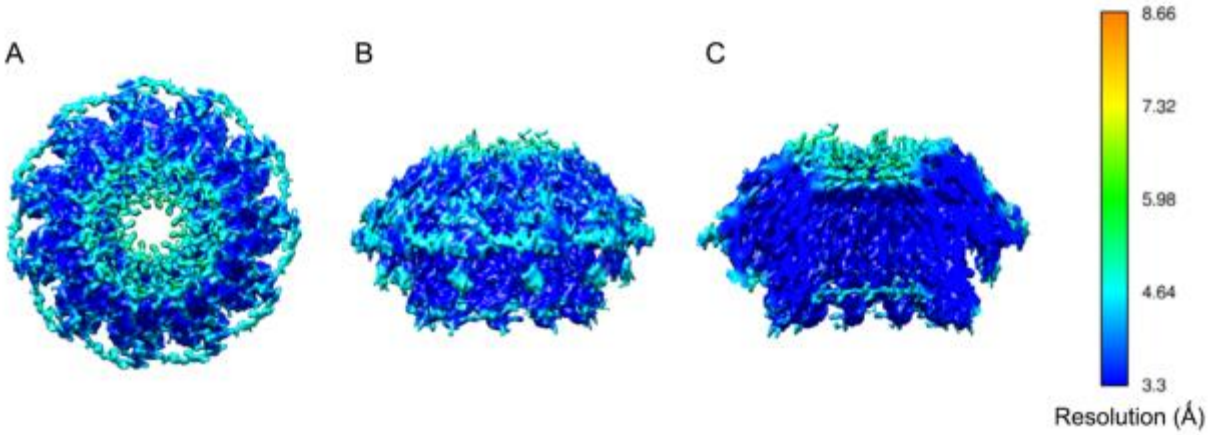


Fig. 45) CryoEM electron density map of the full length RAD52 complex filtered according to ResMap local resolution in top (A) and side (B) views; C, the map shown in B cut longitudinally.

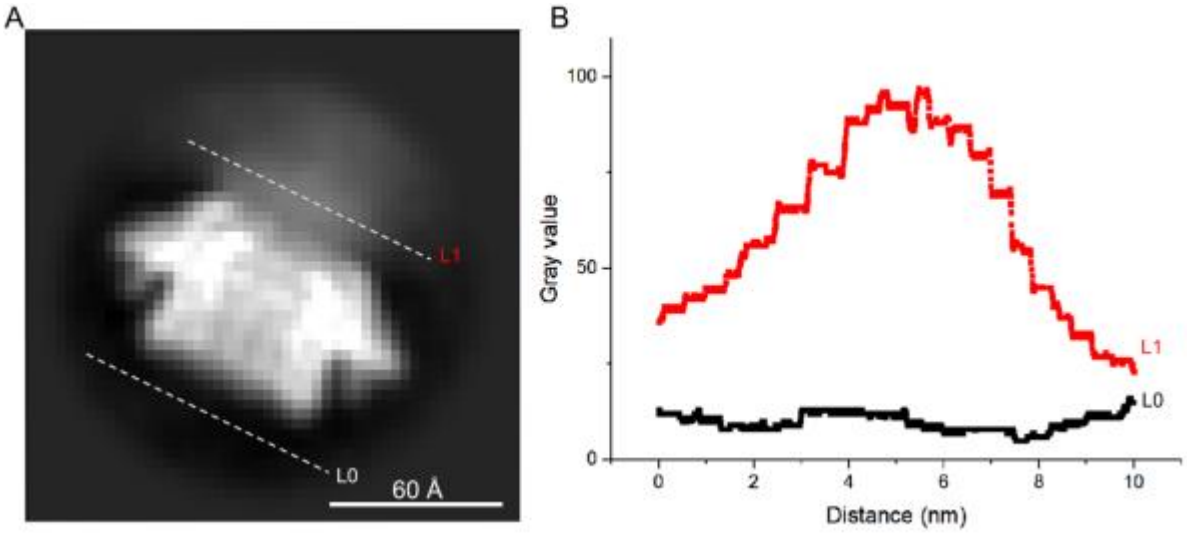


Fig. 46) A) 2D class average showing the human full length RAD52 in side view. An undefined electron density cloud close to the top of the ring is present; B) intensity profile plot trough the lines L0 and L1 in A.

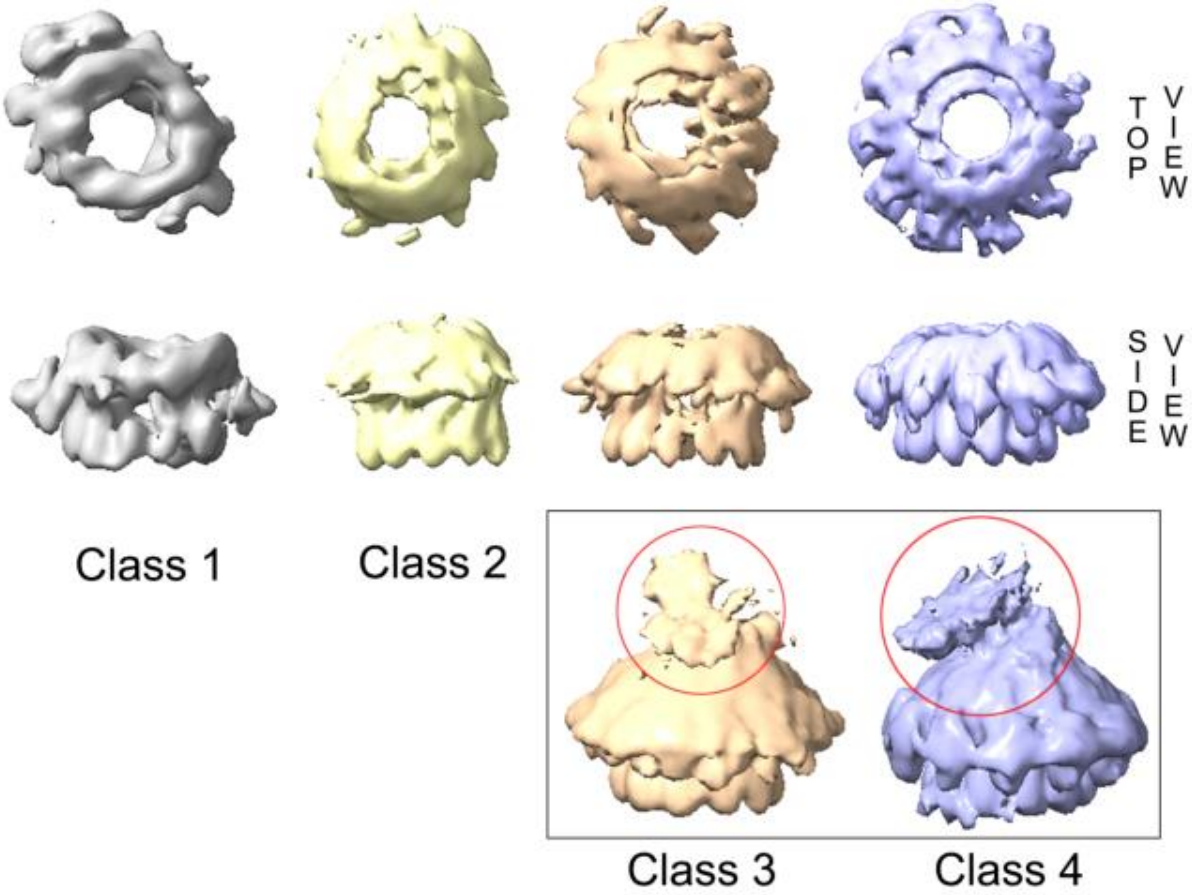


Fig. 47) Human RAD52 full length unsupervised 3D classification showing four 3D classes in top and side views based on 258839 particles. The boxed electron density maps are shown at a lower density threshold. The low-resolution flexible regions putatively corresponding to part of the C-terminal are encircled in red.

4.2. Drug Discovery

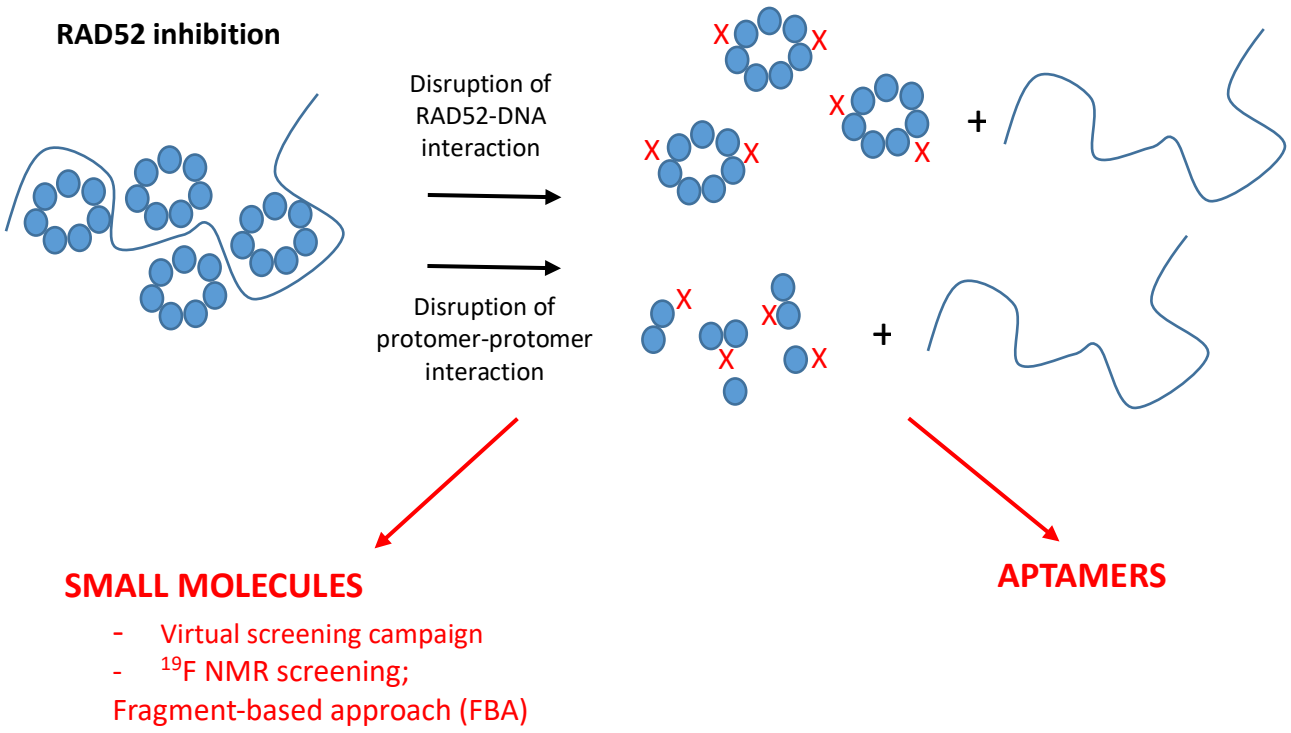


Fig. 48) Image representing the three approaches used to identify novel RAD52 inhibitors and their putative mechanism of actions.

The second part of my PhD project aimed at the identification of novel RAD52 inhibitors. Using three different approaches, the goal was the identification of novel hit compounds able to affect RAD52 activity, either by disrupting RAD52-DNA interaction or protomer-protomer interaction within RAD52 functional ring-shaped unit (Fig. 48).

4.2.1. Identification and biophysical characterization of novel RAD52 inhibitors starting from a virtual screening campaign

The computational unit of our group, through a virtual screening campaign on the crystal structure of RAD52 N-terminal domain (PDB 1KN0), pursued the identification of potential inhibitors of RAD52. The RAD52 pocket targeted in the virtual screening campaign is a pocket essential both for protomer-protomer interaction and DNA-RAD52 interaction, as described in the *materials and methods* section. Notably, the computational study was performed on the N-terminal domain crystal structure and not on the structure of RAD52 FL protein, since no RAD52 FL 3D structure is available so far. However, DNA binding sites are all located in the N-terminal protein domain.

Fifteen compounds were selected, though the computational approach, and subsequently analyzed for their ability to bind RAD52 using NMR spectroscopy (either WaterLOGSY or 1D ^{19}F NMR spectroscopy) and MST. Compounds ability to disrupt RAD52 superstructures or to inhibit DNA binding were tested using Native Gels and FRET, respectively.

4.2.1.1. <u>NMR spectroscopy and MST binding assay</u>

NMR experiments were performed in collaboration with dr. M. Veronesi (IIT Genova). The fifteen compounds were teste by SPAM filter experiments in order to determine their solubility and aggregation state in assay buffer: four compounds showed a solubility lower than 5 µM and were excluded from further analyses (ARN1084, ARN5103, ARN3603, ARN3403). Among the remaining eleven compounds, seven were fluorinated and were tested by ^{19}F T_2 1D NMR experiment to determine their binding to RAD52 (ARN2378, ARN2360, ARN2391, ARN2393, ARN2292,

ARN11679, ARN3758): four out of the seven compounds resulted to bind to the protein, showing a line broadening of theirs ^{19}F signal in presence of RAD52 (see Table 2 and Appendix 7.4, 7.5).

The remaining four compounds (ARN0540, ARN0850, ARN1065, ARN7192), which did not have a fluorine in their scaffold, were tested for their binding to RAD52 using WaterLOGSY and ^{1}H T_2 filter experiments. Briefly, WaterLOGSY experiments exploit the transfer of bulk water magnetization to the chemical compound interacting with the protein (see Appendix 7.5). In WaterLOGSY experiments the NMR signals of small molecules in solution are negative but, when an interaction with the protein is present, their NMR signals become less negative or positive. The four compounds were tested at 100 µM in the absence and presence of 2 µM RAD52. All these four non-fluorinated molecules showed binding to RAD52. In Fig. 49 an example of WaterLOGSY experiments performed on ARN0540 compound is reported. The ^{1}H 1D NMR spectra of the ARN0540 compound is reported in black and the two WaterLOGSY experiments in the absence and presence of RAD52 FL are reported in blue and red, respectively: here the molecule binds the protein and its NMR WaterLOGSY signals become less negative in the presence of RAD52 FL (red spectrum).

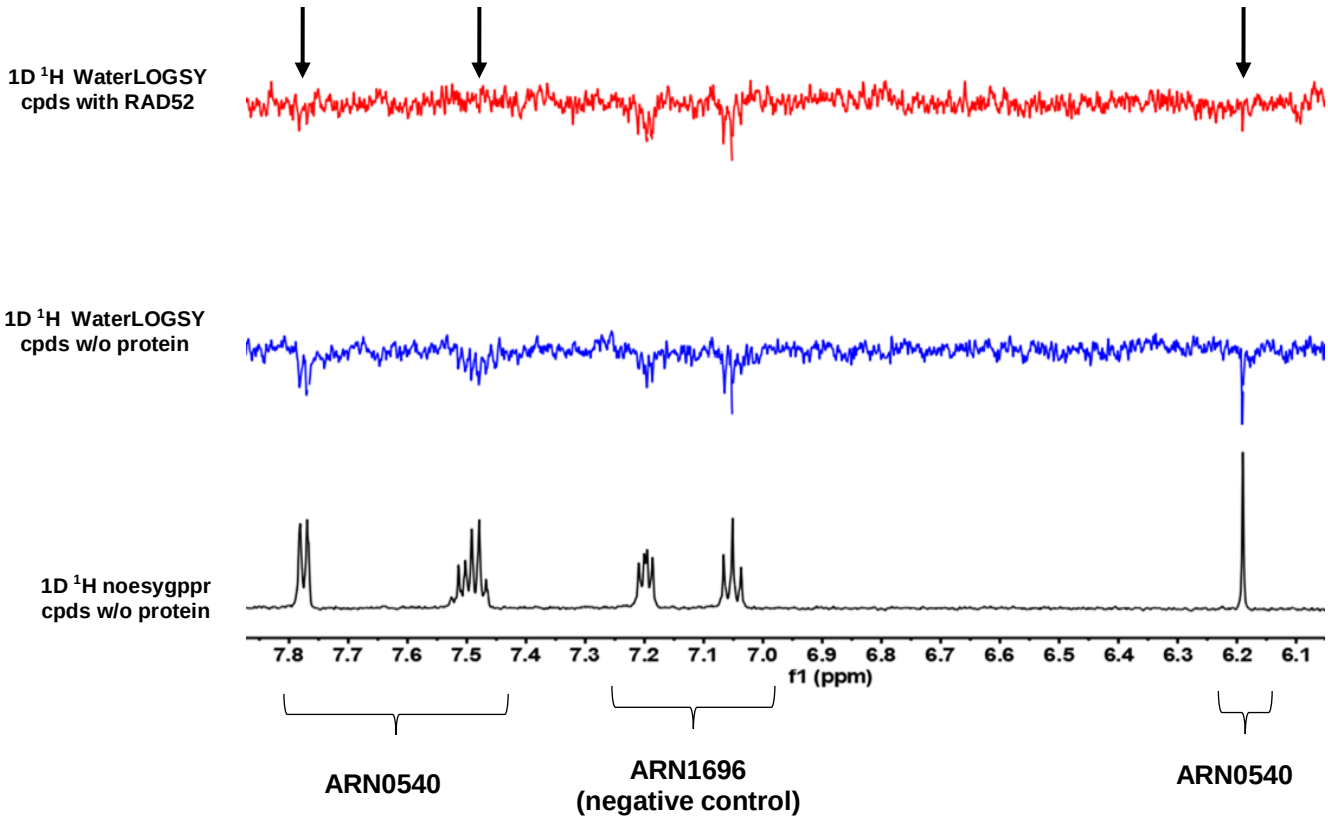


Fig. 49) 1H 1D NMR spectra (black) of ARN0540. WaterLOGSY experiments in the absence (blue) and presence (red) of RAD52 show the binding of the compound to RAD52. Compounds concentration 100 µM, His-hRAD51 concentration 1 µM.

All the eleven soluble compounds were tested for their binding ability to RAD52 FL using also MST, as described in the *materials and methods* section.

In Table 2 data acquired by NMR and MST for the hit compounds obtained from the Virtual Screening campaign are reported.

From these first binding studies compounds ARN1065, ARN2378, ARN2391, ARN2393 and ARN11679 turned out to be promising RAD52 binders, using both NMR spectroscopy and MST. These compounds were therefore selected for further investigations.

Table 2

Compound	Molecular Weight	solubility µM	Aggregation	[cpd] for binding test	Binding NMR	Binding MST
ARN0540	187.20	> 200	no	100	**Yes**	**No**
ARN1084	211.22	80	80 µM		**Not tested**	**Not tested**
ARN0850	224.26	> 40	no	100	**Yes**	**No**
ARN1065	162.19	> 200	no	100	**Yes**	**Yes**
ARN7192	415.53	> 200	> 100 µM	100	**Yes**	**No**
ARN5103	401.48	~ 10	yes		**Not tested**	**Not tested**
ARN2378	232.26	> 200	no	40	**Yes**	**Yes**
ARN2360	237.27	> 200	no	40	**No**	**No**
ARN2391	262.24	<40	no	40	**Yes**	**Yes**
ARN2393	268.33	> 200	no	40	**Yes**	**Yes**
ARN2292	260.31	> 200	no	40	**No**	**Yes**
ARN11679	355.43	> 200	no	40	**Yes**	**Yes**
ARN3758	394.46	> 200	no	40	**No**	**No**
ARN3603	355.40	~ 10	no		**Not tested**	**Not tested**
ARN3403	366.43	< 5			**Not tested**	**Not tested**

4.2.1.2. Native Gel Electrophoresis for RAD52 oligomerization and superstructures states

Native gel electrophoresis were performed as described in the *materials and methods* section, to investigate the effect of the five selected small molecules on RAD52 oligomeric structure and high MW superstructures. Native Gel electrophoresis was performed in the presence or absence of the selected compounds at the maximum concentration possible according to their aqueous buffer solubility. None of the compounds have a disruptive effect on RAD52 superstructures (data not shown).

4.2.1.3. FRET analysis for DNA-RAD52 interaction

Finally, compounds were tested for their inhibitory effect on RAD52-DNA interaction. FRET experiments were set up as reported in the *materials and methods* section. Briefly, whether the compounds had an inhibitory effect on RAD52 binding to ssDNA, the ssDNA would not be able to tightly wrap around RAD52 ring anymore. In this situation, a decrease in FRET effect would be reported due to an increased distance between the donor cy3 and acceptor cy5 fluorophores.

Four of the five compounds selected by NMR and MST binding tests, compounds ARN1065, ARN2378, ARN2393 and ARN11679, showed a decrease in the E_{FRET} value, as expected for compounds able to inhibit the wrapping of the DNA on RAD52 rings, i.e. able to inhibit RAD52-ssDNA interaction (Fig. 50). Nevertheless, these promising data need to be further validated by other techniques (i.e. EMSA) before moving to in cells experiments.

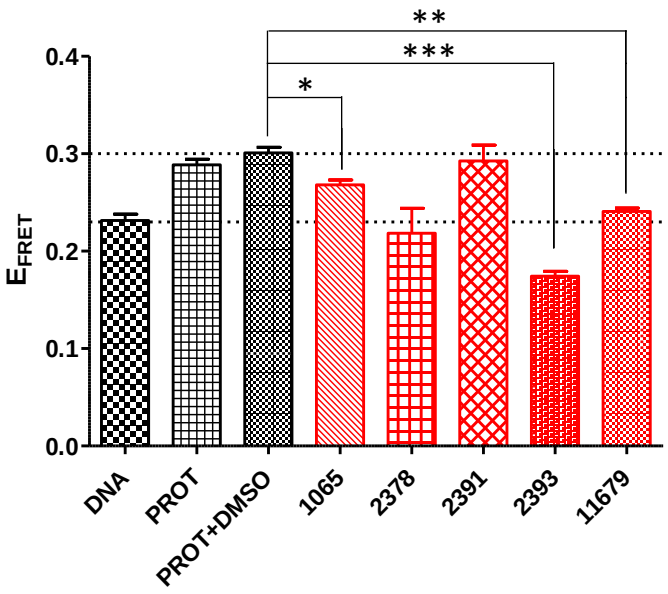


*Fig. 50) FRET assays for identification of compounds effects on RAD52-ssDNA complex disruption; FRET values were measured in samples containing RAD52, dual labelled ssDNA and the selected small molecules (500 µM, with the exception of ARN2391 100 µM); *p-value<0.05; **p-value < 0.005;*** p-value<0.001.*

According to these results, compounds ARN1065, ARN2378, ARN2393 and ARN11679 emerged to be promising starting point for RAD52 inhibitor development. They showed binding evidence both in NMR and MST experiments while FRET experiments suggested their inhibitory role on RAD52-ssDNA interaction. Further characterizations and in cells experiments are required in order to validate the selected promising molecules as potential inhibitors of RAD52 before moving to subsequent developments of their scaffold.

4.2.2. Fragment Based Approach (FBA) screening through 1D ^{19}F NMR spectroscopy for the identification of novel RAD52 inhibitors and characterization of the best hit using biophysical and cellular tools

4.2.2.1. Ligand Based NMR

NMR experiments were performed in collaboration with dr. M. Veronesi (IIT Genova). At the Italian Institute of Technology (IIT) a local Environment of Fluorine (LEF) library constituted by about 900 fluorinated compounds is present. Each compound has a molecular weight ranging from 100 to 350 Da and solubility in buffer higher than 400 µM. All the compounds in the library have passed the SPAM filter[176].

A preliminary direct ^{19} F-NMR ligand-binding screening against RAD52 FL was initially performed, using a portion of 600 fragments of the above mentioned internal LEF Library. Transverse relaxation filter fluorine (^{19}F-R_2) experiments were used in this initial search for possible binders (Appendix 7.4). Specifically, the transverse relaxation rate R_2 is a very sensitive parameter for these studies, due to the large Chemical Shift Anisotropy (CSA) of ^{19}F nucleus and to the large exchange contribution[177]. The observed response (R_{obs}) in the R_2 filter experiments results in a line-broadening of the compound ^{19}F-NMR signal when it binds to the target protein, easy to detect and interpret.

The fragments were screened in mixtures of 20-25 compounds each, at the concentration of 20 µM (CF_3 labeled fragments) and 40 µM (CF labeled fragments), in the presence and absence of 1 µM RAD52 FL. From this screening we identified about 130 fragments showing a line-broadening in the presence of RAD52 and which could be considered as binders. The compounds which were showing a clear binding signal in mixture, were retested and confirmed for their binding to RAD52

as single compounds (at the same concentration) in presence of a non-binder, as negative control and of 2 µM RAD52.

Among all the best candidates, we initially decided to focus on E5 fragment, whose NMR direct binding experiment is reported in Fig. 51. Evidences coming from a parallel project of our group reported E5 as a promising inhibitor of the RAD51-BRCA2 interaction (unpublished data), i.e. the key interaction of the HR DSB repair pathway. If E5 was a promising molecule able to inhibit RAD52 activity, we could aim at the development of a putative "dual inhibitor" capable of simultaneously inhibiting two DNA repair mechanisms at the same time. This idea will be further discussed in the next chapter.

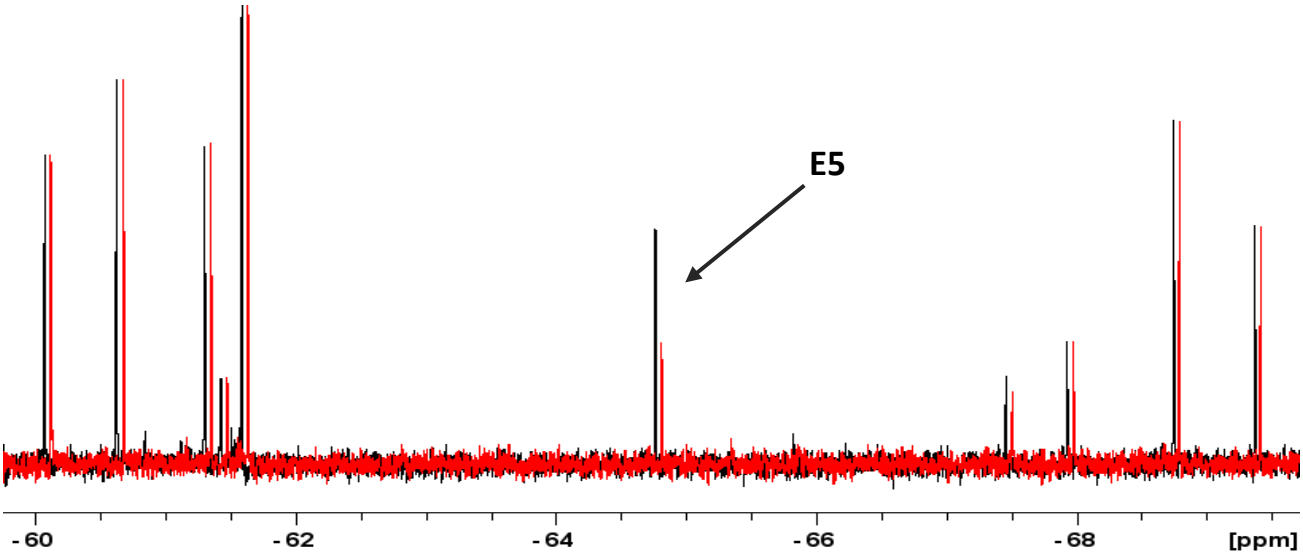


Fig. 51) NMR experiments showing the binding of E5 compound to RAD52. ^{19}F-NMR cpmg spectra of a mixture of 25 fluorinated fragment (20 mM) in absence (black) and in presence of 1 mM (red) RAD52 FL. The arrow indicates the line-broadening of NMR signal of E5 in presence of RAD52, meaning that E5 binds to the protein.

4.2.2.2. MicroScale Thermophoresis of E5 fragment

Microscale Thermophoresis analyses were performed in order to characterize the interaction between the fragment E5 and RAD52 FL (Appendix 7.3). Titration of fluorescently labelled RAD52 FL at a constant concentration with increasing concentrations of E5 compound (up to 1 mM) was used to determine the affinity of the binding. MST analysis of the E5 fragment binding to RAD52 confirmed the interaction of E5 with the protein with an affinity of 218 ± 78 µM (Fig. 52).

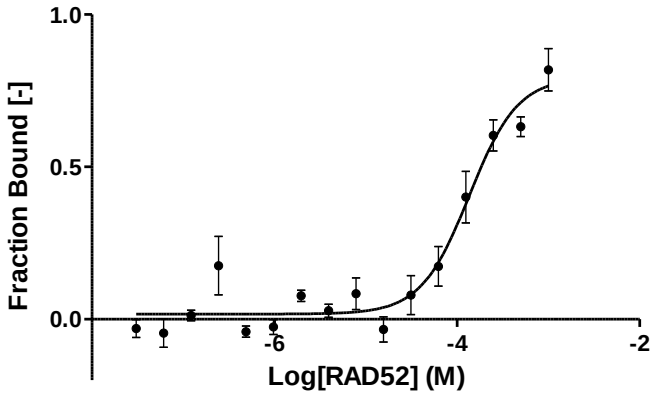


Fig. 52) MST analysis of RAD52-E5 binding. Titration curve of fluorescent-labelled)-RAD52 with increasing concentrations of E5. Sigmoidal fitting curves were obtained using the Affinity Analysis software of Nanotemper Technologies.

4.2.2.3. E5 effect on RAD52-DNA interaction

After validating the binding of E5 to RAD52 protein, E5 effect on RAD52-DNA binding interaction was investigated. EMSA experiments were initially performed as described in *material and methods* section. This technique was selected to easily detect DNA-RAD52 binding inhibition through the observation of variations, in an agarose gel, of fluorescently labelled ssDNA band position in the presence or absence of E5. RAD52 FL - ssDNA samples were incubated with different concentration of E5 (up to 2 mM) and further run on an agarose gel. Results are reported in Fig. 53A. Here, a significant progressive dose-dependent effect of E5 on RAD52 –

ssDNA formation was observed. Notably, E5 effect was comparable with the effect induced by D-I03 used as a reference inhibitor RAD52 (IC_{50} = 25.8 µM) (Fig. 53A lanes 6-8)[151].

E5 inhibition of RAD52-ssDNA interaction was also measured through FRET experiment, as described in *materials and methods* section. As EMSA, FRET technique allowed us to measure E5 inhibition effect on DNA-RAD52 interaction. Whether E5 had an inhibitory effect on RAD52 binding to ssDNA, the ssDNA would not be able to tightly wrap around RAD52 ring anymore. In this condition, there would be a decrease in FRET effect due to an increased distance between the donor cy3 and acceptor cy5 fluorophores. As reported in Fig. 53B, E5 had a significant effect on DNA – RAD52 FL complex formation, showing a decrease in E_{FRET} in comparison with the control sample.

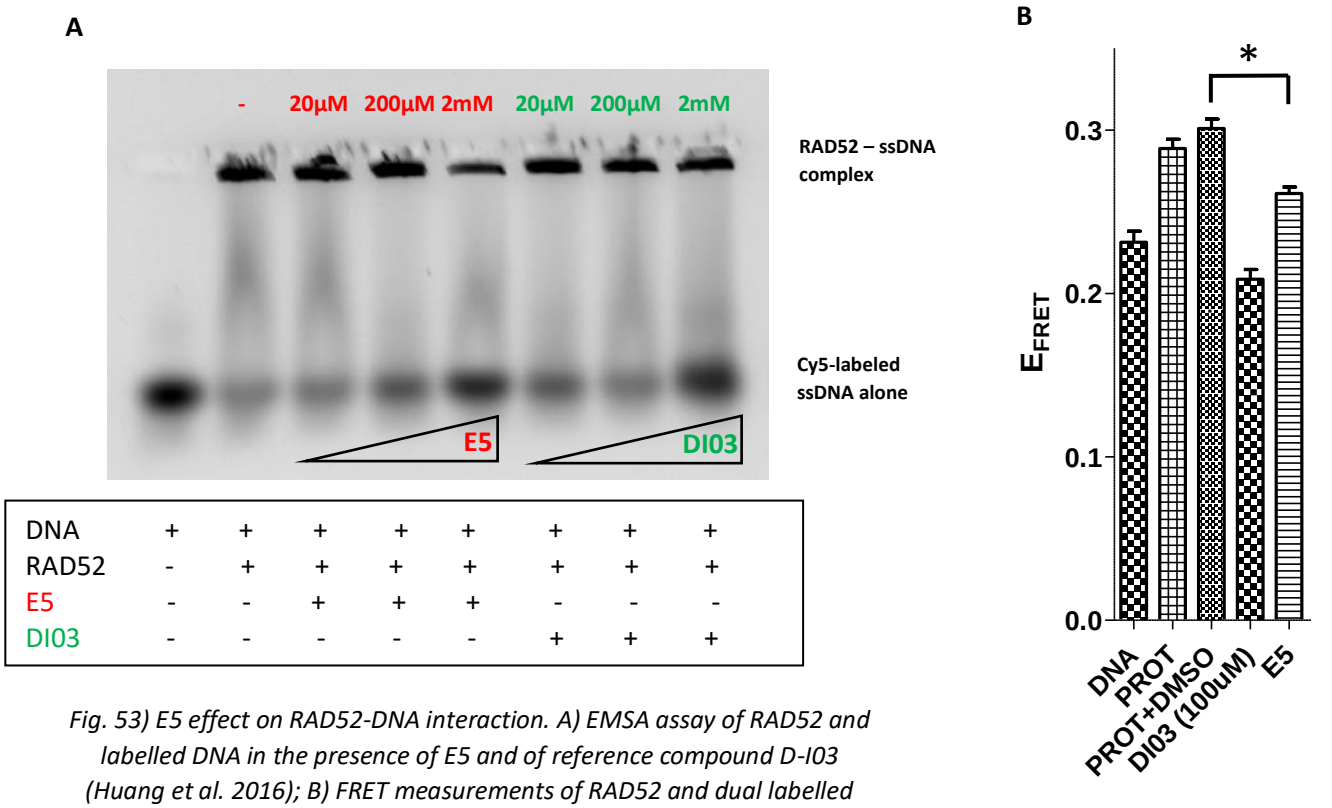


*Fig. 53) E5 effect on RAD52-DNA interaction. A) EMSA assay of RAD52 and labelled DNA in the presence of E5 and of reference compound D-I03 (Huang et al. 2016); B) FRET measurements of RAD52 and dual labelled ssDNA in the presence of E5 compound; * p- value<0.05.*

4.2.2.4. E5 effect on RAD52 oligomerization and high MW superstructures states

The ability of the E5 compound to affect RAD52 heptameric rings units and higher MW superstructures was also investigated through other techniques: Native Gels, SEC and DLS.

All the experiments were performed investigating the superstructures assembly of RAD52 in the presence and absence of the E5 compound (1mM). SEC profile chromatogram, DLS profile and native gel protein migration did not show any significant change upon protein incubation with E5 compound (Fig. 54). No effect induced by E5 on RAD52 superstructures assembly was observed.

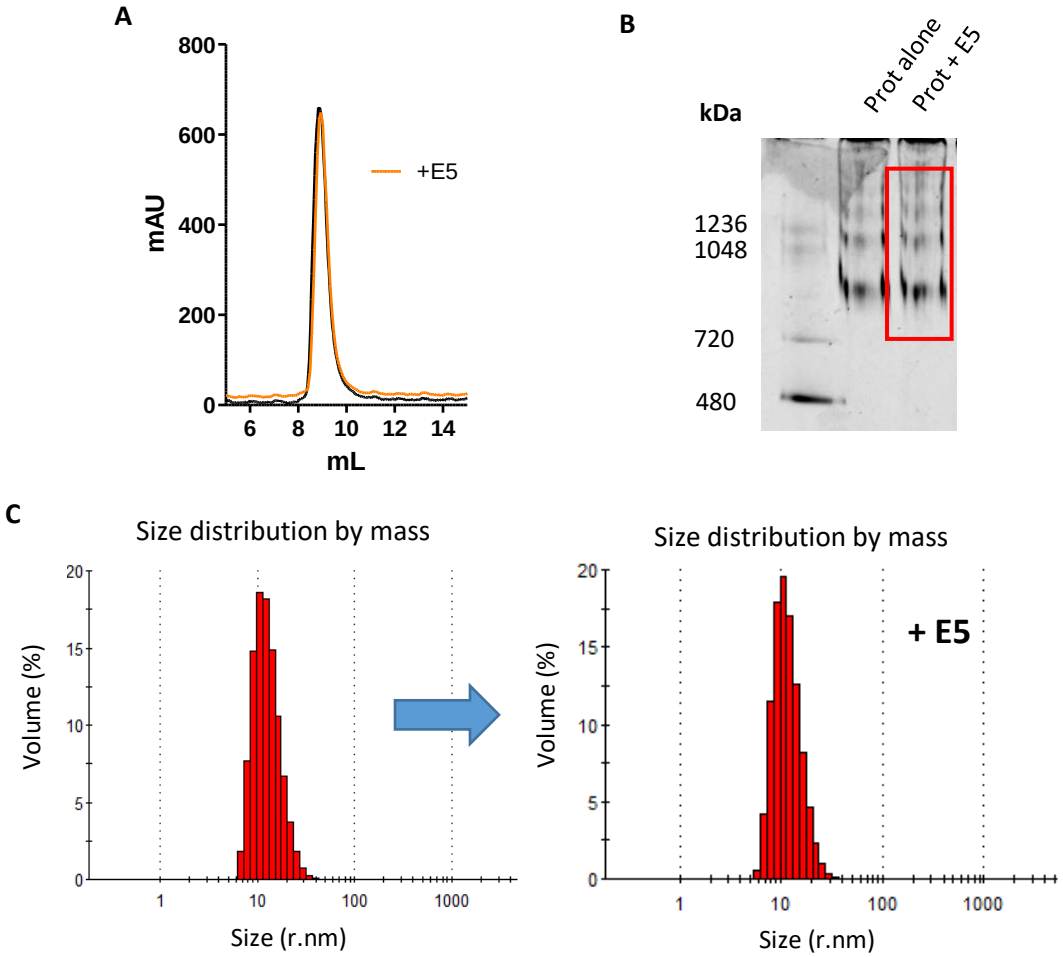


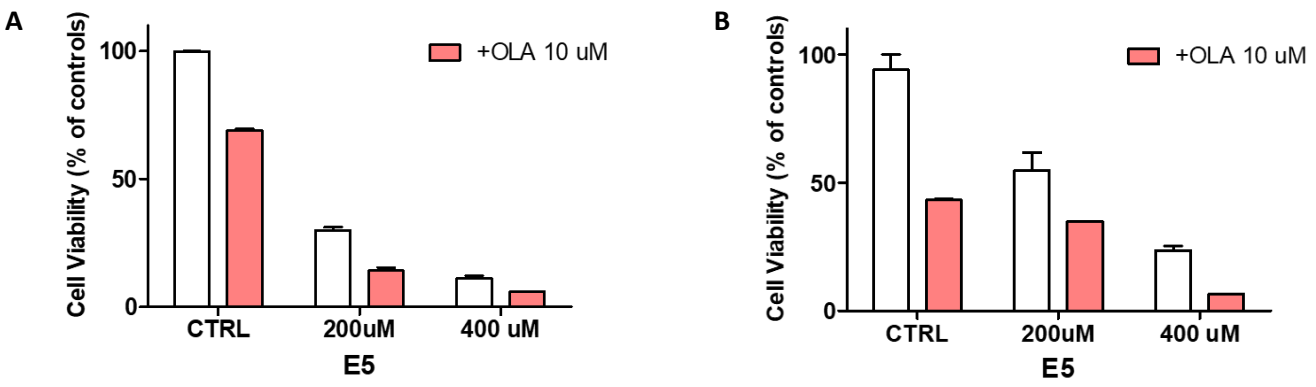
Fig. 54) Tests for E5-induced quaternary structure changes. A) SEC of RAD52 protein alone and after 1.5-hour incubation with 1 mM E5; B) Native gel of RAD52 alone and after 1.5-hour incubation with 1 mM E5; C) DLS analysis of RAD52 alone and after 1.5-hour incubation with 1 mM E5.

4.2.2.5. E5 cellular studies

Viability assay and combination treatment evaluation

For *in cellulo* experiments, we used immortalized pancreatic adenocarcinoma cell lines. We decided to focus on this type of cell lines since pancreatic cancer is one of the most severe cancer with a very poor prognosis and survival, still lacking of any really effective anticancer therapy. Notably, in 2019, FDA approved olaparib as first-line maintenance treatment for BRCA-mutated metastatic pancreatic cancer, exploiting the drug-induced synthetic lethality phenomenon[158,178,179]. For this reason, our idea was the identification of novel inhibitors which could be used in combination with olaparib to trigger synthetic lethality therapies in pancreatic cancers, either BRCA-deficient or proficient cells.

From the previously reported experimental evidences, E5 turned out to be a promising fragment inhibitor of RAD52 protein *in vitro.* E5 was further analyzed for its activity in cells, to study its effect on cell viability and to evaluate potential synergistic effect with olaparib (PARP inhibitor) to improve synthetic lethality therapies. With this aim, E5 effect was tested in pancreatic cancer cell lines either proficient or deficient for BRCA2 activity (i.e. BxPC-3 and Capan-1, respectively). Results of viability assay performed both in BxPC-3 cells and Capan-1 cells (144h) are reported in Fig. 55.



BxPC-3	CI
E5 200µM	0.684247
E5 400µM	0.759243

CAPAN-1	CI
E5 200µM	1.480536
E5 400µM	0.635462

Fig. 55) Viability assay of BxPC-3 and Capan-1 cells treated either with E5 only or co-treated with E5 and olparaib; Both BxPC-3 (A) and Capan-1 (B) show a synergistic effect of E5 with olaparib when used at 200/400 µM and 400 µM, respectively.

According to our data, the combination index of the co-treatment with E5 and olaparib reported a synergistic effect (CI< 0.8[166]) of the two compounds when E5 was used at 200 and 400 µM in BxPC-3 and at 400 µM in Capan-1 cells. These data are in agreement with *in vitro* assays suggesting E5 as a promising starting point to be developed for potential applications in synthetic lethality therapies.

Immunocitochemistry staining of RAD52 and γ-H2AX nuclear foci

To further investigate the effect induced by E5 on DNA damage repair mechanisms and on RAD52 activity, immunofluorescence experiments were performed. In Fig. 56 the results of the immunofluorescence staining of γ-H2AX in BxPC-3 cells are reported. The phosphorylation of H2AX histone is a DSB marker, commonly used for detection of DNA damage sites and for their

quantification[180]. A significant increase of γ-H2AX foci formation was observed when cells were treated with cisplatin (CisPt) (p value < 0.001), as expected from such a sever DNA damage. Interestingly, a decrease in γ-H2AX foci formation was observed in CisPt-E5 treated cells, compared to CisPt only treated cells. E5 alone, however, did not show any significant increase in γ-H2AX foci formation, and thus, in dsDNA damage, in comparison with control cells. As additional control, olaparib-treated cells were analyzed for γ-H2AX foci formation, and, as expected, no increase of γ-H2AX foci formation compared to control cells was observed.

Immunofluorescence experiments showed interesting new data about E5 effects induced on γ-H2AX foci formation. These evidences are only showing a trend, without significative data, though. For this reason, improvements of the assay and optimization of the tested conditions are required to better describe the effects induced by E5 in cells.

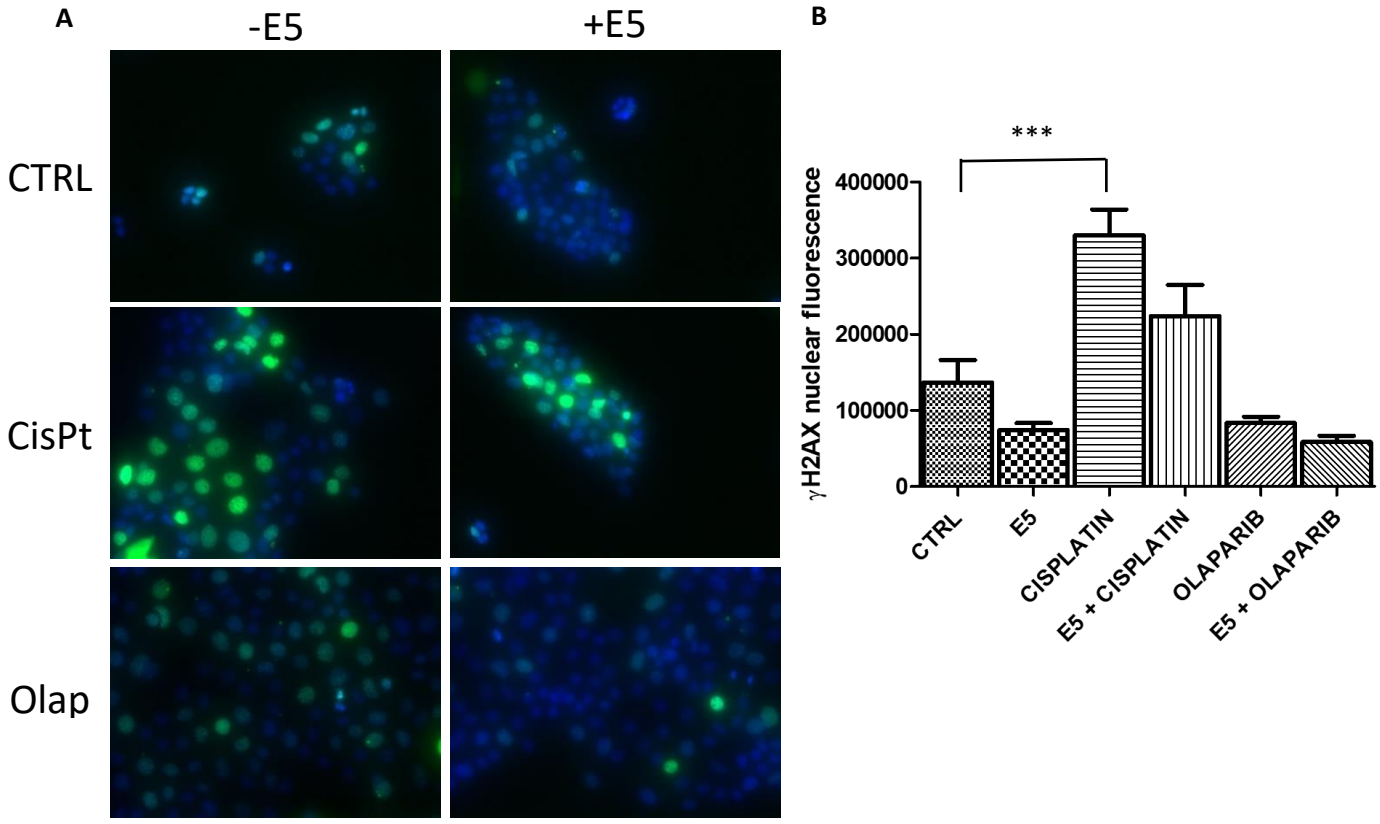


Fig. 56) DNA damage evaluation through Immunofluoresce assays; A) Evaluation of DNA damage through immuno detection of nuclear γ-H2AX foci in BxPC-3 cells exposed to cisplatin, E5 and cisplatin/E5; untreated cells and cells treated with olparaib were used as an additional negative control for DSB; B) Bar graph showing the nuclear fluorescence of γ-H2AX -labelled nuclei. Images were elaborated and analyzed using ImageJ software.

According to these results, E5 emerged as promising starting point for developing a novel RAD52 inhibitor, showing a binding to RAD52 protein corroborated by different techniques and reporting an inhibitory activity on RAD52-DNA interaction. E5 showed moreover interesting cellular activity on viability and interference with DNA damaged sites. Additional studies are required for E5 further characterization.

Additionally, from ^{19}F NMR fragment screening, many hit compounds came out as binders of RAD52 and they will be further characterized for their inhibitory properties on RAD52.

4.2.3. Aptamers in silico design and validation using biophysical and in cellulo tools

Aptamers are a novel class of small nucleic acid ligands, single-stranded RNA or DNA oligonucleotides, that have high specificity and affinity for their targets[181]. They are promising new tools for cancer therapies since they present several advantages: faster and more efficient tissue penetration compared to antibodies; low immunogenicity; high thermal stability and low structural variation; rapid and automated large-scale production process; low production costs. We therefore decided to focus also on this approach as an innovative alternative to canonical small-molecules inhibitors.

Using catRAPID algorithm, developed by prof. G. Tartaglia (IIT Genova) and dr. A. Armaos (IIT Genova) and further described in the *materials and methods* section, aptamers sequences were

designed in order to interact with the DNA binding site of RAD52 and thereby to disrupt the DNA binding activity of RAD52. Eleven aptamers were selected and their sequences are reported in table 3 with their relative catRAPID score. Computational analyses were performed by dr. A. Armaos (IIT Genova).

The secondary structures of the eleven aptamers were investigated to determine how aptamer folding could be affecting aptamer binding to RAD52 and to evaluate if strong binders aptamers had a common structural profile compared to weak/non-binders aptamers. Aptamers ability to bind and inhibit RAD52 activity and cellular internalization properties were also investigated.

Table 3

Aptamer	SEQUENCE	Response k_d (±SEM)	CatRAPID
APT1	GCGGGGCGGGGCGGG	22.3±4.7 nM	6.46
APT2	GCGCGGCGCGGCGCG	NO BINDING	6.17
APT3	GCCGGGCCGGGCCGG	NO BINDING	5.69
APT4	GGGGGGGGGGGG	90.3±15.0 nM	5.73
APT5	GGGCGGGCGGGC	773.3±50.5 nM	5.26
APT6	GCGGGCGGGCGG	126.0±47.1 nM	5.05
APT7	GGCGGGCGGGCG	2.0±0.3 μM	5.07
APT8	GACGGGACGGGACGG	NO BINDING	4.89
APT9	GAGGGGAGGGGAGGG	130.0±11.6 nM	4.03
APT10	GAGCGGAGCGGAGCC	NO BINDING	4.30
APT11	GAGGCGAGGCGAGGC	NO BINDING	4.23

4.2.3.1. Evaluation of aptamers affinity for RAD52 using Bio-Layer Interferometry (BLI)

In order to assess the binding affinity of the selected aptamer sequences for RAD52, BLI experiments were performed as described in *materials and methods* section. BLI technique allows the identification of biomolecular interactions exploiting the change of interference between waves of light depending on mass immobilization on specific biosensors (Appendix 7.2). Using BLI, we assessed the binding affinity of the selected aptamers for RAD52 FL protein. An example (Apt1) of the performed experiments is reported in Fig. 57.

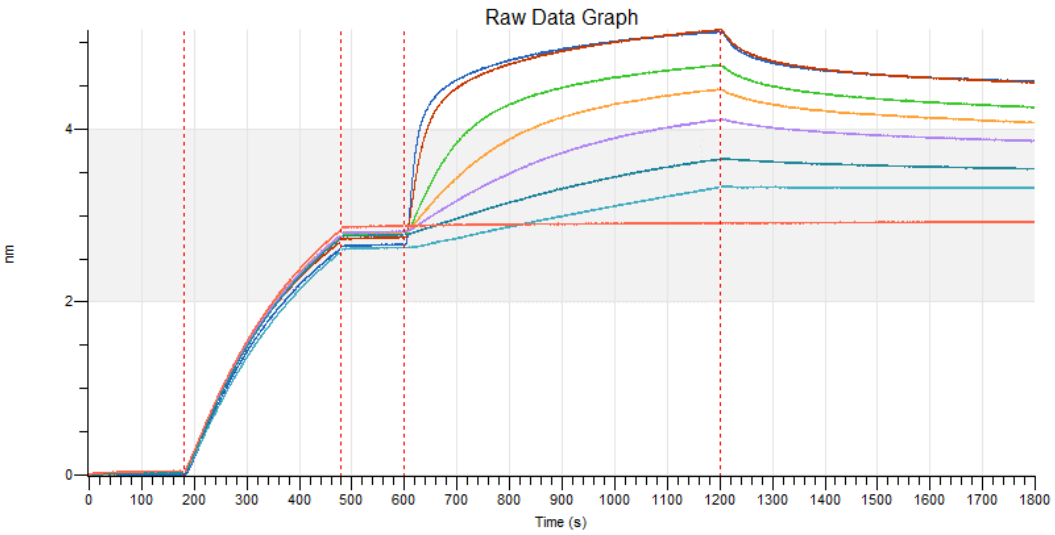
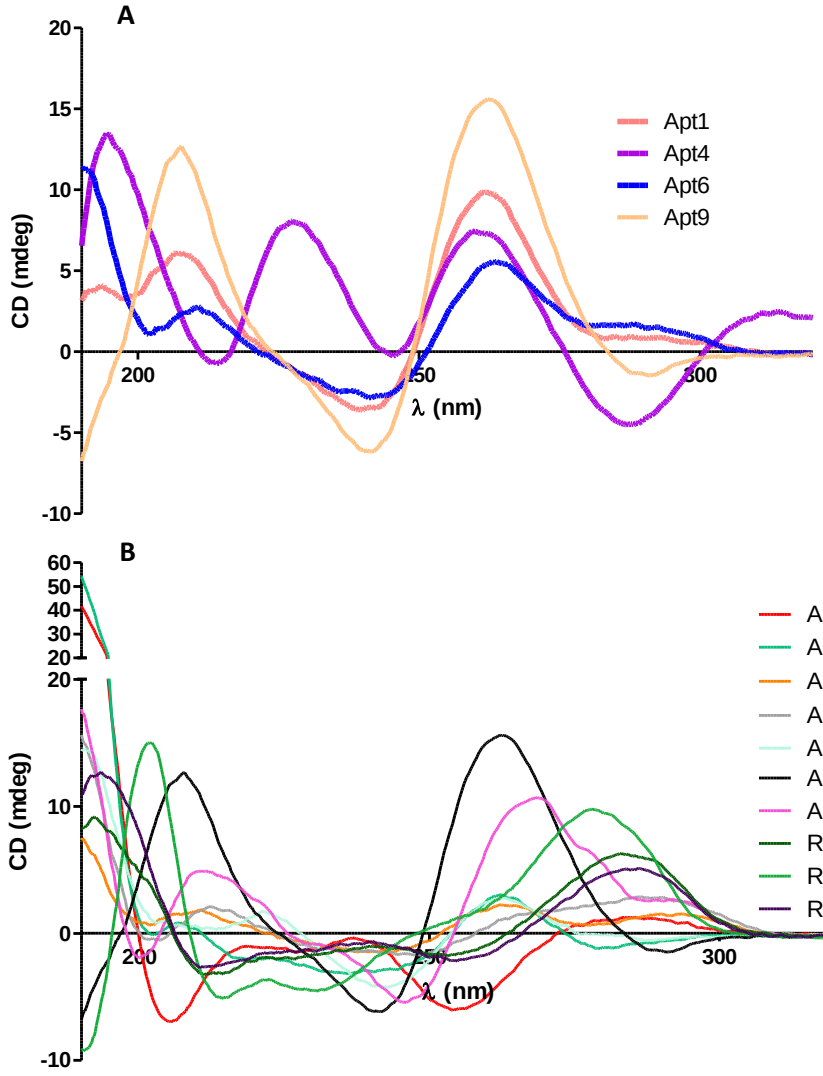


Fig. 57) Example BLI sensorgram reporting Apt1 binding detection; Apt1 was immobilized on the biosensor and was testes with different RAD52 concentrations. The assay phases are divided by vertical dot lines: baseline, loading, washing, association, dissociation.

As reported in table 3, Apt1, Apt4, Apt6 and Apt9 bound RAD52 with a high affinity (K_d < 200 nM); on the contrary, Apt2, Apt3, Apt5, Apt7, Apt8, Apt10, Apt11 showed a weak binding or no binding to RAD52.

4.2.3.2. Aptamers secondary structures evaluation

Aptamers exert their function through their oligonucleotide sequence-related secondary structure[182,183]. For this reason, designed aptamers were characterized using CD spectroscopy, to identify potential recognizable DNA folding conformations that could be possibly associated with experimental binding properties and inhibitory activity on RAD52 (Appendix 7.1).



	Pqsfinder G-quad prediction
Apt1	45
Apt9	45
Apt4	35
Apt6	13

	Pqsfinder G-quad prediction
Apt2	0
Apt3	10
Apt5	0
Apt7	13
Apt8	10
Apt10	0
Apt11	0

Fig. 58) Aptamers secondary structure analysis. Reported plots show strong binders (A) and weak binders (B) CD spectra. In strong binders spectra, a common pattern of curves maxima and minima could be identified; on the contrary, this pattern is not recognizable in weak binders spectra. CD experimental evidences are corroborated by Pqsfinder G-quadruplex predictor, which reports high G-quadruplex prediction scores for 3 out of 4 best binders (A), whereas report low scores for all the weak binders (B).

Figures 58A and 58B report the CD profiles of the aptamers that showed the highest affinity for RAD52 (Apt1, Apt4, Apt6, Apt9) and the ones that showed a lower affinity for RAD52 (Apt2, Apt3, Apt5, Apt7, Apt8, Apt10, Apt11), respectively. Aptamers with the highest affinity for RAD52 showed a high peak at 260 nm; moreover, Apt6 and Apt9 showed a local maximum at 295 nm and Apt1, Apt6, Apt9 showed an additional local minimum at 245 nm. In order to interpret these CD structural data, a comparison with literature evidences of DNA secondary structures was performed. Specifically, among different classes of DNA structures, CD data of DNA highly structured G-quadruplexes were used as reference[184]. In fact, G-quadruplex structures are four-stranded G-rich nucleic acid sequences involved in key genomic functions such as genome stability and transcription regulation, which are characterized by clear and recognizable secondary structures CD spectra profile[185]. G-quadruplex structures can be divided into parallel, antiparallel and hybrid, depending on the orientation of their loops (Fig. 59).

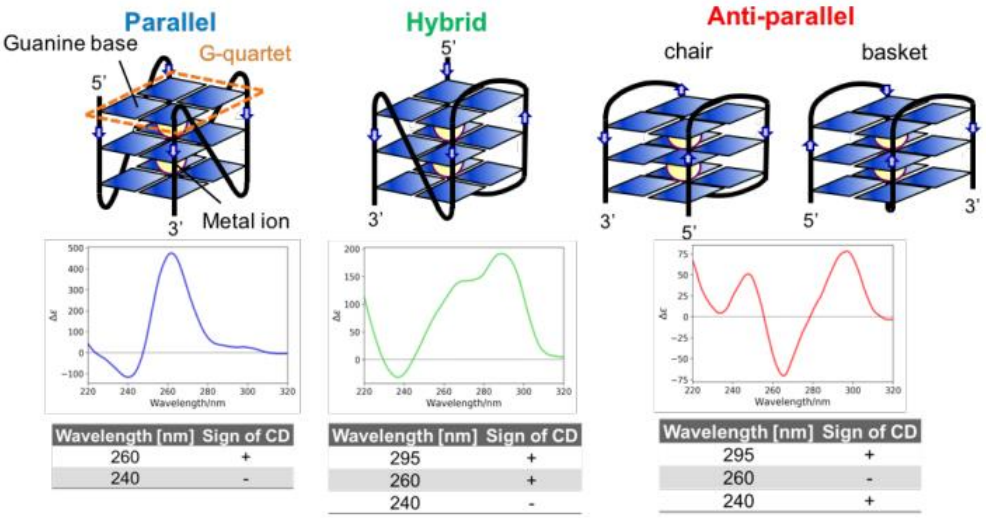


Fig. 59) Types of G quadruplex tertiary structures and their CD spectra; image taken form Jasko application note, Suzuki et al. (2020).

CD profiles suggested that Apt9 resembled a parallel G-quadruplex conformation and Apt1 and Apt6 resembled hybrid G-quadruplex conformations[184]. On the other end, Apt2 from the weak binders group, showed a minimum at ~ 260 nm and a local maximum at ~295 nm, resembling an antiparallel G-quadruplex structure. Surprisingly, Apt10, which did not bind RAD52, showed a CD profile similar to Apt9[184].

These data were corroborated by Potential Quadruplex-forming Sequence finder (Pqsfinder) software[186] (computational studies performed by dr. A. Armaos, IIT Genova). This software is based on an algorithm, which identifies potential stable G-quadruplex structures depending on the G-sequences of a specific DNA region. Hence, pqsfinder is also able to predict the G-quadruplex formation propensity of an aptamer from the oligonucleotide sequence[186–188]. Apt1, Apt9, Apt4 showed the highest Pqsfinder values, whereas all the others aptamers showed a low Pqsfinder value, in agreement with the experimental data. The only exception is Apt6 that showed a low Pqsfinder value even though it had a CD secondary structure profile resembling a hybrid G-quadruplex conformation and it showed a high affinity binding to RAD52.

Overall, the analysis of the CD secondary structure profiles suggests that aptamers that have a highly structured conformation are the best binders of RAD52, in line with computational predictions; among these, Apt1 represent the most promising aptamer for further studies.

4.2.3.3. Apt1 effect on RAD52 oligomerization and high MW superstructures states

Apt1, the aptamer with the highest affinity for RAD52, was further tested for its ability to inhibit RAD52 activity. Apt1 effect on disruption of high MW RAD52 superstructures or disruption of protein oligomers was tested using Native Gels, SEC and DLS.

All the experiments were performed measuring the oligomerization and superstructures state of RAD52 in presence and absence of Apt1 (1µM). SEC profile chromatogram, DLS profile and native gel protein migration did not change upon protein incubation with Apt1 (Fig. 60), suggesting that RAD52 oligomerization and superstructural conformations were not affected by Apt1.

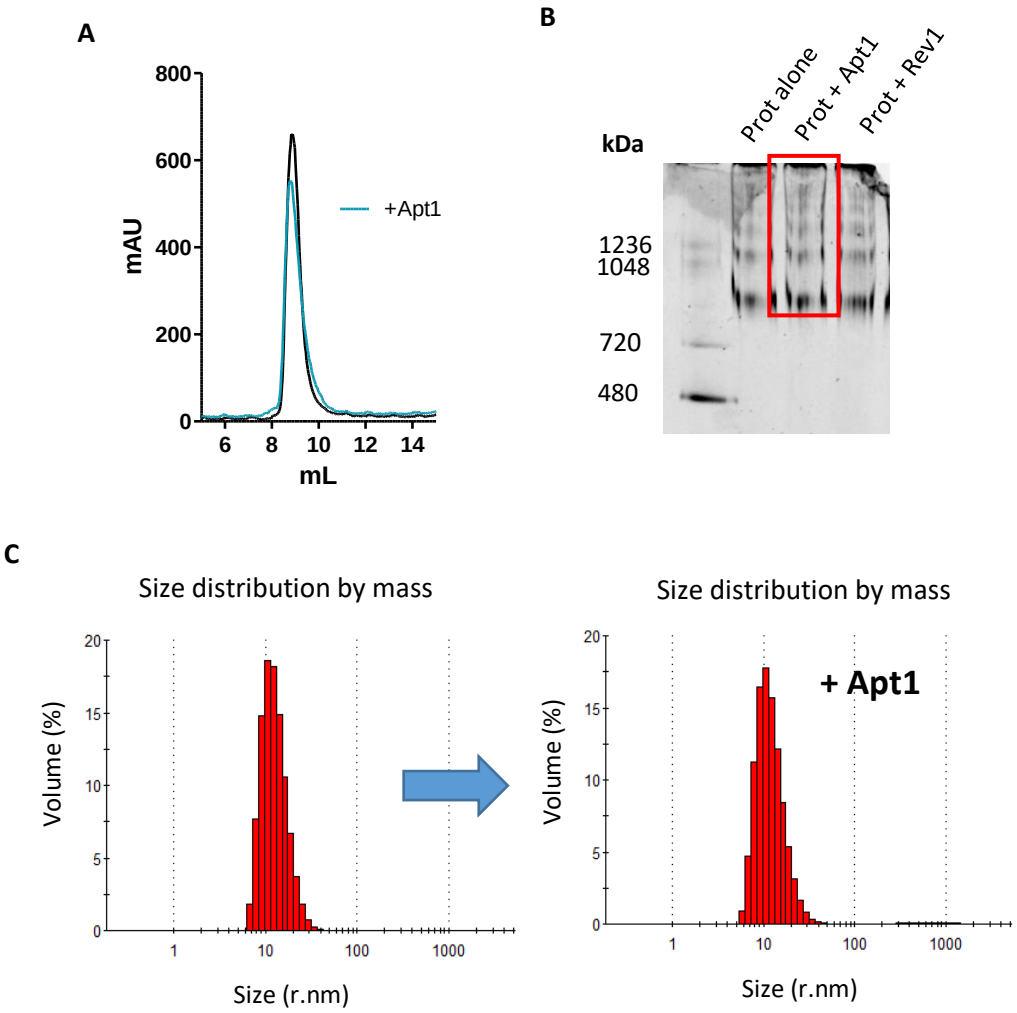


Fig. 60) Tests for Apt1-induced quaternary structure changes. A) SEC of RAD52 protein alone and after 1-hour incubation with 1 µM Apt1; B) Native gel of RAD52 alone and after 1-hour incubation with 1 µM Apt1; C) DLS analysis of RAD52 alone and after 1-hour incubation with 1 µM Apt1.

4.2.3.4. Apt1 internalization in BxPC-3 cell line

Finally, in order to proceed with *in cellulo* tests, the propensity of cancer cells to internalize Apt1, had to be evaluated. Cytofluorimetric experiments were performed as described in *materials and methods* section.

Briefly, BxPC-3 pancreatic cancer cells were treated with TexasRed-labelled Apt1 to quantify, though cytofluorimetry, the number of cells able to internalize Apt1. With preliminary optimization set-ups, an internalization of Apt1 of 61.53% was observed (Fig. 61).

These results showed that Apt1 was well internalized in BxPC-3 pancreatic cancer cells, confirming the possibility of further using Apt1 in activity and viability assays *in cellulo*. These *in cellulo* assays are ongoing in our laboratory.

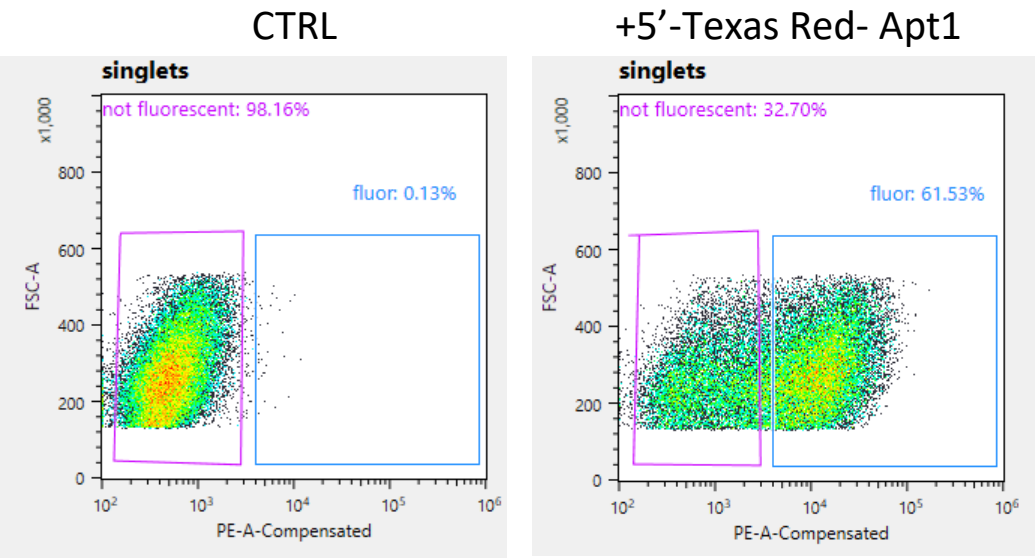


Fig. 61) FACS experiment for cellular internalization of Apt1 in BxPC-3 pancreatic cancer cells. Texas-RED Apt1 was transfected in BxPC-3 cells and the internalized percentage of Apt1 was measured and compared with the control sample, using flow cytometry.

5. Discussion and conclusion

RAD52 is a DNA/RNA binding protein, which mediates several DNA double strand break repair mechanisms. Even though not many information are available on both RAD52 structure and function, in the last few years, RAD52 has been investigated since evidence shows that its inhibition leads to synthetic lethality in the presence of HR deficiencies, in several cancer-related pathologies[54–56]. These evidences make RAD52 a potential pharmacological target for personalized medicine to develop novel therapies, avoid the onset of resistance and enhance the efficacy of already-in-use drugs (i.e. PARPi in BRCA2-deficient cell lines).

My PhD project focused on two main topics: 1) Structural and biophysical characterization of RAD52 to gain novel insights on protein structure and mechanism of action; 2) Application of three different drug discovery approaches to RAD52 protein to identify hit compounds to be eventually developed into novel inhibitors for cancer treatments.

RAD52 Full-Lenght and N-terminal truncated RAD52 [1-212] forms were initially expressed and purified using an optimized purification protocol (i.e. affinity HisTrap column and ionic exchange Heparin HP column for RAD52 FL, HisTrap column for RAD52 [1-212]). The quality of the proteins was assesses by LC-MS. The oligomeric states of both RAD52 FL and RAD52 [1-212] were investigated through size exclusion chromatography. Considering that according to literature data the two proteins form heptameric and undecameric ring respectively[61,63,64,73], the chromatographic elution profiles of the two forms suggested that the two proteins arrange in 2 hepatameric and 1 undecameric functional ring units, respectively. The propensities of the two proteins to form high MW superstructures were then analyzed using DLS, SLS and Native Gel

electrophoresis. The three techniques suggested that RAD52 FL has a high propensity to form high MW superstructures, while a more stable single undecameric structure in solution is observed for RAD52 [1-212]. These results are in agreement with published data, which suggest a higher propensity of RAD52 FL to form superstructures composed of several ring functional units, compared to RAD52 [1-212], which shows a higher stability in aqueous buffer and a lower tendency to form high MW superstructures[63,65,73,189]. These evidences can be explained by the fact that the C-terminal portion of RAD52, completely removed in the RAD52 [1-212] form, can facilitate the formation of high MW quaternary structures. Indeed, this protein propensity to form superstructures can be correlated to the physiological function of RAD52, i.e. assisting DNA homology search and strand annealing, as already reported in the literature[63,73,78,190]. This hypothesis is supported by our AFM experiments performed on RAD52 at different concentrations as well as on RAD52 in the presence of DNA: while concentration increase seems inducing only an increase in superstructures height, i.e. most likely and increase in the number of RAD52 rings stacking, DNA addition induces an increase in RAD52 superstructures size in all three dimensions, suggesting a specific functional role of superstructures formation upon DNA binding.

Given the critical role played by the C-portion of RAD52, not only in the interaction with partner proteins (i.e. RPA, RAD51, etc.), as reported in the literature, but also for the above reported role in the formation of potentially physiological superstructures, we pursued the structural characterization of the full-length RAD52 protein. To date, only the 3D structure of the N-terminal portion of RAD52 has been solved by X-ray crystallography (PDB 1KN0, 1H2I, 5JRB, 5XRZ, 5XS0), while no high-resolution structural information are available on the C-terminal portion. Electron

microscopy investigations were then performed on the full length RAD52 form. Negative staining experiments allowed us to confirm that also RAD52 FL is assembled in oligomeric ring-shaped units[63,67], as well as to optimize the experimental conditions to further proceed with Cryo-EM samples preparation. Once experimental conditions were optimized (concentration, buffer, etc.), cryo-EM grids were prepared for the following data collection and analysis. Cryo-EM experiments allowed to obtain a 3D structure of RAD52 at 3.4 Å resolution. Surprisingly, we discovered that the N-terminal portion of RAD52 FL is arranged in oligomeric ring-shaped units formed by 11 monomers. These data are in contrast with literature hypothesis of RAD52 FL existing as a heptameric ring[61,63,64]. Actually, the assumption that ascribed to the full-length protein a heptameric ring conformation is based on indirect evidences or low resolution data[63,64]. We therefore, for the first time, can assess, from high resolution data, that the N-terminal portion of RAD52 assembles in oligomeric ring-shaped units formed by 11 monomers in the full-length as in the truncated protein forms. Unfortunately, in our high-resolution structure of the full-length RAD52 protein, the C-terminal portion is missing.

In the attempt of explaining this lack of electron density, we used different Intrinsically Disorder Protein (IDPs) predictor softwares to shed light on this issue. The PONDR (Predictor of Natural Disordered Regions) predictor described the C-terminal domain of RAD52 as an intrinsically disordered region (PONDR score >0.5), associated to a low mean hydropathy and a relative high net charge (Fig. 62)[191,192].

The same output was obtained using another prediction tool called IUPred (Fig. 62C)[193,194]. Also from this tool, the protein C-terminal domain was predicted to be intrinsically disordered (average score > 0.5).

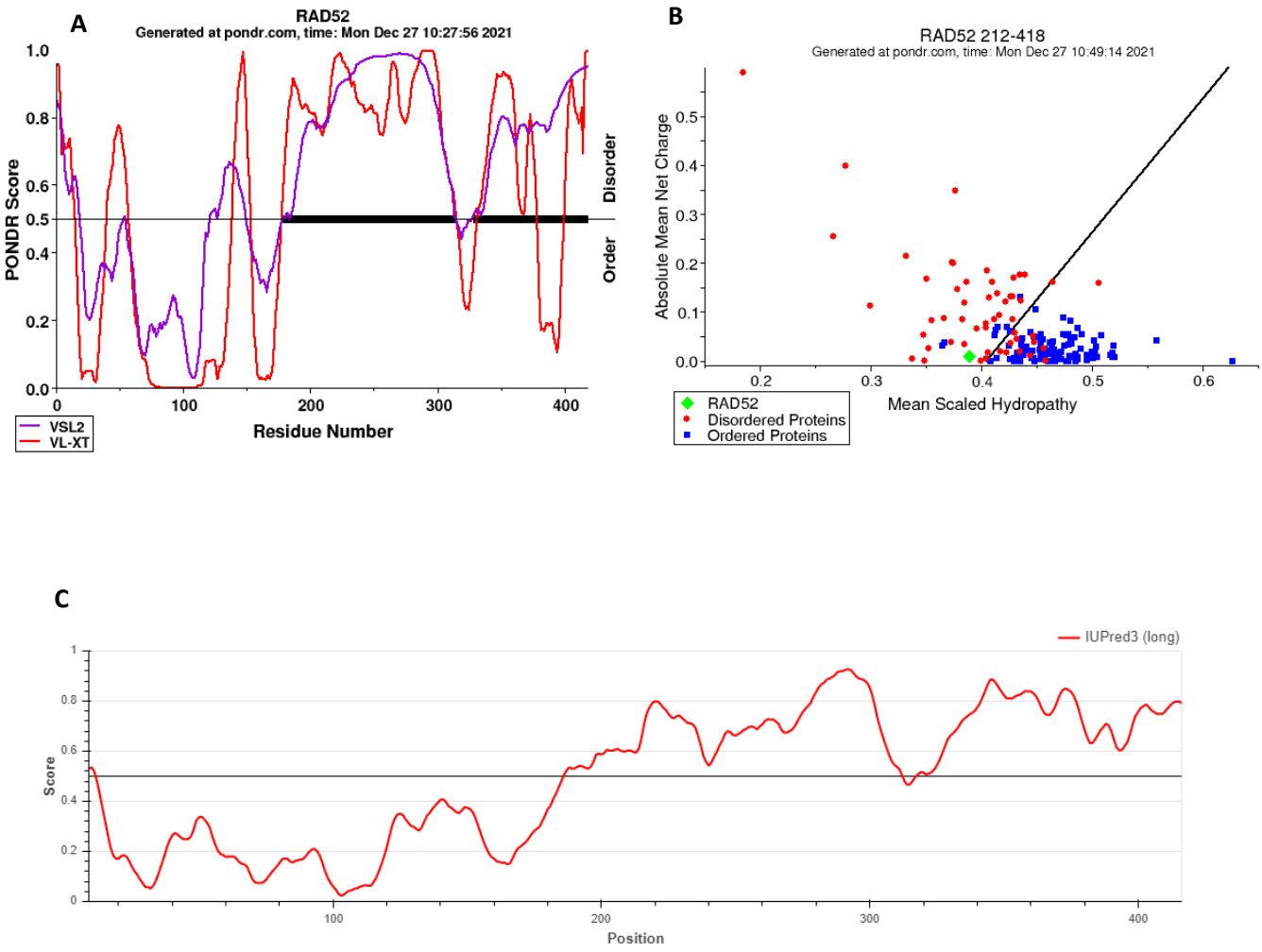


Fig. 62) A) Graph reporting the intrinsically disorder regions of RAD52 using PONDR software; B) PONDR graph reporting RAD52 C-terminal domain absolute mean net charge in function of mean scaled hydropathy, compared with other protein examples (represented as blue and red dots); C) IUPred graph reporting the intrinsically disorder prediction (IUPred score > 0.5); from both PONDR and IUPred softwares, C-terminal part of RAD52 protein is predicted to be intrinsically disordered.

Finally, RAD52 structure was analyzed using also the new algorithm AlphaFold for structure predictions[195]. This prediction reported long loops and disordered regions inside the C-terminal domain (Fig. 63), in agreement with the results obtained from the previously described prediction tools, albeit with a very low model confidence.

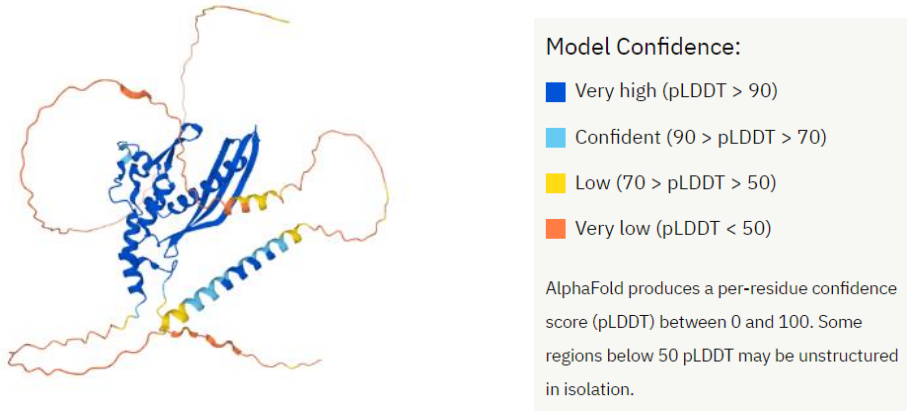


Fig. 63) AlphaFold graphical representation of ternary structure prediction of RAD52 FL. Colors of different structural elements change according the model confidence; Color legend is reported on the right.

These structural hypotheses are in agreement with our experimental data of secondary structure and thermal stability analyses. In fact, thermal stability analyses of RAD52 FL and RAD52 [1-212] showed very similar melting temperatures, suggesting similar α-helix composition of the two proteins (i.e. the melting temperature is quantified following the change in CD value at 222 nm, corresponding to α-helix component). Even though RAD52 FL has a higher MW compared to the RAD52 N-terminal domain, they have the same α-helix composition (compact secondary structure), suggesting that the additional C-terminal domain present in the RAD52 FL is mainly disordered, with few or no structured region components.

In this regard, also CD analysis of RAD52 secondary structure revealed a similar pattern of secondary structure for RAD52 FL and RAD52 [1-212], with a higher RAD52 FL disordered component (random coil 51%) compared to RAD52 [1-212] (random coil 32%).

Assuming that the C-terminal part of RAD52 is very flexible and disordered, as suggested by computational and experimental data, from cryo-EM data it does not fit inside the ring of the protein anyway, but it is likely to form an unordered "cloud" nearby the top of the structured ring. The intrinsic disorder of the C-terminal domain of the protein is closely related to the protein function. RAD52 most likely requires a highly flexible and dynamic C-terminal portion available to rapidly form superstructures by stacking several RAD52 rings, upon DNA binding to the N-terminal portion (see AFM) or to rapidly bind an interacting partner as RAD51, RPA, DSS1, when it is involved in exerting specific functions in a DNA repair pathway. Further investigations will be undertaken to corroborate these hypotheses.

Assuming that RAD52 FL forms undecameric and not heptameric ring-shaped units, as shown by cryo-EM data, the characterization of the recombinant RAD52 FL protein requires a revaluation.

SLS and DLS data showed that RAD52 FL forms at 0.8 mg/mL (16 µM) concentration, high MW superstructures constituted on average by 2-2.5 heptameric rings. Considering RAD52 as an undecameric ring, instead, these high MW superstructures would be formed, on average, by 1.3 -1.6 undecameric ring-shaped units. It should be highlighted that the number of rings that forms RAD52 high MW superstructures is dependent on protein concentration, therefore little effect is observed on the quantification of the high MW superstructures composition if the number of

monomers per ring unit changes. Nevertheless, the existence of RAD52 FL in undecameric, rather than heptameric ring units, certainly affects RAD52 – DNA interaction and its mode of action.

In parallel with the structural characterization of RAD52, also its interaction with the DNA was investigated. Herein, we optimized different techniques to validate and quantify the protein-DNA interaction. These assays were also developed for further drug discovery investigations.

The second main part of this project focused on the identification of inhibitors of RAD52 activity, which could be achieved by interfering with RAD52-DNA binding or by disrupting RAD52 heptameric/undecameric functional units. Three different approaches have been used to address this issue: a virtual screening campaign on a RAD52 N-terminal domain 3D structure available (PDB 1KN0); a fragment-based screening using NMR spectroscopy; aptamers design using computational tools.

A virtual screening campaign, performed on one of the available crystal structures of RAD52 N-terminal domain (1KN0), was pursued using an internal library of commercially available compounds. The main aim was to identify compounds, which could directly target the pocket of Tyr81, involved in protomer-protomer interaction and protein-DNA interaction of RAD52 oligomer (*materials and methods* section). From this screening, 15 molecules were ranked and selected for further biophysical evaluations. All the compounds underwent NMR and MST binding validation. Unfortunately, four of the selected compounds had severe solubility issues. Among the remaining eleven hits, five were validated as RAD52 binders. FRET analyses were also performed in order to identify whether any of selected compounds was able to inhibit RAD52-

DNA interaction and four of the five binders showed a decrease in E_{FRET}, which could be correlated to RAD52-ssDNA binding inhibition.

The four selected compounds will be further tested in cells assays and a SAR study will be pursued for the ones that will show in cells inhibitory activity.

^{19}F NMR FBA screening of an internal library of fluorinated compounds on RAD52 FL allowed the identification of several RAD52 binders. We selected E5 as the most interesting compound for further characterizations. In fact, in a parallel ongoing research work of our group (unpublished data), E5 is reported to inhibit RAD51-BRCA2 interaction by binding RAD51, a critical player of the HR DSB repair pathway. Herein, through *in vitro* and *in cellulo* experiments, we demonstrated that E5 is a potential inhibitor of RAD52, able to prevent its binding to DNA, it does significantly affect viability in cell assays and it displays synergistic effect with olaparib compound (PARPi). The presented data suggest that E5 is a promising starting point that deserves further investigation for the development of a potential "dual" RAD51-RAD52 inhibitor. The simultaneous inhibition of two targets, RAD51 and RAD52, involved in DSB repair mechanisms, triggers synthetic lethality in adenocarcinoma pancreatic cell lines. Further investigations are necessary to corroborate the presented data, but, if proved, E5 may be the progenitor of a class of potent compounds able to generalize the "BRCA2-ness" paradigm, simultaneously inhibiting the HR DSB direct repair mechanism (RAD51) and RAD52-mediated alternative repair pathways (fig. 64).

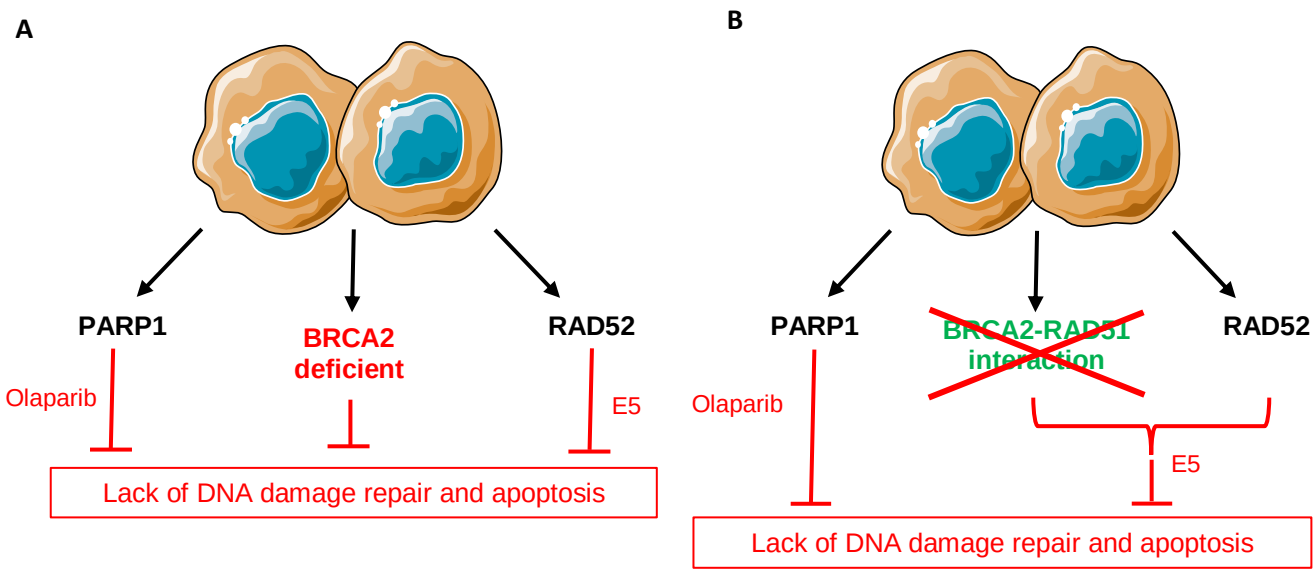


Fig. 64) General overview explaining E5 mechanism of action: A) the Inhibition of RAD52 through E5, with PARP inhibition in BRCA2 deficient-cancer cells is an implementation of synthetic lethality therapy and can avoid resistance occurrence and side effects; B) Since E5 emerged to have effect both on RAD51 and RAD52, E5 treatment could affect two intracellular pathways in parallel, boosting the effect of synthetic lethality strategy with olaparib, generalizing "BRCA-ness" paradigm.

Finally, RAD52 aptamer inhibitors were also designed in order to find a valid alternative to small molecules inhibitors. Aptamers in these years, are emerging as a promising strategy to target intracellular proteins, with better tissues penetration, easier production pipeline and automated multiplexing compared to small molecules[181,182].

Aptamer sequences were designed and ranked through catRAPID algorithm. The sequences were designed in order to target RAD52 DNA binding pocket. Out from the 11 selected sequences, 4 came out with a very promising k_d values (20-200 nM). Notably, all the best binders showed a CD spectrum resembling a G-quadruplex structure, whereas not present in any of the weak binders, but one. Computational analyses and experimental assays were in agreement identifying Apt1, Apt4, Apt6 and Apt9 as promising binders of RAD52. Apt1 as the best identified hit was selected

for further investigations. We evaluated its *in vitro* activity on RAD52 superstructures disruption and its cellular internalization levels in pancreatic adenocarcinoma cancer cells. From these preliminary investigations, Apt1 shows promising features; however, further studies are required to evaluate its inhibitory activity in vitro and in cells.

Two are the main issues of aptamer inhibitors designing strategy: 1) the secondary structure assumed by the oligonucleotides sequences is strictly related to buffer composition and cations concentration[196]: this means that each assay should be performed in a condition as close as possible to the physiological one; however, this is not always possible and it may lead to artefacts; 2) despite their solubility and optimal tissue penetration, aptamers may have problems *in vivo*, due to fast renal clearance, nucleases degradation and toxicity[183]. To solve the renal clearance problem, delivery systems have been developed (such as lipid nanoparticles and GalNAc bio conjugates) to avoid fast aptamers elimination[197,198]; to avoid nuclease activity on aptamers, internal modification, such as fluorine or O'-methyl groups, are commonly performed. Notably, many studies are ongoing on aptamers toxicity, aptamers vectors *in vivo* and on chemical modifications impact[199]. Nevertheless, so far aptamers have shown many promising preliminary evidences in their applications in therapy[181,182]; therefore, it is important to keep on focusing on aptamers study to gain further knowledge about their main hallmarks and to increase their potential applications in novel therapies.

This PhD project has made it possible to acquire valuable knowledge on RAD52 protein, from the structural, functional and medicinal chemistry point of views. Nevertheless, many are the ongoing studies originated from this research work. Studies are set up in order to increase data quality of RAD52 cryo-EM structure also in the presence of interactors, trying to stabilize the C-

terminal domain of the protein. In order to gain additional structural information on RAD52, crystallization trials have been recently designed and will be performed shortly, for both RAD52 alone and in the presence of interactors or promising inhibitors.

Notably, *in cellulo* experiments are also ongoing for RAD52 inhibitors validation. Specifically, we are currently designing stable RAD52 silenced pancreatic cancer cell lines in order to characterize the effect induced by the selected RAD52 inhibitors and to identify their intracellular targets, confirming either their RAD52 specificity or dual effect on RAD51 and RAD52. This strategy have been widely used in literature for RAD52 inhibitors characterization[156,200].

In conclusion, RAD52 is a potential novel therapeutic target for synthetic lethality strategies. We have here identified and characterized potential novel inhibitors (both oligonucleotides and small molecules) using different approaches. These compounds are promising starting points for the development of new RAD52 inhibitors to be uses in synthetic-lethality cancer therapies, with the goal of preventing side effects on normal cells, boosting the toxic effect on cancer cells and avoiding resistance occurrence.

6. References

1. Basu, A. K. DNA Damage, Mutagenesis and Cancer. *Int. J. Mol. Sci.* **19**, (2018).
2. Alhmoud, J. F., Woolley, J. F., Al Moustafa, A.-E. & Malki, M. I. DNA Damage/Repair Management in Cancers. *Cancers (Basel).* **12**, (2020).
3. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* **26**, 52–64 (2016).
4. 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.* **20**, 698–714 (2019).
5. Symington, L. S. Mechanism and regulation of DNA end resection in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **51**, 195–212 (2016).
6. Sartori, A. A. *et al.* Human CtIP promotes DNA end resection. *Nature* **450**, 509–514 (2007).
7. Cejka, P. DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination. *J. Biol. Chem.* **290**, 22931–22938 (2015).
8. Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol. Cell* **50**, 589–600 (2013).
9. Jensen, R. B., Carreira, A. & Kowalczykowski, S. C. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* **467**, 678–683 (2010).
10. Sy, S. M. H., Huen, M. S. Y. & Chen, J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7155–7160 (2009).
11. van der Heijden, T. *et al.* Homologous recombination in real time: DNA strand exchange by RecA. *Mol. Cell* **30**, 530–538 (2008).
12. Chun, J., Buechelmaier, E. S. & Powell, S. N. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the BRCA1-BRCA2-dependent homologous recombination pathway. *Mol. Cell. Biol.* **33**, 387–395 (2013).
13. McVey, M., Khodaverdian, V. Y., Meyer, D., Cerqueira, P. G. & Heyer, W.-D. Eukaryotic DNA Polymerases in Homologous Recombination. *Annu. Rev. Genet.* **50**, 393–421 (2016).
14. Meyer, D., Fu, B. X. H. & Heyer, W.-D. DNA polymerases δ and λ cooperate in repairing double-strand breaks by microhomology-mediated end-joining in Saccharomyces cerevisiae *Proc. Natl. Acad. Sci.* **112**, E6907 LP-E6916 (2015).
15. Kane, D. P., Shusterman, M., Rong, Y. & McVey, M. Competition between replicative and translesion polymerases during homologous recombination repair in Drosophila. *PLoS Genet.* **8**, e1002659 (2012).
16. Sneeden, J. L., Grossi, S. M., Tappin, I., Hurwitz, J. & Heyer, W.-D. Reconstitution of recombination-associated DNA synthesis with human proteins. *Nucleic Acids Res.* **41**, 4913–4925 (2013).

17. Chiruvella, K. K., Liang, Z. & Wilson, T. E. Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* **5**, a012757 (2013).

18. Difilippantonio, M. J. *et al.* DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. *Nature* **404**, 510–514 (2000).

19. Gottlieb, T. M. & Jackson, S. P. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72**, 131–142 (1993).

20. Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol. Cell. Biol.* **20**, 2996–3003 (2000).

21. Ochi, T. *et al.* DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347**, 185–188 (2015).

22. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **124**, 301–313 (2006).

23. Graham, T. G. W., Walter, J. C. & Loparo, J. J. Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol. Cell* **61**, 850–858 (2016).

24. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495–506 (2017).

25. Xie, A., Kwok, A. & Scully, R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol.* **16**, 814–818 (2009).

26. Grundy, G. J. *et al.* APLF promotes the assembly and activity of non-homologous end joining protein complexes. *EMBO J.* **32**, 112–125 (2013).

27. Hanscom, T. & McVey, M. Regulation of Error-Prone DNA Double-Strand Break Repair and Its Impact on Genome Evolution. *Cells* **9**, (2020).

28. Black, S. J., Kashkina, E., Kent, T. & Pomerantz, R. T. DNA Polymerase θ: A Unique Multifunctional End-Joining Machine. *Genes (Basel).* **7**, (2016).

29. Wood, R. D. & Doublié, S. DNA polymerase θ (POLQ), double-strand break repair, and cancer. *DNA Repair* (2016) doi:10.1016/j.dnarep.2016.05.003.

30. Mateos-Gomez, P. A. *et al.* Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* (2015) doi:10.1038/nature14157.

31. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci.* **111**, E5593–E5601 (2014).

32. Kent, T., Chandramouly, G., McDevitt, S. M., Ozdemir, A. Y. & Pomerantz, R. T. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. *Nat. Struct. Mol. Biol.* **22**, 230–237 (2015).

33. Rothenberg, E., Grimme, J. M., Spies, M. & Ha, T. Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20274–20279 (2008).

34. Bhargava, R., Onyango, D. O. & Stark, J. M. Regulation of Single Strand Annealing and its role in

genome maintenance Chromosomal break repair by the Single Strand Annealing (SSA) pathway. *Trends Genet* **32**, 566–575 (2016).

35. Motycka, T. A., Bessho, T., Post, S. M., Sung, P. & Tomkinson, A. E. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J. Biol. Chem.* **279**, 13634–13639 (2004).

36. Bridges, C. B. The Origin of Variations in Sexual and Sex-Limited Characters. *Am. Nat.* (1922) doi:10.1086/279847.

37. DOBZHANSKY, T. Genetics of natural populations; recombination and variability in populations of Drosophila pseudoobscura. *Genetics* (1946) doi:10.1093/genetics/31.3.269.

38. Ashworth, A. & Lord, C. J. Synthetic lethal therapies for cancer: what's next after PARP inhibitors? *Nature Reviews Clinical Oncology* (2018) doi:10.1038/s41571-018-0055-6.

39. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating genetic approaches into the discovery of anticancer drugs. *Science* (1997) doi:10.1126/science.278.5340.1064.

40. Kaelin, W. G. The concept of synthetic lethality in the context of anticancer therapy. *Nature Reviews Cancer* (2005) doi:10.1038/nrc1691.

41. Huang, A., Garraway, L. A., Ashworth, A. & Weber, B. Synthetic lethality as an engine for cancer drug target discovery. *Nature Reviews Drug Discovery* (2020) doi:10.1038/s41573-019-0046-z.

42. Topatana, W. *et al.* Advances in synthetic lethality for cancer therapy: Cellular mechanism and clinical translation. *Journal of Hematology and Oncology* (2020) doi:10.1186/s13045-020-00956-5.

43. Fong, P. C. *et al.* Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers . *N. Engl. J. Med.* (2009) doi:10.1056/nejmoa0900212.

44. Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: A proof-of-concept trial. *Lancet* (2010) doi:10.1016/S0140-6736(10)60893-8.

45. Lord, C. J. & Ashworth, A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nature Medicine* (2013) doi:10.1038/nm.3369.

46. Howard, S. M., Yanez, D. A. & Stark, J. M. DNA Damage Response Factors from Diverse Pathways, Including DNA Crosslink Repair, Mediate Alternative End Joining. *PLoS Genet.* (2015) doi:10.1371/journal.pgen.1004943.

47. Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ -mediated repair. *Nature* (2015) doi:10.1038/nature14184.

48. Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* (2017) doi:10.1126/science.aam7344.

49. D'Andrea, A. D. Mechanisms of PARP inhibitor sensitivity and resistance. *DNA Repair* (2018) doi:10.1016/j.dnarep.2018.08.021.

50. Carneiro, B. A. *et al.* Acquired Resistance to Poly (ADP-ribose) Polymerase Inhibitor Olaparib in

BRCA2-Associated Prostate Cancer Resulting From Biallelic BRCA2 Reversion Mutations Restores Both Germline and Somatic Loss-of-Function Mutations. *JCO Precis. Oncol.* **2**, (2018).

51. Waks, A. G. *et al.* Reversion and non-reversion mechanisms of resistance to PARP inhibitor or platinum chemotherapy in BRCA1/2-mutant metastatic breast cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **31**, 590–598 (2020).

52. Li, S. *et al.* Development of synthetic lethality in cancer: molecular and cellular classification. *Signal Transduction and Targeted Therapy* (2020) doi:10.1038/s41392-020-00358-6.

53. Toma, M., Sullivan-Reed, K., Śliwiński, T. & Skorski, T. RAD52 as a potential target for synthetic lethality-based anticancer therapies. *Cancers* (2019) doi:10.3390/cancers11101561.

54. Feng, Z. *et al.* Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 686–691 (2011).

55. Cramer-Morales, K. *et al.* Personalized synthetic lethality induced by targeting RAD52 in leukemias identified by gene mutation and expression profile. *Blood* (2013) doi:10.1182/blood-2013-05-501072.

56. Lok, B. H., Carley, A. C., Tchang, B. & Powell, S. N. RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination. *Oncogene* (2013) doi:10.1038/onc.2012.391.

57. Sullivan-Reed, K. *et al.* Simultaneous Targeting of PARP1 and RAD52 Triggers Dual Synthetic Lethality in BRCA-Deficient Tumor Cells. *Cell Rep.* (2018) doi:10.1016/j.celrep.2018.05.034.

58. Rossi, M. J., Didomenico, S. F., Patel, M. & Mazin, A. V. RAD52 : Paradigm of Synthetic Lethality and New Developments. **12**, 1–16 (2021).

59. Kito, K., Wada, H., Yeh, E. T. H. & Kamitani, T. Identification of novel isoforms of human RAD52. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1489**, 303–314 (1999).

60. Lok, B. H. & Powell, S. N. Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement. *Clin. cancer Res. an Off. J. Am. Assoc. Cancer Res.* **18**, 6400–6406 (2012).

61. Kagawa, W. *et al.* Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol. Cell* **10**, 359–371 (2002).

62. Shen, Z., Cloud, K. G., Chen, D. J. & Park, M. S. Specific interactions between the human RAD51 and RAD52 proteins. *J. Biol. Chem.* **271**, 148–152 (1996).

63. Ranatunga, W. *et al.* Human RAD52 Exhibits Two Modes of Self-association. *J. Biol. Chem.* **276**, 15876–15880 (2001).

64. Stasiak, A. Z. *et al.* The human Rad52 protein exists as a heptameric ring. *Curr. Biol.* **10**, 337–340 (2000).

65. Van Dyck, E., Hajibagheri, N. M. A., Stasiak, A. & West, S. C. Visualisation of human Rad52 protein and its complexes with hRad51 and DNA. *J. Mol. Biol.* **284**, 1027–1038 (1998).

66. Van Dyck, E., Stasiak, A. Z., Stasiak, A. & West, S. C. Visualization of recombination intermediates produced by RAD52-mediated single-strand annealing. *EMBO Rep.* **2**, 905–909 (2001).

67. Ranatunga, W., Jackson, D., Flowers, R. A. & Borgstahl, G. E. O. Human RAD52 protein has extreme thermal stability. *Biochemistry* **40**, 8557–8562 (2001).

68. Symington, L. S. Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair . *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).

69. Saotome, M. *et al.* Structural Basis of Homology-Directed DNA Repair Mediated by RAD52. *iScience* **3**, 50–62 (2018).

70. Kagawa, W., Kurumizaka, H., Ikawa, S., Yokoyama, S. & Shibata, T. Homologous Pairing Promoted by the Human, Rad52 Protein. *J. Biol. Chem.* **276**, 35201–35208 (2001).

71. Kagawa, W. *et al.* Identification of a second DNA binding site in the human Rad52 protein. *J. Biol. Chem.* **283**, 24264–24273 (2008).

72. Dyck, E. Van, Stasiak, A. Z., Stasiak, A. & West, S. C. Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* **401**, 371–375 (1999).

73. Chandramouly, G. *et al.* Small-Molecule Disruption of RAD52 Rings as a Mechanism for Precision Medicine in BRCA-Deficient Cancers. *Chem. Biol.* **22**, 1491–1504 (2015).

74. Grimme, J. M. & Spies, M. FRET-based assays to monitor DNA binding and annealing by Rad52 recombination mediator protein. *Methods Mol. Biol.* **745**, 463–483 (2011).

75. Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by Rad52 Protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl. Acad. Sci.* **95**, 6049 LP – 6054 (1998).

76. Mortensen, U. H., Bendixen, C., Sunjevaric, I. & Rothstein, R. DNA strand annealing is promoted by the yeast Rad52 protein. *Proc. Natl. Acad. Sci.* **93**, 10729 LP – 10734 (1996).

77. Lloyd, J. A., McGrew, D. A. & Knight, K. L. Identification of residues important for DNA binding in the full-length human Rad52 protein. *J. Mol. Biol.* **345**, 239–249 (2005).

78. Hanamshet, K. & Mazin, A. V. The function of RAD52 N-terminal domain is essential for viability of BRCA-deficient cells. *Nucleic Acids Res.* **48**, 12778–12791 (2020).

79. Grimme, J. M. *et al.* Human Rad52 binds and wraps single-stranded DNA and mediates annealing via two hRad52-ssDNA complexes. *Nucleic Acids Res.* **38**, 2917–2930 (2010).

80. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).

81. Sugiman-Marangos, S. N., Weiss, Y. M. & Junop, M. S. Mechanism for accurate, protein-assisted DNA annealing by Deinococcus radiodurans DdrB. *Proc. Natl. Acad. Sci.* **113**, 4308 LP – 4313 (2016).

82. Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C. & Wigley, D. B. Structure of the single-strand annealing domain of human RAD52 protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13492–13497 (2002).

83. Ma, C. J., Kwon, Y., Sung, P. & Greene, E. C. Human RAD52 interactions with replication protein a and the RAD51 presynaptic complex. *J. Biol. Chem.* **292**, 11702–11713 (2017).

84. Jalan, M., Olsen, K. S. & Powell, S. N. Emerging roles of RAD52 in genome maintenance. *Cancers*

(Basel). **11**, (2019).

85. Brouwer, I. *et al.* Human RAD52 Captures and Holds DNA Strands, Increases DNA Flexibility, and Prevents Melting of Duplex DNA: Implications for DNA Recombination. *Cell Rep.* **18**, 2845–2853 (2017).

86. Welty, S. *et al.* RAD52 is required for RNA-templated recombination repair in post-mitotic neurons. *J. Biol. Chem.* **293**, 1353–1362 (2018).

87. Mazina, O. M., Keskin, H., Hanamshet, K., Storici, F. & Mazin, A. V. Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Mol. Cell* **67**, 19-29.e3 (2017).

88. Yasuda, T. *et al.* Novel function of HATs and HDACs in homologous recombination through acetylation of human RAD52 at double-strand break sites. *PLoS Genet.* (2018) doi:10.1371/journal.pgen.1007277.

89. Yasuda, T. *et al.* Human SIRT2 and SIRT3 deacetylases function in DNA homologous recombinational repair. *Genes to Cells* (2021) doi:10.1111/gtc.12842.

90. Charifi, F. *et al.* Rad52 SUMOylation functions as a molecular switch that determines a balance between the Rad51- and Rad59-dependent survivors. *iScience* (2021) doi:10.1016/j.isci.2021.102231.

91. Saito, K. *et al.* The putative nuclear localization signal of the human RAD52 protein is a potential sumoylation site. *J. Biochem.* (2010) doi:10.1093/jb/mvq020.

92. Cramer, K. *et al.* BCR/ABL and other kinases from chronic myeloproliferative disorders stimulate single-strand annealing, an unfaithful DNA double-strand break repair. *Cancer Res.* (2008) doi:10.1158/0008-5472.CAN-08-1101.

93. Fernandes, M. S. *et al.* BCR-ABL promotes the frequency of mutagenic single-strand annealing DNA repair. *Blood* (2009) doi:10.1182/blood-2008-07-172148.

94. Honda, M., Okuno, Y., Yoo, J., Ha, T. & Spies, M. Tyrosine phosphorylation enhances RAD52-mediated annealing by modulating its DNA binding. *EMBO J.* (2011) doi:10.1038/emboj.2011.238.

95. New, J. H., Sugiyama, T., Zaitseva, E. & Kowalczykowski, S. C. Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A. *Nature* **391**, 407–410 (1998).

96. Sugiyama, T., Kantake, N., Wu, Y. & Kowalczykowski, S. C. Rad52-mediated DNA annealing after Rad51-mediated DNA strand exchange promotes second ssDNA capture. *EMBO J.* **25**, 5539–5548 (2006).

97. Hanamshet, K., Mazina, O. M. & Mazin, A. V. Reappearance from obscurity: Mammalian Rad52 in homologous recombination. *Genes (Basel).* **7**, 1–18 (2016).

98. Rijkers, T. *et al.* Targeted inactivation of mouse RAD52 reduces homologous recombination but not resistance to ionizing radiation. *Mol. Cell. Biol.* **18**, 6423–6429 (1998).

99. Yamaguchi-Iwai, Y. *et al.* Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Mol. Cell. Biol.* **18**, 6430–6435 (1998).

100. Mahajan, S., Raina, K., Verma, S. & Rao, B. J. Human RAD52 protein regulates homologous

recombination and checkpoint function in BRCA2 deficient cells. *Int. J. Biochem. Cell Biol.* **107**, 128–139 (2019).

101. McIlwraith, M. J. & West, S. C. DNA Repair Synthesis Facilitates RAD52-Mediated Second-End Capture during DSB Repair. *Mol. Cell* **29**, 510–516 (2008).

102. Sugiyama, T. & Kantake, N. Dynamic Regulatory Interactions of Rad51, Rad52, and Replication Protein-A in Recombination Intermediates. *J. Mol. Biol.* **390**, 45–55 (2009).

103. Blasiak, J. Single-strand annealing in cancer. *Int. J. Mol. Sci.* **22**, 1–24 (2021).

104. Chen, L., Nievera, C. J., Lee, A. Y.-L. & Wu, X. Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J. Biol. Chem.* **283**, 7713–7720 (2008).

105. Al-Minawi, A. Z., Saleh-Gohari, N. & Helleday, T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res.* **36**, 1–9 (2008).

106. Stefanovie, B. *et al.* DSS1 interacts with and stimulates RAD52 to promote the repair of DSBs. *Nucleic Acids Res.* **48**, 694–708 (2020).

107. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).

108. Kondratick, C. M., Washington, M. T. & Spies, M. Making Choices: DNA Replication Fork Recovery Mechanisms. *Semin. Cell Dev. Biol.* **113**, 27–37 (2021).

109. Malacaria, E. *et al.* Rad52 prevents excessive replication fork reversal and protects from nascent strand degradation. *Nat. Commun.* **10**, 1–19 (2019).

110. Mijic, S. *et al.* Replication fork reversal triggers fork degradation in BRCA2-defective cells. *Nat. Commun.* **8**, 859 (2017).

111. Murfuni, I. *et al.* Survival of the replication checkpoint deficient cells requires MUS81-RAD52 function. *PLoS Genet.* **9**, e1003910 (2013).

112. Malkova, A. & Ira, G. Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.* **23**, 271–279 (2013).

113. Costantino, L. *et al.* Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* **343**, 88–91 (2014).

114. Sotiriou, S. K. *et al.* Mammalian RAD52 Functions in Break-Induced Replication Repair of Collapsed DNA Replication Forks. *Mol. Cell* **64**, 1127–1134 (2016).

115. Sharma, N. *et al.* Distinct roles of structure-specific endonucleases EEPD1 and Metnase in replication stress responses. *NAR cancer* **2**, zcaa008 (2020).

116. Bhowmick, R., Minocherhomji, S. & Hickson, I. D. RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. *Mol. Cell* **64**, 1117–1126 (2016).

117. Helmrich, A., Ballarino, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).

118. Lukas, C. *et al.* 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* **13**, 243–253 (2011).

119. Spies, J. *et al.* 53BP1 nuclear bodies enforce replication timing at under-replicated DNA to limit heritable DNA damage. *Nat. Cell Biol.* **21**, 487–497 (2019).

120. Zhang, J.-M., Yadav, T., Ouyang, J., Lan, L. & Zou, L. Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. *Cell Rep.* **26**, 955-968.e3 (2019).

121. Shen, Y. *et al.* RNA-driven genetic changes in bacteria and in human cells. *Mutat. Res.* **717**, 91–98 (2011).

122. Keskin, H. *et al.* Transcript-RNA-templated DNA recombination and repair. *Nature* **515**, 436–439 (2014).

123. McDevitt, S., Rusanov, T., Kent, T., Chandramouly, G. & Pomerantz, R. T. How RNA transcripts coordinate DNA recombination and repair. *Nat. Commun.* **9**, 1–10 (2018).

124. Chandramouly, G. *et al.* Polθ reverse transcribes RNA and promotes RNA-templated DNA repair. *Sci. Adv.* **7**, (2021).

125. Su, Y. *et al.* Human DNA polymerase η has reverse transcriptase activity in cellular environments. *J. Biol. Chem.* **294**, 6073–6081 (2019).

126. Meers, C. *et al.* Genetic Characterization of Three Distinct Mechanisms Supporting RNA-Driven DNA Repair and Modification Reveals Major Role of DNA Polymerase ζ. *Mol. Cell* **79**, 1037-1050.e5 (2020).

127. Aymard, F. *et al.* Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **21**, 366–374 (2014).

128. Teng, Y. *et al.* ROS-induced R loops trigger a transcription-coupled but BRCA1/2-independent homologous recombination pathway through CSB. *Nat. Commun.* **9**, (2018).

129. Yasuhara, T. *et al.* Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* **175**, 558-570.e11 (2018).

130. Chen, H. *et al.* m5C modification of mRNA serves a DNA damage code to promote homologous recombination. *Nat. Commun.* **11**, 2834 (2020).

131. Wang, J. *et al.* RAD52 Adjusts Repair of Single-Strand Breaks via Reducing DNA-Damage-Promoted XRCC1/LIG3α Co-localization. *Cell Rep.* **34**, 108625 (2021).

132. Sakofsky, C. J. & Malkova, A. Break induced replication in eukaryotes: mechanisms, functions, and consequences. *Crit. Rev. Biochem. Mol. Biol.* **52**, 395–413 (2017).

133. Lieberman, R. & You, M. Corrupting the DNA damage response: A critical role for Rad52 in tumor cell survival. *Aging (Albany. NY).* **9**, 1647–1659 (2017).

134. Hironaka, K., Factor, V. M., Calvisi, D. F., Conner, E. A. & Thorgeirsson, S. S. Dysregulation of DNA repair pathways in a transforming growth factor alpha/c-myc transgenic mouse model of accelerated hepatocarcinogenesis. *Lab. Invest.* **83**, 643–654 (2003).

135. Treuner, K., Helton, R. & Barlow, C. Loss of Rad52 partially rescues tumorigenesis and T-cell maturation in Atm-deficient mice. *Oncogene* **23**, 4655–4661 (2004).

136. Lieberman, R., Pan, J., Zhang, Q. & You, M. Rad52 deficiency decreases development of lung squamous cell carcinomas by enhancing immuno-surveillance. *Oncotarget* **8**, 34032–34044 (2017).

137. Kohzaki, M., Ootsuyama, A., Sun, L., Moritake, T. & Okazaki, R. Human RECQL4 represses the RAD52-mediated single-strand annealing pathway after ionizing radiation or cisplatin treatment. *Int. J. Cancer* **146**, 3098–3113 (2020).

138. Shi, T.-Y. *et al.* RAD52 variants predict platinum resistance and prognosis of cervical cancer. *PLoS One* **7**, e50461 (2012).

139. Ho, V. *et al.* Aberrant Expression of RAD52, Its Prognostic Impact in Rectal Cancer and Association with Poor Survival of Patients. *Int. J. Mol. Sci.* **21**, (2020).

140. Kan, Y., Batada, N. N. & Hendrickson, E. A. Human somatic cells deficient for RAD52 are impaired for viral integration and compromised for most aspects of homology-directed repair. *DNA Repair (Amst).* **55**, 64–75 (2017).

141. Adamson, A. W. *et al.* The RAD52 S346X variant reduces risk of developing breast cancer in carriers of pathogenic germline BRCA2 mutations. *Mol. Oncol.* **14**, 1124–1133 (2020).

142. Li, Z. *et al.* Association of a functional RAD52 genetic variant locating in a miRNA binding site with risk of HBV-related hepatocellular carcinoma. *Mol. Carcinog.* **54**, 853–858 (2015).

143. Naccarati, A. *et al.* Double-strand break repair and colorectal cancer: gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. *Oncotarget* **7**, 23156–23169 (2016).

144. Sharma, A. B. *et al.* XAB2 promotes Ku eviction from single-ended DNA double-strand breaks independently of the ATM kinase. *Nucleic Acids Res.* **49**, 9906–9925 (2021).

145. Hromas, R. *et al.* The endonuclease EEPD1 mediates synthetic lethality in RAD52-depleted BRCA1 mutant breast cancer cells. *Breast Cancer Res.* **19**, 122 (2017).

146. Xu, Y. *et al.* RAD52 aptamer regulates DNA damage repair and STAT3 in BRCA1/BRCA2-deficient human acute myeloid leukemia. *Oncol. Rep.* **44**, 1455–1466 (2020).

147. Al Zaid Siddiquee, K. & Turkson, J. STAT3 as a target for inducing apoptosis in solid and hematological tumors. *Cell Res.* **18**, 254–267 (2008).

148. Bruserud, Ø., Nepstad, I., Hauge, M., Hatfield, K. J. & Reikvam, H. STAT3 as a possible therapeutic target in human malignancies: lessons from acute myeloid leukemia. *Expert review of hematology* vol. 8 29–41 (2015).

149. Gunn, A. & Stark, J. M. I-SceI-based assays to examine distinct repair outcomes of mammalian chromosomal double strand breaks. *Methods Mol. Biol.* **920**, 379–391 (2012).

150. Olney, J. W. *et al.* Excitotoxicity of L-dopa and 6-OH-dopa: implications for Parkinson's and Huntington's diseases. *Exp. Neurol.* **108**, 269–272 (1990).

151. Huang, F. *et al.* Targeting BRCA1-and BRCA2-deficient cells with RAD52 small molecule inhibitors. *Nucleic Acids Res.* **44**, 4189–4199 (2016).

152. Hengel, S. R. *et al.* Small-molecule inhibitors identify the RAD52-ssDNA interaction as critical for

recovery from replication stress and for survival of BRCA2 deficient cells. *Elife* **5**, 1–30 (2016).

153. Al-Mugotir, M. *et al.* Selective killing of homologous recombination-deficient cancer cell lines by inhibitors of the RPA:RAD52 protein-protein interaction. *PLoS One* **16**, 1–22 (2021).

154. Sullivan, K. *et al.* Identification of a Small Molecule Inhibitor of RAD52 by Structure-Based Selection. *PLoS One* **11**, 1–11 (2016).

155. Li, J. *et al.* Compound F779-0434 causes synthetic lethality in BRCA2-deficient cancer cells by disrupting RAD52-ssDNA association. *RSC Adv.* **8**, 18859–18869 (2018).

156. Yang, Q., Li, Y., Sun, R. & Li, J. Identification of a RAD52 Inhibitor Inducing Synthetic Lethality in BRCA2-Deficient Cancer Cells. *Front. Pharmacol.* **12**, 1–10 (2021).

157. Tseng, W. C. *et al.* Targeting hr repair as a synthetic lethal approach to increase dna damage sensitivity by a rad52 inhibitor in brca2-deficient cancer cells. *Int. J. Mol. Sci.* **22**, (2021).

158. Bagnolini, G. *et al.* 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.* **63**, 2588–2619 (2020).

159. Roberti, M. *et al.* Rad51/BRCA2 disruptors inhibit homologous recombination and synergize with olaparib in pancreatic cancer cells. *Eur. J. Med. Chem.* **165**, 80–92 (2019).

160. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

161. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

162. Wu, Y., Siino, J. S., Sugiyama, T. & Kowalczykowski, S. C. The DNA binding preference of RAD52 and RAD59 proteins: Implications for RAD52 and RAD59 protein function in homologous recombination. *J. Biol. Chem.* **281**, 40001–40009 (2006).

163. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* (1958) doi:10.1063/1.1716296.

164. Dalvit, C. *et al.* Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **18**, 65–68 (2000).

165. Fischel, J.-L., Formento, P. & Milano, G. Epidermal growth factor receptor double targeting by a tyrosine kinase inhibitor (Iressa) and a monoclonal antibody (Cetuximab). Impact on cell growth and molecular factors. *Br. J. Cancer* **92**, 1063–1068 (2005).

166. dos Santos Ferreira, A. C., Fernandes, R. A., Kwee, J. K. & Klumb, C. E. Histone deacetylase inhibitor potentiates chemotherapy-induced apoptosis through Bim upregulation in Burkitt's lymphoma cells. *J. Cancer Res. Clin. Oncol.* **138**, 317–325 (2012).

167. Falchi, F., Caporuscio, F. & Recanatini, M. Structure-based design of small-molecule protein-protein interaction modulators: the story so far. *Future Med. Chem.* **6**, 343–357 (2014).

168. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).

169. Beard, H., Cholleti, A., Pearlman, D., Sherman, W. & Loving, K. A. Applying physics-based scoring

to calculate free energies of binding for single amino acid mutations in protein-protein complexes. *PLoS One* **8**, 1–11 (2013).

170. Li, J. *et al.* The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. *Proteins* **79**, 2794–2812 (2011).

171. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).

172. Agostini, F. *et al.* catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* **29**, 2928–2930 (2013).

173. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).

174. Toma, M., Sullivan-Reed, K., Śliwiński, T. & Skorski, T. RAD52 as a potential target for synthetic lethality-based anticancer therapies. *Cancers (Basel).* **11**, (2019).

175. Nogueira, A., Fernandes, M., Catarino, R. & Medeiros, R. RAD52 functions in homologous recombination and its importance on genomic integrity maintenance and cancer therapy. *Cancers (Basel).* **11**, (2019).

176. Dalvit, C., Caronni, D., Mongelli, N., Veronesi, M. & Vulpetti, A. NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Current Drug Discovery Technologies* vol. 3 115–124 (2006).

177. Dalvit, C., Fagerness, P. E., Hadden, D. T. A., Sarver, R. W. & Stockman, B. J. Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. *J. Am. Chem. Soc.* **125**, 7696–7703 (2003).

178. Golan, T. *et al.* Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. *N. Engl. J. Med.* **381**, 317–327 (2019).

179. Pecoraro, C. *et al.* GSK3β as a novel promising target to overcome chemoresistance in pancreatic cancer. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* **58**, 100779 (2021).

180. Kuo, L. J. & Yang, L.-X. Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **22**, 305–309 (2008).

181. Sun, H. *et al.* Oligonucleotide aptamers: new tools for targeted cancer therapy. *Mol. Ther. Nucleic Acids* **3**, e182 (2014).

182. Yang, S. *et al.* Oligonucleotide Aptamer-Mediated Precision Therapy of Hematological Malignancies. *Mol. Ther. Nucleic Acids* **13**, 164–175 (2018).

183. Kovacevic, K. D., Gilbert, J. C. & Jilma, B. Pharmacokinetics, pharmacodynamics and safety of aptamers. *Adv. Drug Deliv. Rev.* **134**, 36–50 (2018).

184. Del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **57**, 7171–7175 (2018).

185. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2**, 123–136 (2020).

186. Hon, J., Martínek, T., Zendulka, J. & Lexa, M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* **33**, 3373–3379 (2017).

187. Puig Lombardi, E. & Londoño-Vallejo, A. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* **48**, 1–15 (2020).

188. Miskiewicz, J., Sarzynska, J. & Szachniuk, M. How bioinformatics resources work with G4 RNAs. *Brief. Bioinform.* **22**, bbaa201 (2021).

189. Benson, F. E., Baumann, P. & West, S. C. Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. *Nature* **391**, 401–404 (1998).

190. Saotome, M., Saito, K., Onodera, K., Kurumizaka, H. & Kagawa, W. Structure of the human DNA-repair protein RAD52 containing surface mutations. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **72**, 598–603 (2016).

191. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. & Dunker, A. K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **61 Suppl 7**, 176–182 (2005).

192. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).

193. Erdős, G. & Dosztányi, Z. Analyzing Protein Disorder with IUPred2A. *Curr. Protoc. Bioinforma.* **70**, e99 (2020).

194. Erdős, G., Pajkos, M. & Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).

195. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

196. Chang, Y.-M., Chen, C. K.-M. & Hou, M.-H. Conformational changes in DNA upon ligand binding monitored by circular dichroism. *Int. J. Mol. Sci.* **13**, 3394–3413 (2012).

197. Chen, L. *et al.* The isolation of an RNA aptamer targeting to p53 protein with single amino acid mutation. *Proc. Natl. Acad. Sci.* **112**, 10002 LP – 10007 (2015).

198. Heo, K. *et al.* An aptamer-antibody complex (oligobody) as a novel delivery platform for targeted cancer therapies. *J. Control. Release* **229**, 1–9 (2016).

199. Ganson, N. J. *et al.* Pre-existing anti-polyethylene glycol antibody linked to first-exposure allergic reactions to pegnivacogin, a PEGylated RNA aptamer. *The Journal of allergy and clinical immunology* vol. 137 1610-1613.e7 (2016).

200. Hengel, S. R., Spies, M. A. & Spies, M. Small-Molecule Inhibitors Targeting DNA Repair and DNA Repair Deficiency in Research and Cancer Therapy. *Cell Chem. Biol.* **24**, 1101–1119 (2017).

201. Greenfield, N. J. Analysis of circular dichroism data. *Methods Enzymol.* **383**, 282–317 (2004).

202. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).

203. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular interaction

studies using microscale thermophoresis. *Assay Drug Dev. Technol.* **9**, 342–353 (2011).

204. Entzian, C. & Schubert, T. Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). *Methods* **97**, 27–34 (2016).

205. López-Méndez, B., Uebel, S., Lundgren, L. P. & Sedivy, A. Microscale Thermophoresis and additional effects measured in NanoTemper Monolith instruments. *Eur. Biophys. J.* **50**, 653–660 (2021).

206. Rainard, J. M., Pandarakalam, G. C. & McElroy, S. P. Using Microscale Thermophoresis to Characterize Hits from High-Throughput Screening: A European Lead Factory Perspective. *SLAS Discov.* **23**, 225–241 (2018).

207. Baaske, P., Wienken, C. J., Reineck, P., Duhr, S. & Braun, D. Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angew. Chem. Int. Ed. Engl.* **49**, 2238–2241 (2010).

208. Li, Q. Application of Fragment-Based Drug Discovery to Versatile Targets . *Frontiers in Molecular Biosciences* vol. 7 180 (2020).

209. Hann, M. M., Leach, A. R. & Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856–864 (2001).

210. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **8**, 876–877 (2003).

211. Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discov.* **12**, 5–7 (2013).

212. Singh, M., Tam, B. & Akabayov, B. NMR-Fragment Based Virtual Screening: A Brief Overview. *Molecules* **23**, (2018).

213. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science (80-.).* (1996) doi:10.1126/science.274.5292.1531.

214. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3**, 39–53 (2003).

215. Meyer, B. *et al.* Saturation transfer difference NMR spectroscopy for identifying ligand epitopes and binding specificities. *Ernst Schering Res. Found. Workshop* 149–167 (2004) doi:10.1007/978-3-662-05397-3_9.

216. Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discov.* **12**, 644 (2013).

217. Mayer, M. & Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **123**, 6108–6117 (2001).

218. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* **21**, 349–359 (2001).

219. Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov. Today* **14**, 1051–1057 (2009).

7. Appendix

7.1. Circular Dichroism (CD)

Circular dichroism spectroscopy is a technique based on the measurements of the difference in the absorption of left and right circularly polarized light in optically active (chiral) substances[201,202].

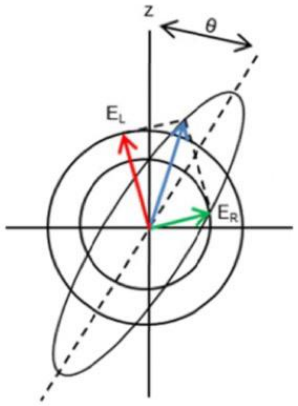


Fig. A1) Different absorption of the left- and right- hand polarized component leads to optical rotation and ellipticity (CD); figure taken from Jasco Learning Center (https://jascoinc.com/learning-center/theory/spectroscopy/).

Briefly, electromagnetic waves contain electric and magnetic field components that oscillate perpendicularly in the direction of a light beam's propagation. The directionality of these components defines the waves' polarization. In unpolarized light or white light, the electric and magnetic fields oscillate in many different directions. In linearly polarized light, the electromagnetic wave oscillates along a single plane, while in circularly polarized light two electromagnetic wave planes are at a 90° phase difference to one another and this plane rotates as the light beam propagates.

The linearly polarized light can be described as the superimposition of a right-handed and left-handed circular polarized light of equal intensities. When passing through an optically active medium, these two components of the polarized light acquire different velocities, according to their refractive index (this phenomenon is called birefringent) and are differently absorbed. In other words, as the circular polarized components of the linear polarized light travels through a chiral material, they change their oscillation and their resulting electric field vectors trace out an

ellipse, rather than a line. Circular Dichroism is based on the measurements of this absorption difference between left- and right- handed waves (E_L and E_R, respectively).

CD is reported either in units of ΔE, the difference in absorbance of left (E_L) and right (E_R) polarized lights by an asymmetric molecule, or in degrees ellipticity, which is defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse. Molar ellipticity [θ] is the most common used unit and it is expressed in deg. *cm^2/dmol.

CD is a method highly used for proteins and oligonucleotides to roughly estimate the secondary structure composition. CD is useful to characterize systems, which undergo structural changes after perturbation of their chemical environment or after the addition of a ligand or a drug.

The peptide bond is inherently asymmetric and it is always optically active. It is the principal absorption group in the far UV (260-190 nm), where its angles phi and psi rotate depending on the protein conformation. Two are the most significant absorption bands: one comes from the weak but broad component of n→ π* transition centered around 220 nm and the second one from the more intense π→ π* transition around 190 nm. Different types of protein secondary structures give rise to specific CD spectra[201,202].

Each of the secondary structure component (α-helical, β-sheet, turn, and unordered) has a peculiar CD spectrum (Fig. A2). Random coils present a minimum at 198 nm and a maximum at 212 nm; β-sheets present a minimum at 218 nm and a maximum at 196 nm; α-helices present two minima at 208 and 222 nm and a maximum at 192 nm. For unknown proteins, CD spectra analysis requires spectra deconvolution procedures based on a set of reference spectra with known secondary structure.

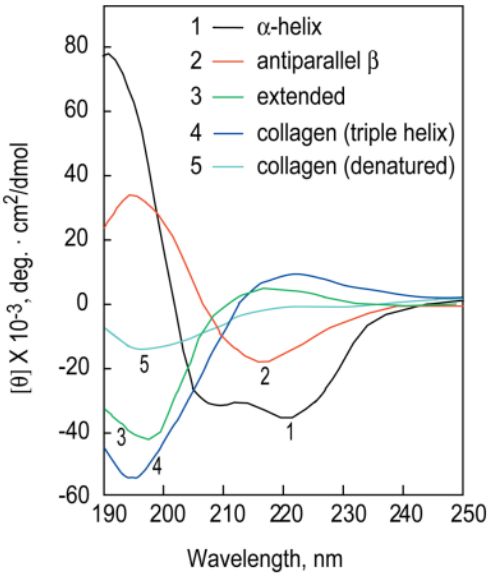


Fig. A2) Circular dichroism (CD) spectra of polypeptides and proteins with some representative secondary structures; peptides and proteins with specific CD spectra profile are used to predict secondary structure of unknown proteins. Image taken from Greenfield et al. (2006).

Similar analyses can be performed for DNA and RNA structures. Here, the chirality of sugar groups bound to DNA/RNA bases can induce a CD absorption band for the base pair $\pi \rightarrow \pi^*$ transition in 200-300 nm range. At variance with proteins, for nucleic acid structures not many reference examples are described and it is therefore more difficult to unequivocally associate CD profiles to specific secondary structures. However, some general features of DNA CD spectra have been described, so far. B-DNA, which is the most common DNA form found in cell large 2.4 nm and with a 10.5 base pairs per turn, shows a positive maximum at 290 nm, negative maximum around 245 nm, and a broad positive peak between 260-280 nm. The A-DNA, which is a more compact, right-handed double helix structure, exhibits a positive peak at 260 nm and a negative peak at 210 nm. The Z-form structure is a left-handed double helix that displays a negative band at 290 nm, a positive peak at 260 nm, and negative maximum ~200 nm[184].

7.2. Bio-Layer Interferometry (BLI)

BLI is a label-free technology that allows to obtain real-time rapid and sensitive measurements of affinities and binding kinetics of a variety of interactions, including protein-protein, protein-nucleic acids and protein-small molecules interactions.

BLI measures kinetics and biomolecular interactions exploiting the difference in two interference patterns between waves of light depending on mass immobilization on specific biosensors to measure kinetics and biomolecular interactions.

Specifically, BLI is based on the use of fiber-optic biosensors. White light is transmitted through biosensors toward a tip, which is formed by two layers, one biocompatible layer on the surface of the tip and one internal reference layer. White light is reflected from each of the two layers and the reflected beams can interfere constructively or destructively at different wavelengths in the spectrum. The interference pattern is measured by the charge couple device (CCD) array detector. When the target molecule is immobilized on the sensor surface, the thickness of the tip layer increases leading to a subsequent increase in the distance between the two reflective layers, creating a shift in the interference pattern of the reflected lights. This change is reported on a sensorgram as a change in wavelength (nm). The same principle of shift in the interference of reflected lights can be used to detect the number of analyte molecules associating and dissociating from the biosensors surface functionalized with target molecules, providing real-time kinetics data on molecular interactions (Fig. A3).

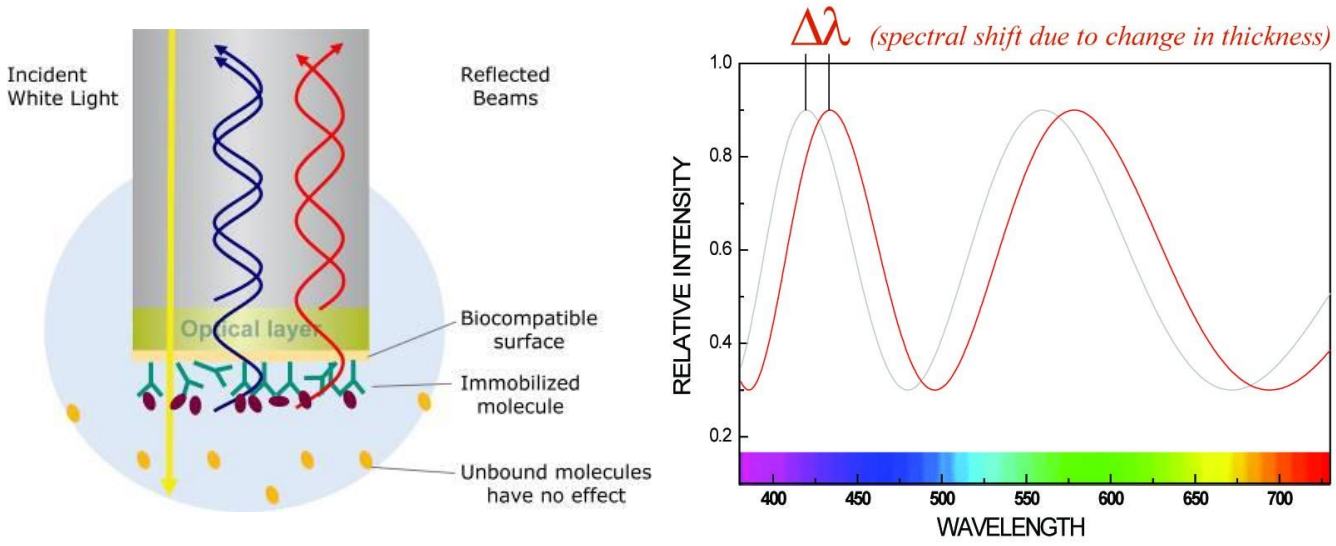


Fig. A3) Schematic representation of the biophyisical principle of Bio-layer Interferometry. An incident white light is reflected from two different layer on a biosensor tip and create constructive or destructive interactions; depending on the immobilized mass on the biolayer, a spectral shift of the reflected light can be reported and measured to determine kinetics properties of the analyte. Image adapted from Pall Fortè Bio Training Slides.

A typical binding kinetics experiment is reported in Fig. A4.

A classical BLI assay usually starts with an equilibration step of the biosensors to record a reference baseline. Biosensors are then immersed in the solution containing the target molecule of interest to immobilize it on the biosensors surfaces using different immobilization methods (i.e. capture-based or direct immobilization). After an additional washing step in buffer to remove all the unspecific binders and to assess assay drift, the biosensors are immerged in the solutions containing the analyte ligand, whose binding needs to be investigated (association step). Finally, the dissociation step is performed dipping the biosensors in a solution containing only buffer.

Real-time monitoring of the above described steps allows the extrapolation of all the kinetics data, such as k_a and k_d, which could not be detectable in end point assays.

The rate of association is a function of the decreasing concentration of unbound ligand molecules as analyte binding occurs and it is described by the following equation:

$$Y=Y_0+A(1-e^{-k_{obs}*t})$$

Where Y is the level of binding and Y_0 is the binding level before the association step; A is the slope and t is the time; k_{obs} is the observed rate constant, reflecting the overall rate of association and dissociation of the ligand to the immobilized molecule.

The dissociation step is described by the following equation:

$$Y=Y_0+Ae^{-k_d*t}$$

Where the k_d is the dissociation rate constant. It is the direct measurement of the stability of the complex and is related to the fraction of the complex that dissociates per second.

From the association (k_{obs}) and dissociation (k_d) equations, the association rate constant k_a can be calculated:

$$k_a = \frac{k_{obs} - k_d}{[\text{Analyte}]}$$

Finally, the affinity constant (K_D) can be calculated through the ratio between k_d and k_a.

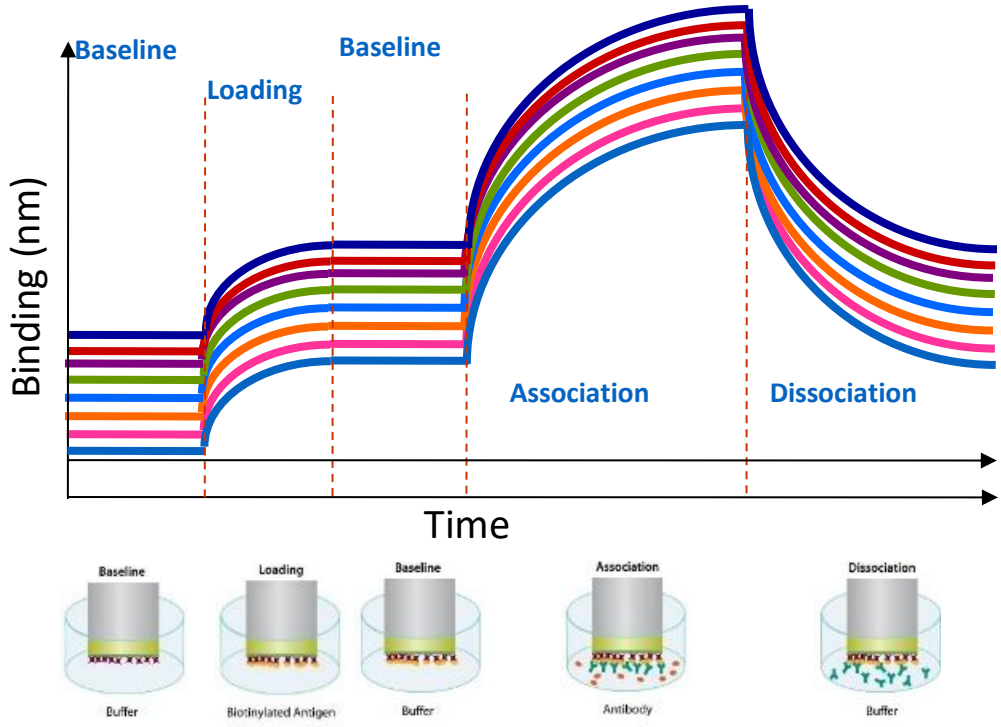


Fig. A4) Schematic representation of an example of BLI sensorgram where binding level (in nm) is expressed as function of time (s); changes in the sensorgram traces are correlated to different immobilization level of target (loading phase) and analyte (association); BLI sensorgram can also follow the dissociation phase, where the mass on the biosensor decreases proportionally to analyte K_d. Image adapted from Pall Fortè Bio Training Slides.

7.3. MicroScale Thermophoresis (MST)

MST is a powerful technique to quantify biomolecular interactions (Fig. A5). It is based on "thermophoresis", the directed movement of a labelled molecules in a temperature gradient, which strongly depends on size, charge, hydration shell or conformation[203]. MST have the advantage to be performed in solution, without any immobilization step required, using a very low nanomolar concentrations of the target molecule. Kinetic analyses can be performed for binding affinities (K_d) ranging from pM to mM range.

To perform MST analysis, 16 capillaries (volume ~4-10 µL) are simultaneously filled with solution containing a constant concentration of labelled protein and increasing concentrations of a potential ligand. After that, using an IR laser, a thermal gradient is generated in each of the capillaries containing target protein or target protein-ligand solution (Fig. A5). Upon this IR laser-induced ΔT, the actual thermophoresis, described by the thermal diffusion coefficient D_T, begins, balancing the mass diffusion (described by D coefficient). The sample behavior in this dynamic equilibrium in each capillary can be described by the diffusion coefficient ratio, also known as Soret coefficient (S_T): $S_T = D_T/D$.

The local change in temperature induces a migration of the observed species and the movement can be described by the following equation:

$$\frac{c_{hot}}{c_{cold}} = e^{-ST\Delta T}$$

Where C_{hot} represents the concentration of the labelled protein in the heated region, and C_{cold} is the concentration in the initial cold region[203–205].

S_T is very sensitive to small changes at the molecular interface and it is expected to depend on surface area, surface entropy and net charge. Whereas the binding of a ligand to the labelled molecule alters at least one of these parameters, a change in thermophoretic mobility is reported and the binding can be quantified by measuring the change in thermophoresis at different ligand concentrations[204,206].

Specifically, for quantitatively analyze the binding profile of the potential binding interactor, the change in MST signal is expressed as the change in the normalized fluorescence (ΔF_{norm}), where

each data point is defined as $F_{norm} = F_1/F_0$, where F_1 is the fluorescence of the sample after a given MST-laser *on time* and F_0 the fluorescence of the sample before the IR laser activation.

Ligand-dependent changes in MST are plotted as F_{norm} values against ligand concentration to obtain a dose-response binding curves, from which the binding K_d can be extrapolated.

Notably, in recent publications, MST have been generalized with the term of Temperature-Related Intensity Change (TRIC), since it is known that the MicroScale Thermophoresis phenomenon is only one of the many events that may affect fluorescence signal changes[205]. For instance, fluorophores show temperature dependent variation of their fluorescence intensity, depending on their chemical environment[207]. Binding events, which alter molecular conformations, physicochemical properties or protein dynamics can be directly detected based on changes in the temperature dependency of proximate fluorophores.

MST is a powerful technique which we used either to screen different target compounds with our target proteins or to obtain preliminary kinetics information, exploiting its fast set-up and low amount of sample required.

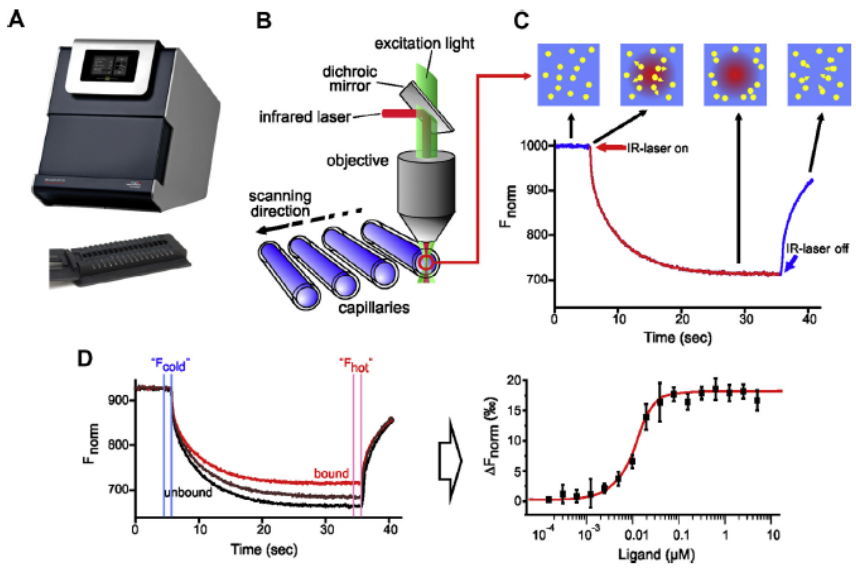


Fig. A5) Schematic representation of MST setup and experiments. A) Monolith NT.115 instrument (NanoTemper Technologies Gmbh); B) Representation of one MST experiment; C) Representation of a signal from a MST experiment: when laser is on, a rapid change of in fluorophore properties is measured. The thermophoresis of the molecules is detected in the temperature gradient. ; D) Example of a binding experiment where differences in the thermophoretic traces of molecules in bound and unbound state are present; the fluorescence traces are normalized and are reported as function of ligand concentration to calculate the Kd value; Image taken from Jerabek-Willemsen et al (2013).

7.4. Fragment Based Drug Discovery (FBDD)

FBDD is a fragment-based approach (FBA) to develop potent compounds from fragments. This concept arises from the idea that a potential inhibitor can be considered as the combination of fragments that bind different parts of a target protein. The FBDD can start from chemical fragments with low binding affinities to the target, low complexity in chemical structure and low molecular weight and generate a more complex and active compound, whose affinity and effect can be derived from the sum of its single parts[208].

FBDD approach facilitates the investigation of the chemical space of protein binding sites using small fragments instead of using big drug-like molecules. Additionally, it is important to mention

that the use of fragments in drug discovery, instead of big complex chemical entities was proved to increase the probability to discover hits compounds, due to the better fit in the different binding pockets of the target protein[209]. Moreover, it is easier to cover a larger chemical space with the screening of small molecules libraries respect to big drug-like compounds libraries.

There is not a unique term for "fragment". It is suggested that fragments should follow the rules-of-three in which compounds have a molecular weight less than 300 Da, ClogP value less than three, and less than three hydrogen bond donors and acceptors[210].

Notably, some fragment hits do not adhere to these restrictions indicating that this rule can be applied only as a guideline.

According to this definition the molecular weight of fragments compounds should be relatively small, which gives rise to high ligand efficiency and provides more opportunities for growing the hits[211,212]. Moreover, always thanks to their low size, it is possible to check fragment properties in buffer solution before testing them in proper biophysical assays[176].

Nevertheless, due to their chemical simple scaffolds and reduced functionality, active compounds (hits), coming from FBA, usually show weak affinity for the target protein in the micro- milli-molar range. For this reason, fragments usually need to be screened at high concentration; sensitive methods and structural information are required for determination of active fragments (binders) to start their development gaining a higher potency. The most common approaches used for FBA measurements are NMR, X-ray crystallography, SPR and fluorescence spectroscopy. Each approach presents both advantage and disadvantages[208].

In this research work the FBDD was investigated by NMR.

7.5. NMR fragment Based Approach (FBA)

NMR has become the most popular technique for application in FBDD, as it can detect weak bindings (K_d in the low mM-range) between fragments and target macromolecules[213,214]. This technique allows to measure binding effects in solution, without any target labelling ,molecules modifications or immobilization, leaving the system unaltered [215]. Moreover, the high sensitivity of NMR allows the identification of a hit also when only 5% of a fragment interacts with the target, therefore the detection of hits with solubilities lower than their potencies is possible[216].

Two are the approaches more commonly used in NMR screening: target-based and ligand-based[212]. The target-based method, even if able to provide precise structural information about ligand binding epitope on the target, requires a very complex time-consuming experimental set up with a significant amount of double (^{15}N, ^{13}C) and triple (^{1}D, ^{15}N, ^{13}C) labelled proteins if previous NMR spectra assignments are not available in NMR data banks. Moreover, this approach has limitation linked to protein size.

On the other end, the ligand-based approach can be applied to a broader range of unlabeled targets at low concentration. This method shows changes in the NMR spectrum of the observed ligand, when it interacts with the target protein. The most used ligand-observed NMR experiments are: Saturation Transfer Difference (STD), Water-Ligand Observed via Gradient SpectroscopY (WaterLOGSY), ^{1}H transverse relaxation time (T2p) and ^{19}F NMR-based experiments (Fig. A6).

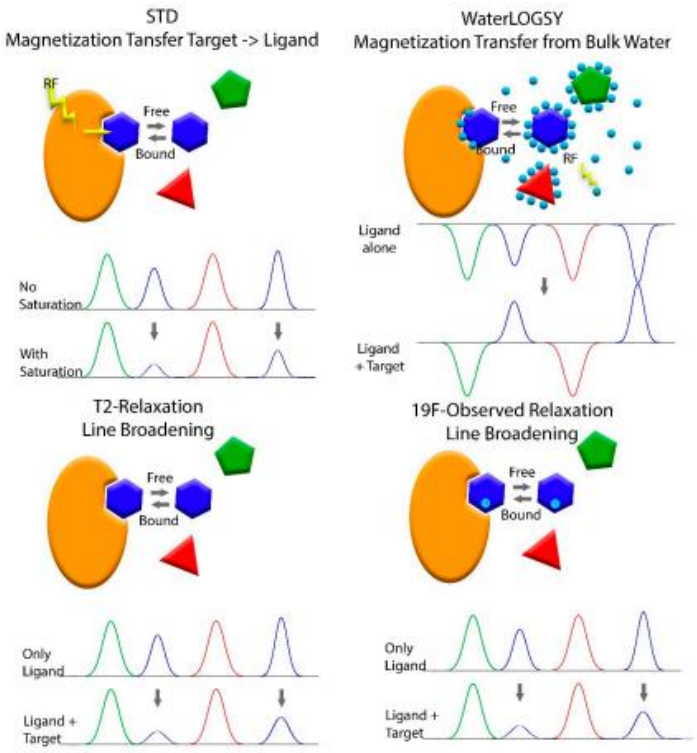


Fig. A6) Schematic representation of different ligand-based methods used in NMR fragment-based screening; image reports examples of NMR spectra obtained for each method. Image taken from Sing et al. (2017).

One of the most widely used ligand-based methods is the saturation transfer difference (STD) experiment[217]. This experiment uses a series of selective pulses to saturate signals of the protein that are in an area of the spectrum with no signals originating from the fragments (e.g., methyl protons of the proteins below 0 ppm). The saturation is transferred throughout the protein and to any bound fragment, causing a decrease in signal intensity. The comparison between 1D spectra of the fragments recorded with and without the saturation pulses allows the identification of fragments that bind to the protein.

In waterLOGSY experiments, water molecules, and not the protein, are selectively excited [164,218]. The proton (^{1}H) magnetization from excited water molecules is then transferred to the observed ligand either via direct interaction or indirectly through an initial transfer to protons located on the protein surface and then onto the protein-bound ligand. The direct interaction of water with free ligand leads to an increase of the observed ligand signal due to a direct magnetic interaction between ligand protons and water, resulting in a positive Nuclear Overhausen Effect (NOE) (conventionally, signals of this free unbound state of the ligands are displayed as negative in the waterLOGSY spectrum). The indirect magnetization transfer of a bound ligand through protein magnetization leads, on the contrary, to a decrease of the ligand signal correlated with a negative NOE (signals are conventionally displayed as positive in the waterLOGSY spectra). The differences in magnetization transfer behavior when ligand is in the bound and non-bound states are related to the tumbling rates of molecules in solutions, which are fast for small unbound ligands and slow for macromolecules (i.e. proteins) and their bound ligands. Therefore, the waterLOGSY effect on a ligand depends on these two NOE effects. Since the ligand is not in high excess compared to protein concentration and the NOE negative effect is much larger in comparison with the positive NOE, the signals in the waterLOGSY spectrum of a ligand in the presence of its protein target depend much on the NOE negative effect and are positive. For this reason, binder molecules can be easily discriminated from non-binder molecules.

This method can also be applied to evaluate the aggregation state of ligand molecules.

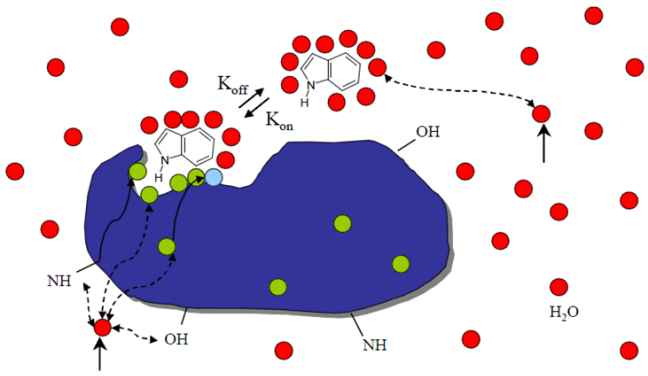


Fig. A7) Representation of a binding condition measured in waterLOGSY experiment; the protein is shown with the buried cavities and the active binding site. Ligand is shown in the bound and free states. Excitation of bulk water (circles) is shown with a solid arrow and some of different magnetization transfer pathways are shown with dashed lines. Figure taken from Dalvit et al. (2001)

Another highly used NMR experiment used in FBA is the T2 relaxation line broadening. Specifically, a ligand bound to a much larger molecule, such as a its protein target, will adopt the relaxation properties of the large target in its bound state, moving with a much slower tumbling and a faster relaxation time. The Carr Purcell Maiboom Gill (CPMG) is a NMR pulse sequence that can be used to measure the relaxation time T2 of the fragments in their unbound form and in a sample containing the target biomolecule. Fragments that bind to the target have reduced relaxation times. In a NMR spectrum, the relaxation time is indirectly proportional to the line width at half height of the NMR signal. Therefore, the shorter relaxation times of ligands bound to large targets lead to a line broadening. Since the ligand peak consists of a large contribution from the unbound state with a sharp narrow line and a smaller contribution from the bound state with a broad line, the main effect observed in the 1D NMR spectrum is a reduction in intensity of the ligand signal.

The T2 relaxation line broadening can be pursued also using ^{19}F labelled ligands. Measuring the ^{19}F relaxation rates of the ligand in the presence or absence of the biomolecule will allow the identification of those ligands that bind to the target by observing a shorter relaxation time, i.e. line broadening, reduction in intensity of ligand signals as for the proton experiments. Notably, ^{19}F nucleus presents additional advantages: 100% natural abundance, high sensitivity, favorable transverse relaxation and large dispersion of chemical shifts that allows the screening of large chemical mixtures avoiding problems of signals overlapping in highly crowded spectra[219].

In my PhD project, ^{19}F NMR spectroscopy was used as screening technique to identify novel fragments to exploit as starting point for RAD52 inhibitors development. Given the high sensitivity and large chemical shift window of fluorine signals, we were able to screen fluorinated mix of fragments with an easy detection of binding /non-binding state on the target.

GSK-3β Project

In the following section, I present a project that I focused on during the first period of my PhD. As it is a secondary project, I report it in a preliminary manuscript-like form. The related article has been submitted and accepted to the "International Journal of Molecular Sciences".

8. *Identification of novel GSK-3β hits using competitive biophysical assays*

Abstract

GSK-3β is an evolutionarily conserved serine-threonine kinase dysregulated in many pathologies such as diabetes II, Alzheimer's disease and cancer. This protein is a validated pharmacological target and several inhibitors have been already reported. Even though, some of them have been, or are currently being, tested in pre-clinical and clinical studies, most GSK-3β inhibitors have not reached clinical phase due to a variety of encountered issues including toxicity and safety. Nevertheless, one of the main issues we are still facing in this field is the identification of highly selective GSK-3β inhibitors.

In this work, we describe the identification and characterization, through structural and biophysical approaches, of two new promising GSK-3β fragment inhibitors, identified through a novel drug discovery workflow.

Recombinant GSK-3β was expressed in *High Five* cells and purified in a two-step purification process (affinity and cation exchange chromatography). A library of fragments was screened against the purified recombinant protein using a 1D ^{19}F NMR ligand-based screening approach in the presence or absence of saturating AMP-PNP (non-hydrolysable ATP analogue) concentrations. This approach aimed at the identification of small molecules capable of adapting to specific protein hotspots other than the ATP binding pocket (allosteric binders) or to the ATP binding pocket, but with an affinity able of competing with AMP-PNP. The advantage of this approach is that small molecules can better fit a pocket, therefore they should in principle be more specific, and competition with a known binder allows the selection of the most potent hit fragments directly during the initial screening campaign.

The identified potential binders of GSK-3β were validated and further selected through *in vitro* biophysical and activity assays (MicroScale Thermophoresis, Protein Thermal Shift, Time Resolved – Fluorescence Resonance Energy Transfer), based on kinetic and thermodynamic characterizations of their bindings and inhibitory actions on GSK-3β. Out of this screening, two fragments (G12 and G5) were selected. The X-ray 3D structures of these two compounds in complex with GSK-3β were solved and they allowed the identification of both compounds bindings in the ATP pocket, thus assessing their ATP-competitive behavior. The ATP competition was confirmed growing co-crystals of GSK-3β also in the presence of AMP-PNP increasing

concentrations. According to the observed electron densities, the two compounds were progressively displaced from their binding site in an AMP-PNP concentration-dependent fashion. A customized phosphorylation activity assay on a kinases panel highlighted G12 and G5 selectivity on GSK-3β compared to other kinases. The selective behavior of the two compounds was explained by superimposing the G12 - GSK-3β, G5 - GSK-3β 3D structures with available structures of other kinases. Through a computational analysis of our internal database, G12 and G5 compound analogues were identified and tested for their bindings and activities on the protein. Out of seven molecules, two compounds (ARN1484 and ARN9133) showed a similar inhibitory activity compared to G12 and G5. Also the 3D structures of these compounds in complex with GSK-3β were solved.

In conclusion, with our study we identified promising fragments, inhibitors of GSK-3β, worth to be further developed into more potent lead compounds. In addition, we propose a novel drug discovery screening pipeline, which, can be generalized to the search of potent and selective inhibitors for any ATP dependent enzyme.

Preliminary manuscript: identification of novel GSK3β hits using competitive biophysical assays

Introduction

GSK-3β is a serine-threonine kinase known first for its ability to regulate glycogen synthesis. Nowadays, it has been reported to play a critical role in many other intracellular pathways, giving it a widespread importance. A part from glycogen synthesis, GSK-3β is a component of the signaling pathways involving for instance insulin, growth factors and neurothrophins (Wnt/β-catenin pathway)[1]. Moreover, it plays a critical role in regulation of transcriptional factors able to control the expression of different genes[2] . Indeed, this widespread activity has led to identifying GSK-3β dysregulation as the cause of many diseases developments, such as Alzheimer's disease, bipolar disorder, diabetes, cardiovascular diseases and cancer[3,4].

GSK-3β has a typical structural kinase fold (Fig. 1): a N-terminal β-strand domain (residues 25-138), consisting of seven antiparallel β-strands, and a C-terminal α-helical domain (residues 136-343). The ATP binding site is located at the interface between the two domains (hinge region) and is constituted by a glycine-rich loop (G-X-G-X-X-G), which interacts with the phosphate atoms of ATP. The activation loop (residues 200-226) is well ordered and runs along the surface of the substrate binding groove, and the C-terminal residues (344-382) are located outside the core kinase fold[5].

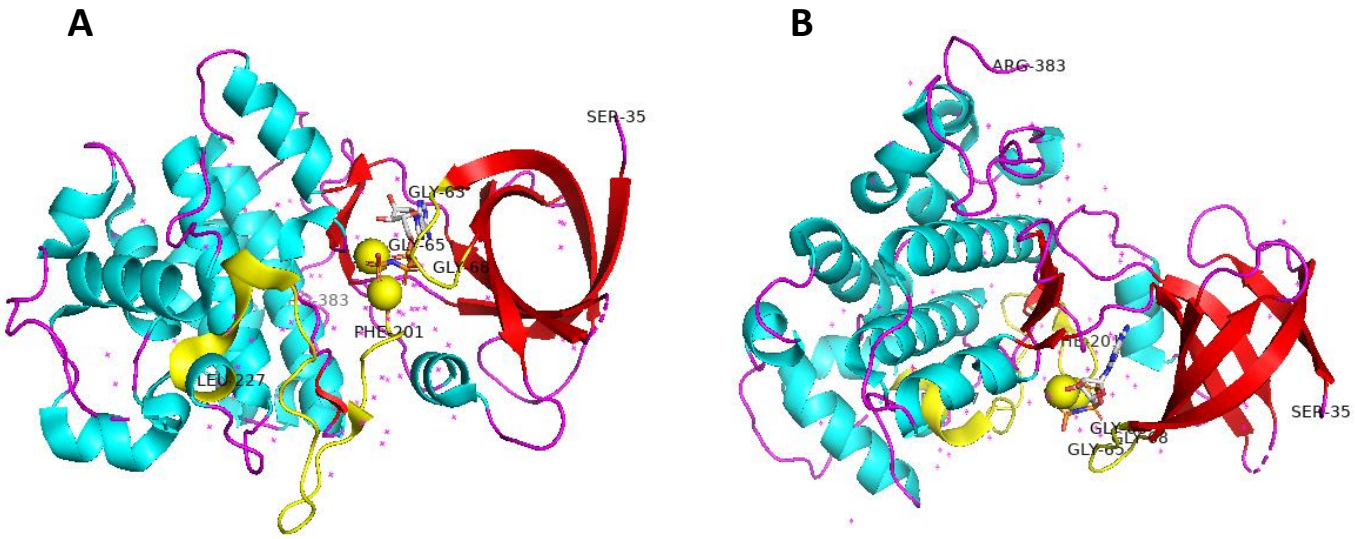


Fig. 1) A) Structure of GSK-3β from residue 35 to the visible C-terminus reside 384. N-terminal β-strands are reported in red, C-terminal α-helices are reported in blue. Glycine-rich loop and activation loop are reported in yellow. B) View is rotated by 90 ° around the horizontal axis. Images were prepared using PyMOL 2.4 software (PDB 1PYX).

Although GSK-3β has structural features similar to many other kinases, it has distinctive characterizing features related to its activity. First, this kinase can only act on a pre-phosphorylated substrate sequence S/T-X-X-X-S/T(P), which induces a conformational change that aids the positioning of the substrate in the active catalytic domain of the kinase[5,6].

Secondly, this kinase is constitutively active, being inhibited only in response to a specific stimulus[7]. The phosphorylation on Ser9, induced by extracellular signals, negatively regulate GSK-3β kinase activity causing the self-association of the GSK-3β N-terminal tail to its substrate binding pocket therefore hampering the interaction with the substrates. This inhibitory GSK-3β phosphorylation is required to maintain normal cell homeostasis[8].

On the other hand, phosphorylation of GSK-3β on Tyr216 induces a conformational change that allows the interaction with the substrate and facilitate protein phosphorylation activity [9,10].

The central role of GSK-3β in many intracellular pathways and the high number of its substrates (i.e. 100+[3,11]) makes this enzyme a very interesting and puzzling target for drug discovery pipeline. In fact, in these years a plethora of GSK-3β inhibitors have been developed for many diseases treatments[12–16] .

Promising therapies with GSK-3β inhibitors are already available and they have been widely studied for bipolar disorder[17,18], Alzheimer's disease[19], inflammatory diseases[20], AIDS[21], diabetes[22,23] and cancer[24]. Furthermore, according to the literature data available so far, some GSK-3β inhibitors are being studied in clinical trials for the therapeutic treatment of AD[25] and Mild Cognitive Impairment[26].

Maleimide derivatives for example are reported to be ATP competitive GSK-3β specific inhibitors[27]. SB-216763 is a maleimide derivative with IC_{50} of 34 nM and it is highly selective for GSK-3β[28]. Indole derivatives are another class of GSK-3β inhibitors, which, on the contrary, has high affinity (50-100 nM) also for the closely related CDK5/p25[29]. Lastly, also pyrazine derivatives have very high affinity (<1 nM)[30] for GSK-3β. Many other ATP competitive and non-ATP competitive are nicely summarized by Pandey et al.[31].

The tridimensional structures of several GSK-3β-inhibitors complexes are available in the Protein Data Bank (PDB). An inspection of these structures shows that most ATP competitive inhibitors commonly share hydrogen-bonding interactions with the backbone atoms of the hinge region residues ASP133-Pro136. Maleimide derivatives, which a part from the high affinity also show a high selectivity for GSK-3β, are additionally involved in either direct or water mediates hydrogen bonding interactions with Lys85, Glu97 and Asp200, besides the interactions with the hinge region residues Arg144 and Glu137 of the protein. The interactions with these last five residues are proposed to be responsible for the selectivity of Maleimide analogues towards GSK-3β[30,32].

Up to now, many are the GSK-3β inhibitors described, but a very slow progress of these compounds to the clinic is reported, due to adverse effects caused by cellular off-targets[31]. For these reason, novel strategies to design selective and potent GSK-3β inhibitors are a medical need. In this perspective of identification of innovative inhibitors, we designed and validated a drug discovery pipeline, which allowed to search for allosteric inhibitors or to screen and simultaneously select chemical fragments with a significant affinity for GSK-3β. Specifically, fragments, given their small size, fit in the protein hotspots much better compared to bigger drug-like molecules, but they usually display very low affinities[33,34]. For this reason, we focused on fragments that given their small size are able to enter into specific protein pockets and, at the same time, we overcame the issue of their low affinity using a competitor compound to set a cut-off to exclude too weak fragment binders.

In this work, we identified and characterized four novel GSK-3β inhibitor molecules. Furthermore, we proposed and validated a pipeline for the development of novel kinase inhibitors. This approach started with a monodimensional ^{19}F NMR fragments library screening[33,35–39] in the presence of saturating concentration of non-hydrolysable ATP analogue, AMP-PNP. The goal was the identification of chemical fragments able of binding the target in pockets others than that of the ATP or at least with an affinity higher than the one of AMP-PNP[40]. Selected compounds were then characterized and their inhibitory activity was evaluated through biochemical and biophysical assays. Moreover, X-ray crystal structures of the two identified potential hits (G5 and G12) and two analogues (ARN9133 and ARN1484) in complex with GSK-3β were solved. The results helped us understanding the interaction properties of these compounds within the protein and allowed us to clarify the mechanism of compounds interaction and inhibitory activity.

Materials and Methods

Expression and Purification of recombinant GSK-3β

Expression step of GSK-3β protein was performed by prof. P. Storici and dr. B. Giabbai (Elettra Sincrotrone Trieste). The DNA sequence of human GSK-3β full length (1-420) was cloned in pFB-LIC-Bse vector (kindly provided by dr. Opher Gileadi, SGC-Oxford). The resulting transfer vector was sequence verified and transformed into E.*coli* DH10Bac cells to obtain the recombinant bacmid-DNA. Sf9 cells (Expression Systems LLC, Davies - USA) were seeded in a six-well plate at $1.5x10^6$ cells/well in ESF-921™ medium (Expression systems) and transfection was performed using FuGENE® HT reagent (Promega). Plate was incubated at 27°C and recombinant baculovirus was harvested 55 hours post-transfection (P0 stock). The high titer virus stock (P2) was generated by two rounds of amplification and used for the His-tagged GSK-3β expression. *High Five* (H5) cells (Expression Systems LLC, Davies - USA) were infected with P2 stock at an initial density of $1.5x10^6$ cells/ml in ESF-921™ medium, incubated at 27°C and harvested by centrifugation 72 hours after infection. Cell pellets of $900x10^6$ cells were resuspeded and thawed in lysis buffer (20 mM TRIS pH 8, 0.5 M NaCl, 10 mM Imidazole, 1 mM DTT, 5 mM MgCl2, 0.5x protease inhibitor

EDTA free *(Roche),* 5% glycerol, 0.01% Tween20) and lysed with sonicator (12' pulse at 60-70% intensity). After sonication, lysate solution was incubated for 20 minutes at 4°C with DNAse I 5 ug/mL (*Sigma*), then centrifuged for 1h at 4°C at 30000g. After centrifugation, supernatant solution containing protein of interest underwent a two-step-purification procedure. First, the clarified supernatant was incubated for 2h with Ni-NTA agarose resin (*Qiagen*) and washed out with a binding buffer containing 10 mM Imidazole (20 mM TRIS pH 8, 0.5 M NaCl, 10 mM Imidazole, 5% glycerol, 1 mM DTT). Elution was performed with buffer containing 0.3 M imidazole (20 mM TRIS pH 8, 0.5 M NaCl, 0.3 M Imidazole, 5% glycerol, 1 mM DTT). Eluted proteins were then diluted in 20 mM Hepes and 1 mM DTT only, to reach the correct salt concentration of ~40 mM NaCl in order to load it on a cationic exchange column *HiTrap HP SP (Cytiva)* (loading buffer: 20 mM Hepes pH 7.5, 30 mM NaCl, 5% glycerol, 1 mM DTT). At this point, the protein was purified and cleaned from solution impurities with a step gradient elution (elution buffer: 20 mM Hepes pH 7.5, 1 M NaCl, 5% glycerol, 1 mM DTT), first with 8% elution buffer (washing step), then with 13%, and, finally, increasing gradually elution buffer concentration up to 100%, allowing all GSK-3β isoforms to elute. Only the first peak of the chromatography eluate, at 100-130 mM NaCl, corresponding to phosphorylated (pTyr216) and most active GSK-3β isoform, was used for further experiments[6]. Obtained protein aliquots were collected and stored at -80°C.

^{19}F NMR ligand-based binding screening R_2 filter experiments ($T_2\rho$)

NMR experiments were performed by dr. M. Veronesi (IIT Genova). All NMR screening experiments were recorded at 298 K with a Bruker FT NMR Avance III 600 MHz spectrometer, equipped with a 5 mm CryoProbe™ QCI ^{1}H/^{19}F–^{13}C/^{15}N–D with an automatic sample changer SampleJet™ with temperature control. All GSK-3β experiments were performed at a low enzyme concentration (750 nM/1 µM) in 60 mM Hepes pH 7.5, 25 mM NaCl, 10 mM $MgCl_2$, 2 mM TCEP, 8% D2O (for lock signal), 0.003% Triton X100 (for the coating of NMR tube wall). About 350 fluorinated fragments, belonging to the internal LEF library were screened at 40 µM in mixtures of 25 compounds each one in the presence of 1mM AMP-PNP (Sigma Cat. Numb. A2647) and in the absence (control) or presence of 1 µM GSK-3β.

Promising hits were tested as single compounds in the presence of two non-binder (negative controls) in the same experimental screening conditions to confirm the binding observed in mixture. Bindings of the confirmed hits were further analyzed by testing the compounds at 20 µM in absence or in the presence of 750 nM GSK-3β and in absence or in the presence of saturating concentrations of AMP-PNP (1 mM and 3 mM). For each sample a 1D ^{19}F with ^{1}H experiment and ^{19}F R_2 filter experiments were recorded with the Carr-Purcell-Meiboom-Gill (cpmg) scheme[41,42] with a time interval of 23.5 ms between the 180° pulses and different total length (94, 188, 282 and 376 ms respectively). All the NMR experiments were run with proton decoupling using the Walts 16 composite pulse sequence with a 90° pulse of 120 ms, with a spectral width of 50 ppm, an acquisition time of 0.58 s, a relaxation delay of 5 s and a number of

scans of 128. The spectra were transformed using a line broadening of 1 Hz before the Fourier transformation. All the fluorine chemical shifts were referred to the $CFCl_3$ signal in water.

Microscale Thermophoresis binding detection

MST experiments were conducted in triplicate on a Monolith NT.115 Pico system (NanoTemper Technologies).

MST technique was used both for hit binding validation and for analyses and definition of k_ds (where possible). In particular, different labelling procedures and dyes were tested and selected for optimizing the output of the performed experiments. Binding and affinity of compounds of interest were tested in different buffers and labelling conditions as direct binding or displacement binding experiments (in the presence of AMP-PNP saturating condition).

GSK-3β was labelled with the His-tag RED-tris-NTA label 2nd generation (Nanotemper Technologies) following Nanotemper technology protocol. Briefly, protein labelling was performed in buffer NaCl 150 mM, $MgCl_2$ 5 mM, Hepes pH 7.5 20 mM, PEG8000 0.1%, Tween80 0.001%. Protein and dye solutions were mixed in a 2:1 ratio (200 nM and 100 nM respectively) and incubated for 30 minutes at room temperature. The resulted labelled protein was separated from the free-dye through 10 minutes centrifugation at 15000g at 4°C. The 100 nM labelled protein solution was then used at a concentration of 20 nM to reach around 7000 fluorescence counts in each screening experiment.

GSK-3β was labelled also using RED-NHS ammine dye (Nanotemper Technologies) following manufacturer's protocol. Briefly, fresh prepared dye RED-NHS ammine dye (470 µM stock in DMSO) was mixed in a 3:1 ratio with protein GSK-3β in a 200 µL final volume. The solution was incubated for 30 minutes at RT in the dark. To remove the "unreacted" free dye from the solution of the labelled protein an exchange buffer column provided by Nanotemper labelling kit was used. Protein aliquots could be stored at -80°C. Obtained ammine-labelled protein was used at 20 nM concentration to reach around 6000 fluorescence counts.

Compounds bindings were tested using various concentration of compounds. The highest concentration reached of compounds for binding check analyses (with or without AMP-PNP saturating conditions) was between 200 and 400 µM, depending on solubility features of the compound itself. Specifically, for binding checks, MST measurements were performed using 8 capillaries with a constant concentration of labelled protein, detecting the effect of the compounds on the fluorescence signal (with or without AMP-PNP saturation 3 mM). Protein solution treated with DMSO only represented the negative control. For binding affinity assays, MST measurements were performed using 16 capillaries with constant concentration of labelled protein and 16 different concentration of compound in absence or presence of saturating condition of AMP-PNP (3 mM).

Experiments of binding check were performed testing different MST assay buffers and labelling dyes. In particular, first screenings were performed using 20 nM His-tag labelled protein. Afterwards, to confirm binding and to better define k_d of promising compounds, ammine labelling method was preferred. These last analyses were performed with the following assay buffer: Phosphate buffer 0.1 M pH 7.2, 5 mM MgCl2, 0.001% Tween80, 0.1% PEG8000. Experiments were performed in triplicates.

Surface Plasmon Resonance (SPR)

SPR measurements were carried out on a Pioneer FE (Pall ForteBio) at 25 °C. GSK-3β was immobilized on a HisCap SensorChip (Pall ForteBio) and then left to equilibrate in immobilization buffer (20 mM Hepes pH 7.5, 150 mM NaCl, 0.01% Tween20). To improve the chip surface stability, an approach based on the capture-coupling protocol proposed by Rich and colleagues was used[43]. The SensorChip surface was activated with 500 µM $NiCl_2$ in immobilization buffer at a flow rate of 10 µL/min. After the activation of the surface, GSK-3β diluted in immobilization buffer at 7.5 µg/mL was immobilized by injecting 100 µL at a flow rate of 10 µL/ min. The surface was stabilized by amine coupling with the injection of 7.5 µL of 0.2 M NHS/0.4 M EDC diluted 1:10 in H2O at a flow rate of 15 µL/min, followed by an injection of 35 µL of 1.0 M ethanolamine pH 8.0 at a flow rate of 5 µL/min to block unreacted groups. The chip surface was cleaned of Ni^{2+} ions by injection of 60 µL of EDTA in immobilization buffer at a flow rate of 20 µL/min. GSK-3β was alternatively immobilized on flow cell (FC) 1 or FC 3, with FC 2 used as reference. Typical immobilization levels ranged from 3500 to 4500 RU. Binding experiments were performed in binding buffer (50 mM Tris−HCl pH 7.5, 250 mM NaCl, 0.01% Tween20) supplemented with 5% DMSO. Tested compounds were solubilized in 100% DMSO and then diluted in binding buffer by serial doubling. In binding assays, a flow rate of 40 µL/min was set up. Association was measured for 5 minutes and dissociation for up to 20 minutes. Data were then elaborated using Pioneer Qdat Software (Pall ForteBio) using a 1:1 stochiometric ratio model.

Activity assay of GSK-3β for inhibitory activity of selected compounds, ATP competition, substrate competition (Time-Resolved Fluorescence Resonance Energy)

Activity assays were performed in collaboration with dr. D. Russo and dr. I. Penna (IIT Genova). GSK-3β kinase assay was run in 384 well microplates (OptiPlate™-384, White, Perkin Elmer) in a total reaction volume of 20 µL. The inhibitory potency against human recombinant GSK-3β was evaluated using the LANCE® Ultra (Perkin Elmer) time-resolved fluorescence resonance energy transfer (TR-FRET) by measuring the phosphorylation of the specific substrate human Muscle Glycogen Synthase (ULight-GS (Ser641/pSer657)), according to the manufacturer's instructions. For the screening and dose-response curves, test compounds, staurosporine (reference compound) or DMSO (control) were mixed with the enzyme (2 nM) in a buffer containing 50 mM Hepes (pH 7.5), 1 mM EGTA, 10 mM $MgCl_2$, 2 mM DTT and 0.01% Tween20. The reaction was

initiated by adding 50 nM of the substrate ULight-PASVPPSPSLSRHSSPHQ(pS)ED and 1 µM ATP. The mixture was incubated for 1 hour at 23°C. Afterwards, the reaction was stopped by adding 8 mM EDTA. After 5 min, the anti-phospho-GS antibody labeled with europium chelate was added. 1 hour later, the kinase reaction was monitored by irradiation at 320 nm, and the fluorescence measured at 615 and 665 nm, using EnVision 2014 Multilabel Reader (PerkinElmer). The calculated signal ratio at 665/615 nm was proportional to the extent of ULight-GS phosphorylation. The compounds were screened at 4 concentrations (1 – 5 – 50 – 100 µM). For selected compounds, dose – response curves, ranging from 30 nM up to 200 µM, were performed. Dose-response curves were run in three independent experiments, each performed in three technical replicates. IC_{50} values (concentrations causing half-maximal response or enzyme inhibition) were determined by non-linear regression analysis of the Log [concentration]/response curves generated with mean replicate values using a four parameter Hill equation curve fitting with GraphPad Prism 8 (GraphPad Software Inc., CA – USA).
To study the GSK-3β kinetics, the reaction mixture, varying concentrations of ATP (0.25 – 0.5 – 1 – 2 – 4 µM) or substrate (12.5 – 25 – 50 – 100 – 200 nM) versus test samples (20 and 50 µM), was incubated for 5, 15, 30, and 60 min at 23°C, followed by the addition of the 8 mM EDTA and the anti-phospho-GS antibody according to the manufacturer's protocol.
For GSK-3β kinetic experiments, initial velocities (V_0) were determined and fitted to Michaelis-Menten equation. To directly visualize G12 and G5 inhibition mode, a Lineweaver-Burk plot was generated according to the values obtained from the Michaelis-Menten analysis. The slope corresponded to K_m/V_{max}, the intercept on the vertical axis to $1/V_{max}$, and the intercept on the horizontal axis to $-1/K_m$. Moreover, at the reciprocal of the smallest value of substrate concentration ($X=1/[S_{min}]$) was associated a value of Y representing the equation $Y=(1/V_{max})(1.0 + K_m/[S_{min}])$. Graphs and data analysis were performed using GraphPad Prism 7 software.

Thermal Shift Assay hit binders characterization

Thermal shift experiments were conducted in technical triplicates and recorded with *ViiA7™* real-time PCR instrument. The analysis protocol of Partch and colleagues (2015) was followed, developed for fast stability screening of buffers and interactors for recombinant proteins[44]. Briefly, the recombinant protein GSK-3β at a final concentration of 5 µM was mixed with the dye SYPRO Orange (2x final concentrated) in the selected assay buffer (NaCl 300 mM, Hepes pH 7.5 20 mM, $MgCl_2$ 5 mM, DTT 1 mM) and let it incubate with hit compounds at 2 different concentrations, 200 and 500 µM. Moreover, experiments were performed with GSK-3β also in the presence of AMP-PNP at a final concentration of 200 µM. A 96-well plate was prepared with samples in triplicate; each sample-well was composed by 40 µL of protein solution and 10 µL of compound solution (1 µL compound stock 50x diluted in 9 µL MilliQ water). DMSO final concentration in each sample was 2%. Blank controls of the experiment were designed, the first with protein, SYPRO Orange 2x and 2% DMSO and the second with compounds/DMSO, SYPRO

Orange 2x without protein. After filling plate wells with the required 50 µL solution and resuspending properly, the plate covered with ad adhesive sheet was centrifuged 800g for 2 minutes at 25°C to remove bubbles eventually formed during preparation. As soon as the plate was placed into the Applied Biosystems ViiA7 real-time PCR instrument the following parameters were selected:
EXPERIMENTAL PROCEDURE AND DESIGN: Set up *Fast 96-well block (0.1 mL)*; Experiment type *MELT CURVE;* Reagents used to detect target sequence *OTHER*; Ramp speed *STANDARD;* Reporter *ROX;* Quencher *NONE;* RUN METHOD (melt curve profile): *"Step and hold" 1:00;* an initial 2:00 hold at 25 °C, ramping up in increments of 1 °C to a final temperature of 95 °C (with a 2:00 hold).
Output data were collected and analyzed with Excel software and Graphpad Prism 7.

Protein-molecule complex crystallization and X-ray data collection

Crystallization procedures and X-ray data collection were performed in collaboration with dr. S. Tripathi (IIT Genova). Purified GSK-3β protein was concentrated to 3-4 mg/ml. Crystals were grown with the hanging drop vapor diffusion method, using 20 mM Hepes pH 7.5, 50 mM $MgCl_2$ and 15-20% PEG3350 as crystallization buffer. Protein crystals appeared within 1-2 days. After that, crystals were soaked in a solution, which was constituted of the same crystallization buffer supplemented with 6-8% glycerol and 1 mM final concentration of G5, G12, ARN1484, ARN9133. These crystals were frozen in liquid nitrogen, after 5-6 hours of soaking.

Data collection was performed in Macromolecular crystallography beamline in Trieste, Italy. Data reduction was performed using iMOSFLM[45], scaling was performed using AIMLESS[46] and refinement was performed using REFMAC[47]. Ligand drawing was performed using ligand in CCP4 suit. For ligand site identification, OMIT maps were used[48]. Model modification and visualization was performed using Coot[49]. Images were finally constructed using Pymol[50].

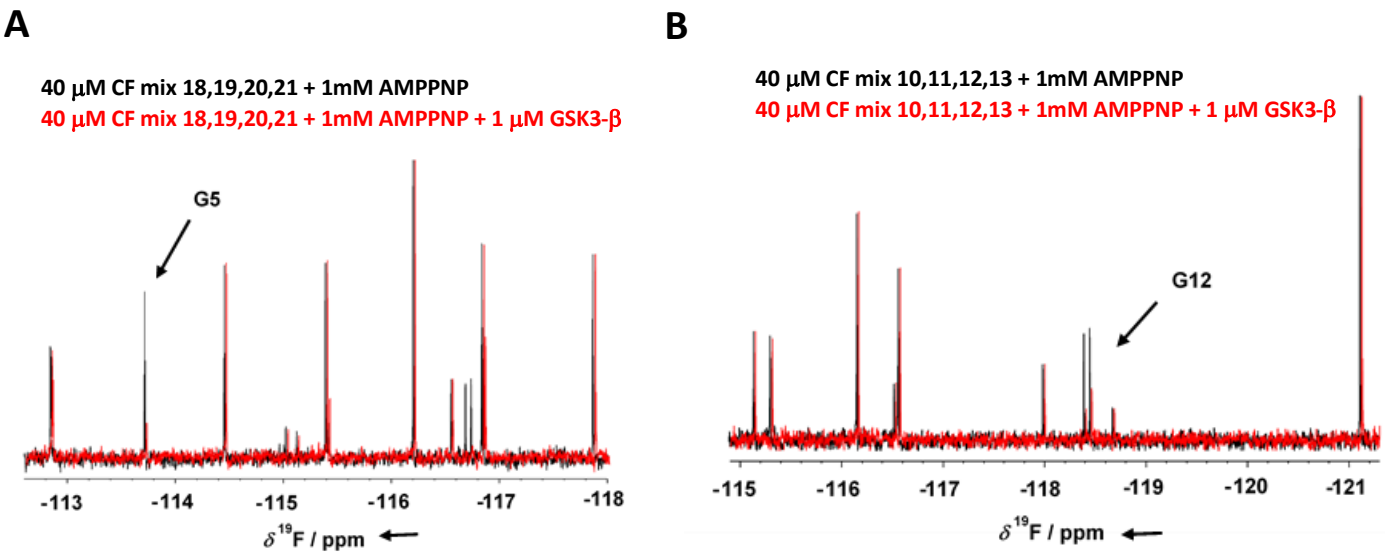
Results

^{19}F NMR ligand-based binding screening and hits identification

In order to identify novel hit compounds able to inhibit GSK-3β, a subset of 350 fluorinated compounds (average MW of 270 Da) of the internal LEF compounds library was initially screened against the target protein in mixtures of 25 each, by ^{19}F NMR in the presence of saturating AMP-PNP concentrations (1 mM), a non-hydrolizable ATP analogue. AMP-PNP concentrations used in the screening were selected based on previously performed GSK-3β AMP-PNP saturation experiments (data not shown). NMR experiments were performed by dr. M. Veronesi (IIT Genova).

GSK-3β was pre-incubated at room temperature for 10 minutes in the presence of 1 mM AMP-PNP prior to addition to the mixtures. Binding compounds (hits) could be easily identified in the

NMR screening comparing ^{19}F cpmg spectra of the mixture in the absence and in the presence of the protein as shown in Fig. 2A and 2B (black and red traces, respectively). The NMR signals of the molecules interacting with the protein showed a significant reduction in their intensity in the presence of the protein. On the contrary, NMR signals of the non-binding molecules were not affected by the presence of the target protein. Molecules selected from this first NMR screening were then tested as single compound under the same screening conditions in the presence of two non-binding compounds as negative controls. This step was performed to avoid false positive results, since, some fragments in mixtures can have a cooperative behavior that can favor their non-specific binding. Indeed, fragments, which alone did not confirm the binding observed in the mixture, were excluded from the following analyses. The remaining promising compounds, still showing binding to GSK-3β when in single, were retested also in the presence of 3 mM AMP-PNP (among these, G12 and G5, whose NMR spectra is reported in Fig. 2C and 2D, respectively). Finally, 32 compounds were still showing binding to GSK-3β also in the presence of higher concentrations of AMP-PNP. These hits were selected to be further analyzed and characterized with other biophysical techniques.



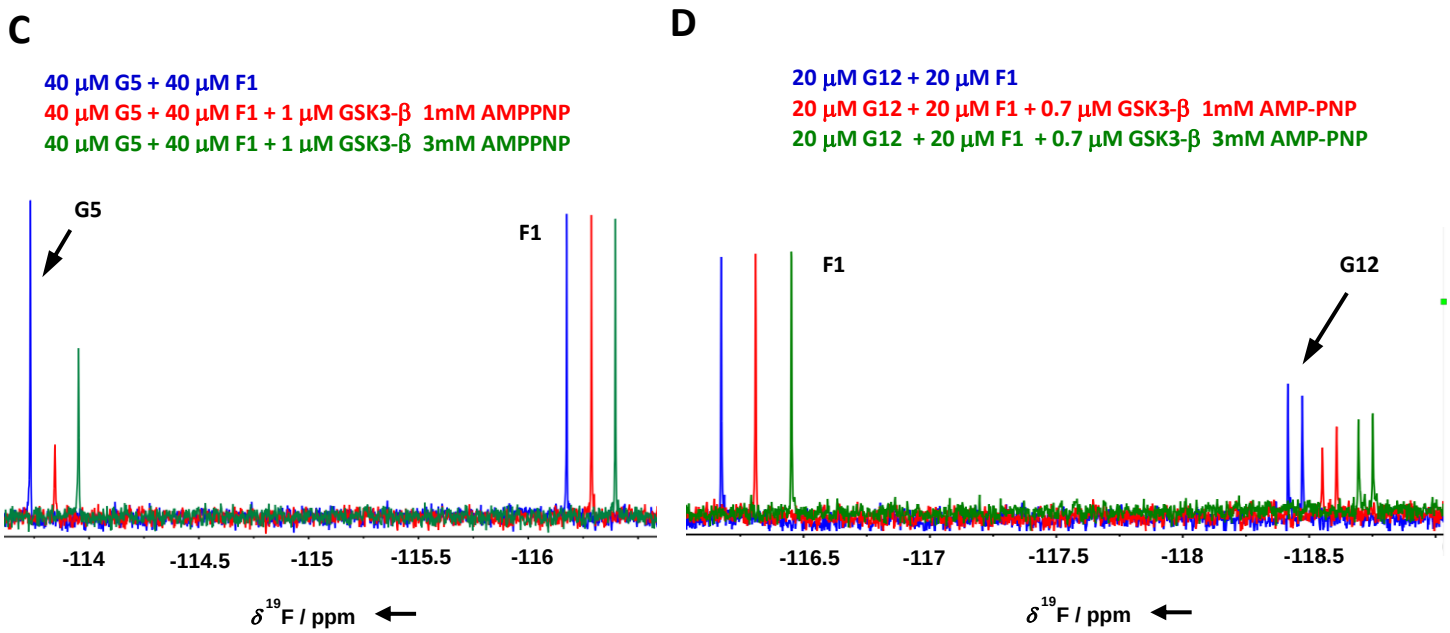


Fig. 2) A and B) Fragment mixtures containing G12 and G5, in the absence (black) or presence (red) of GSK-3β. Binding compounds G12 and G5 give clear changes in resonance signal when in the absence or in the presence of the protein; C and D) Fragments G12 and G5 bind GSK-3β in single and they are not displaced completely by AMP-PNP. The images show G12 and G5 $T_2\rho$ spectra in the absence of GSK-3β (blue), in the presence of GSK-3β and 1 mM AMP-PNP (red) and in the presence of GSK-3β and 3 mM AMP-PNP (green). F1 compound signal is used as negative internal control.

Hits selection and validation and K_d determination through MicroScale Thermophoresis analysis

The 32 compounds, identified through NMR, were further validated in MST binding experiments. Where solubility issues or fluoresce interferences were not present, subsequent MST experiments, at increasing compounds concentrations, were set up to determine the affinities of the compounds for the target protein.

Binding check experiments were performed as direct bindings or as displacement assays in the presence of 3 mM AMP-PNP (data not shown). The compounds which were showing good response in direct binding assays were thirteen. Among these, only four hits (i.e. G12, G5, B3 and E2) maintained their ability to bind GSK-3β upon incubation with AMP-PNP. Only for G12 it was possible to build a complete binding affinity curve and to identify a binding affinity parameter k_d for GSK-3β of 66.5 ± 31.8 µM (Fig 3). Low solubility of compounds B3 and E2 and G5, under the conditions used for these tests, did not allow to obtain complete binding curves (buffer 0.1 M pH 7.2, 5 mM $MgCl_2$, 0.001% Tween80, 0.1% PEG8000; DMSO not higher than 5% to avoid unspecific effects of DMSO on the protein folding and activity). Further data analyses suggested that

solubility issues for these compounds may be coupled to an overestimation of the affinity parameters due to the steric hindrance induced by the fluorescent label on GSK-3β lysines (23 lysines are present in GSK-3β sequence). Indeed, MST experiments provided for staurosporine binding to GSK-3β a k_d of ~ 67 nM (data not shown), compared to less than 10 nM reported in the literature[29,32,51].Binding data were therefore confirmed by a second label-free biophysical approach, to prevent label-induced steric hindrance (i.e. SPR).

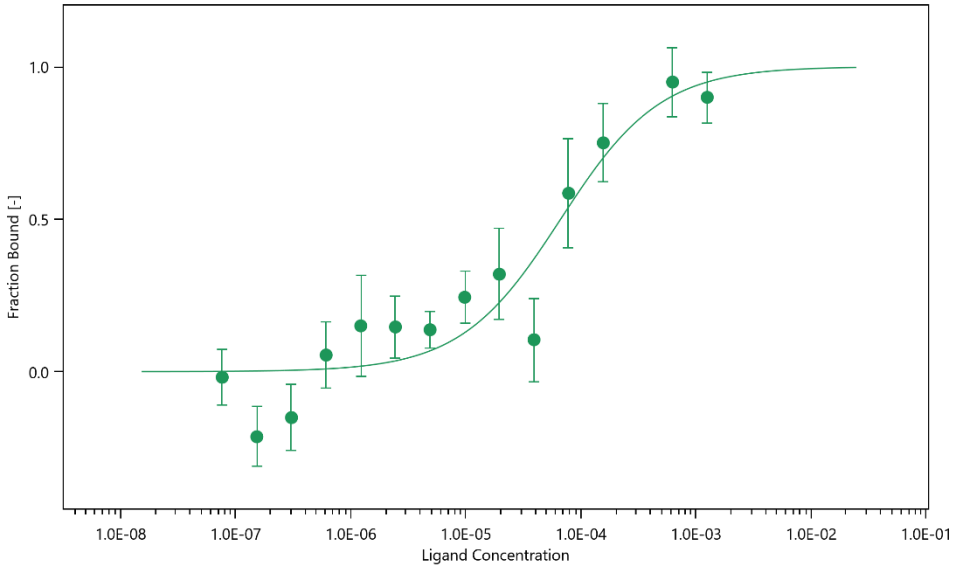


Fig. 3) Graph reporting MST experiments performed with GSK-3β for G12 binding kinetics definition. Sigmoidal fitting curves were obtained using the Affinity Analysis software of Nanotemper Technologies.

Measurements of hits binding to GSK-3β through SPR

Hit compounds were validated and further characterized by SPR. SPR experiments were performed as described in the *materials and methods* section and allowed to confirm G12 and G5 binding to GSK-3β, with a K_D value (steady state) of 11.22 ± 0.02 µM and 7.79 ± 0.05 µM, respectively. SPR sensorgram of G12 is reported in Fig. 4. Notably, these experiments are preliminary data and still require minor technical optimizations.

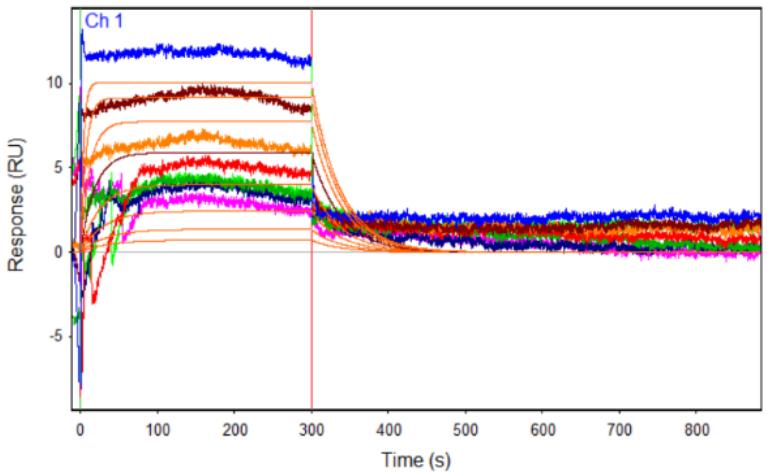


Fig. 4) SPR sensorgram of binding assay of G12 on GSK-3β. G12 showed a K_D value of 11.22 µM.

Inhibitory activity of the identified hits

The inhibitory activity of the four identified hit compounds was tested at four concentrations (1 – 5 – 50 – 100 µM). While Compounds B3 and E2 showed a poor ability of inhibiting GSK-3β even at the highest concentration, G12 and G5 were selected for IC_{50} determination, since they showed a significant percentage of GSK-3β inhibition higher than 80% at 100 µM (Table 1).

G12 and G5 showed a comparable potency with an IC_{50} of 15.25 ± 1.34 µM and 14.81 ± 0.55 µM, respectively (Fig. 5).

Table 1

	G12		**G5**		**B3**		**E2**	
	% Inhibition		% Inhibition		% Inhibition		% Inhibition	
Dose	Average	STDev	Average	STDev	Average	STDev	Average	STDev
1 µM	8,5	1,0	0,3	0,1	2,5	0,9	ni	-
5 µM	15,2	2,7	13,4	2,0	4,9	1,5	ni	-
50 µM	71,4	1,7	69,4	2,0	27,1	4,9	8,2	1,9
100 µM	88,0	1,2	84,5	0,4	46,6	6,0	18,2	1,5

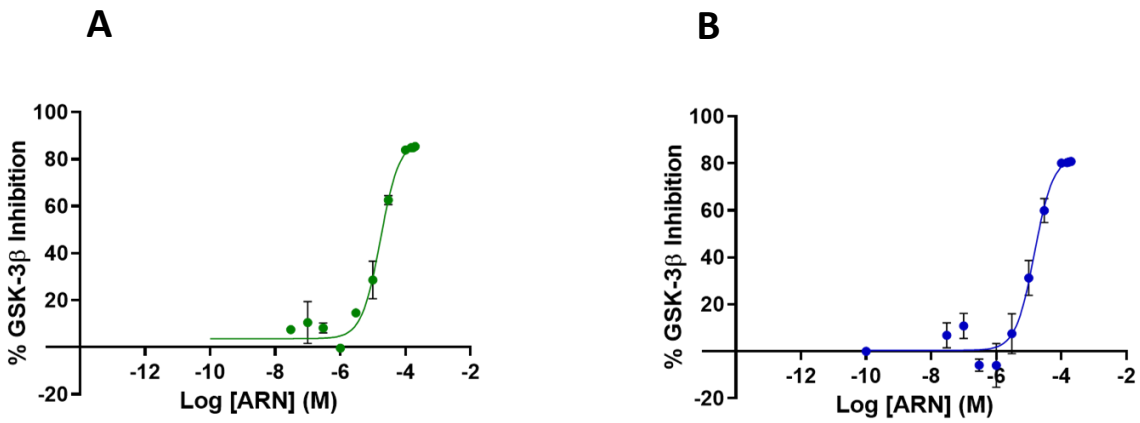


Fig. 5) Dose-response curves for GSK-3β inhibition by G12 (A) and G5 (B). G12 and G5 compounds showed an IC_{50} of 15.25 and 14.81 µM, respectively.

To have insights on the GSK-3β inhibitory mechanism, G12 and G5 (20 and 50 µM) were tested to competitively replace ATP or the GSK-3β substrate ULight-GS (Ser641/pSer657), as described in the *materials and methods* section.

First, under a constant concentration of the substrate ULight-GS (50 nM), ATP concentrations were varied from 0.25 to 4 µM and G12 and G5 were tested at 20 and 50 µM. We observed an increase in K_m constant (*Michaelis-Menten* constant, which numerically corresponds to the substrate concentration at which the reaction rate is half of the maximal velocity V_{MAX}), but unaltered $1/V_{max}$ value, when the concentrations of G12 and G5 increased (Fig 6A), suggesting a competition between ATP and the two compounds. Additionally, G12 and G5 (at 20 and 50 µM) were tested keeping constant the ATP concentration (1 µM), and varying the substrate ULight-GS concentration from 12.5 to 200 nM. At increasing concentrations, both compounds showed an increased $1/V_{max}$ value, but they differed in effect on K_m, which was unaltered for G12 while it increased for G5 (Fig 6B): this suggested that G12 did not compete with the substrate while G5 showed a mixed competition mechanism.

Activity data therefore allowed us to show that despite the two selected molecules have very similar IC_{50}s and they are both ATP competitive, G12 does not compete with the subtract, while G5 shows a mixed competition towards the subtract.

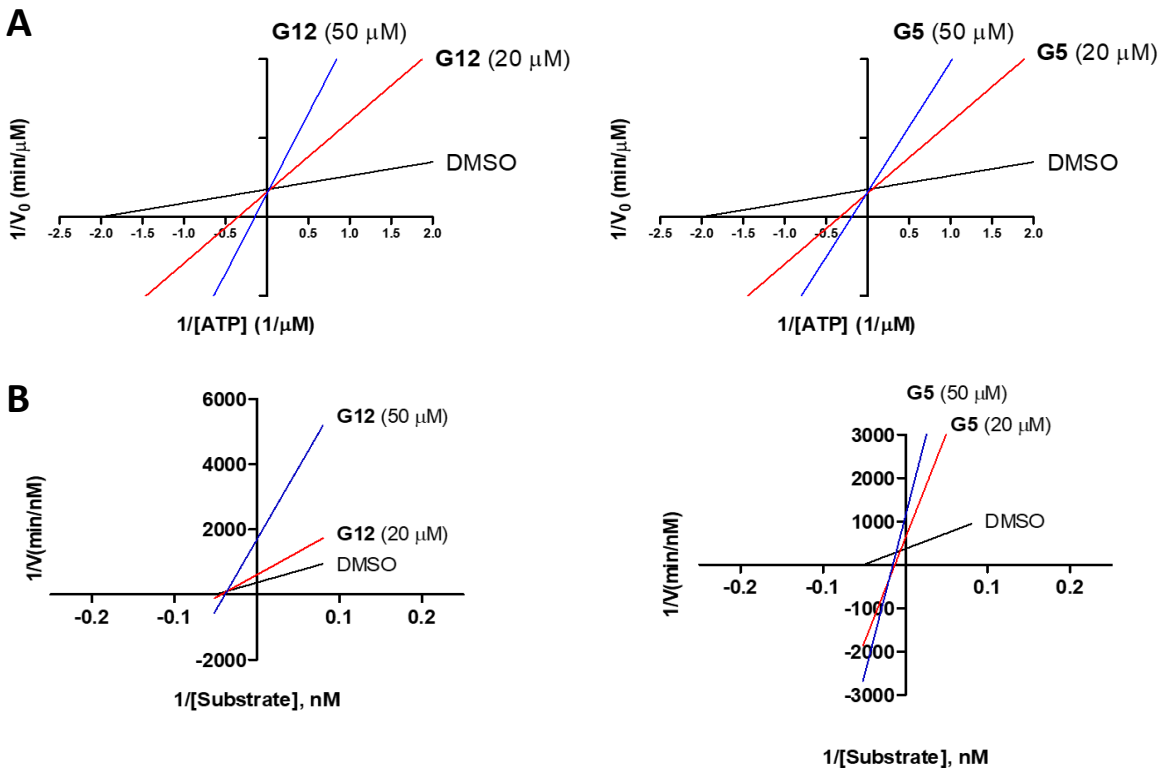


Fig. 6) Lineweaver–Burk plots of GSK-3β kinetic data at two concentrations of G12 and G5 (20 and 50 μM). A) Linear regression plotting of 1/V against 1/ATP at a given concentration of compound. Intersecting at the same point on the y-axis indicates competitive inhibition with respect to ATP; B) Linear regression plotting of 1/V against 1/Substrate at a given concentration of compound. Intersecting at the same point on the x-axis indicates noncompetitive inhibition (G12), while for G5 the x intercept shifts right suggesting a mixed inhibition with respect to the substrate.

Protein Thermal shift stabilization effect study

Given the poor inhibitory activity of compounds B3 and E2, only G12 and G5 were selected for further biophysical and structural investigations. Thermal shift experiments allowed to observe that both compounds have a stabilizing effect on the protein structure upon binding (Fig. 7). In fact, the binding of each fragment to GSK-3β led to an increase of protein melting temperatures (T_m). A T_m increase of 2.5 °C and 3.2 °C was observed upon binding of G12 and G5, respectively. Reported effects were compared and normalized to the effect induced by DMSO alone on GSK-3β denaturation. The same experiments were performed also in the presence of AMP-PNP to unravel possible effect of this molecule on protein denaturation. Notably, observed changes in T_m were not affected by the presence of AMP-PNP.

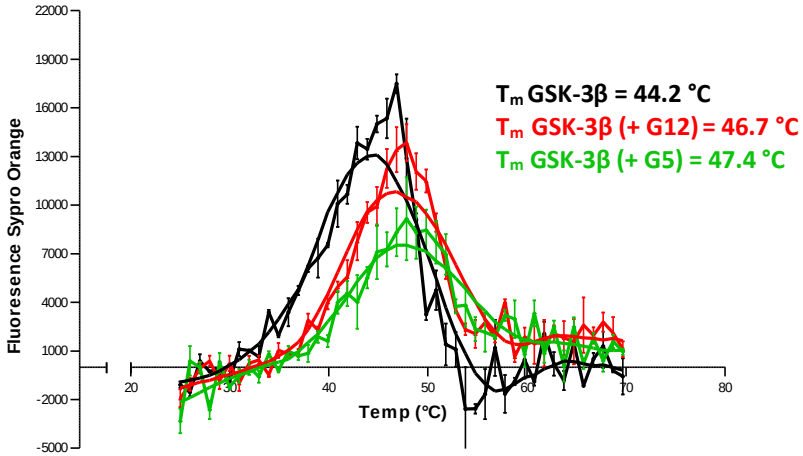


Fig. 7) Thermal Shift Assays of GSK-3β in the presence of G12 and G5 compounds. G12 (red trace) and G5 (green trace) stabilize GSK-3β protein and lead to an increase in T_m of 2.5 °C and 3.2 °C, respectively.

Crystal structure and complex definition

In order to further characterize the binding of the two identified compounds to GSK-3β, structural investigations were pursued through X-ray crystallography. Crystallization procedures and X-ray data elaboration were performed in collaboration with dr. S. Tripathi (IIT Genova). Briefly, crystals of GSK-3β in complex with G5 or G12 were prepared by soaking compounds into the crystals of the protein in the apo form (no added compounds). Crystals were obtained after 1-2 days by the hanging drop vapor diffusion method (see *materials and methods* section), after buffer screening and optimization procedures. Data collection was performed at the Elettra Synchrotron, Trieste. Structures of G5 and G12 in complex with GSK-3β were solved with the molecular replacement method at a 2.5 and 2.2 Å resolution, respectively (PDB to be deposited). Data collection and refinement statistics are presented in Table 2.

Unambiguous electron density corresponding to both G5 and G12 were observed in the ATP binding pocket of GSK-3β (Fig 8A and 10A).

Interactions demonstrated by G5 are reminiscent of GSK-3β specific maleimide analogs. The methyl sulphonyl group of G5 interacts with Lys85 and Asp200 sidechains via hydrogen bonding interactions (Fig 8C). The core pyrimidine 2-amine is involved in a direct hydrogen bonding interaction with the backbone oxygen of Asp133 and nitrogen of Val135 (Fig 8B). The conserved water molecule next to Thr138 is observed close to the putative fluoro-phenyl site (Fig 8D). Fluorine of fluoro-phenyl part of G5 is involved in direct hydrogen bond with Arg141 (Fig 8D). The Gln185 and Arg141, which are hydrogen bonded to the conserved water molecule and the

maleimide derivative in the structure of PDB 1Q4L, are flipped away from water in our structure (Fig 9A).

The binding position of G12 in the ATP pocket completely differs from G5. Electron density corresponding to G12 is represented in Fig. 10. Direct hydrogen bonding interaction is observed between pyridine nitrogen and carboxamide nitrogen with the backbone atoms of Tyr134 and Val135, respectively (Fig. 10B). The propanoyl oxygen of G12 is involved in hydrogen bonding interaction with Arg141 (Fig. 10B). Electron density corresponding to fluoro-phenyl group is not very well defined, suggesting that this portion of G12 molecule is lacking of adequate stabilizing interactions with the protein. A water mediated hydrogen bond network is observed between carboxamide oxygen of G12 and backbone oxygen of Gln185 (Fig. 10C). A similar interaction has been already observed and proposed to justify specificity of ruthenium-based inhibitors against GSK-3β[52]. Furthermore, to validate G12 binding position on GSK-3β, crystals of GSK-3β in the presence of G12 and increasing concentrations of AMP-PNP were grown. The collected electron density data confirmed the ATP competitive behavior of G12, showing the compound being progressively displaced from the ATP-binding site in an AMP-PNP dependent fashion (Fig. 11).

Table 2

	G5	G12
Data reduction		
Resolution (Å)	2.5-48.8 (2.5-2.59)	2.2-89 (2.2-2.26)
Rmerge	0.388 (3.619)	0.319 (3.86)
Mean(I/sd(I))	7.9(0.8)	6.9 (0.6)
Completeness	100(100)	95.6 (87.8)
Unique Reflections	44494 (4582)	61582(4107)
Multiplicity	13.4 (12.9)	12.2(11.4)
$CC_{1/2}$	0.992 (0.39)	0.993(0.37)
Refinement		
Resolution (Å)	2.5-48.8	2.2-89
Rwork/Rfree	0.225/0.268	0.24/0.27
RMS bonds (Å)	0.0086	0.0075
RMS angles (º)	1.808	1.573

[#]Numbers in parentheses are values in the highest resolution shell.

[†] $R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, *where $I_i(hkl)$ is the intensity of the i^{th} measurement and $\langle I(hkl)\rangle$ is the mean intensity for that reflection.*

[‡] $R_{work} = \sum_{hkl}\big||F_{obs}| - |F_{calc}|\big|/\sum_{hkl}|F_{obs}|$, *where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure-factor amplitudes, respectively.*

[+]R_{free} was calculated with 5.0% of reflections in the test set.

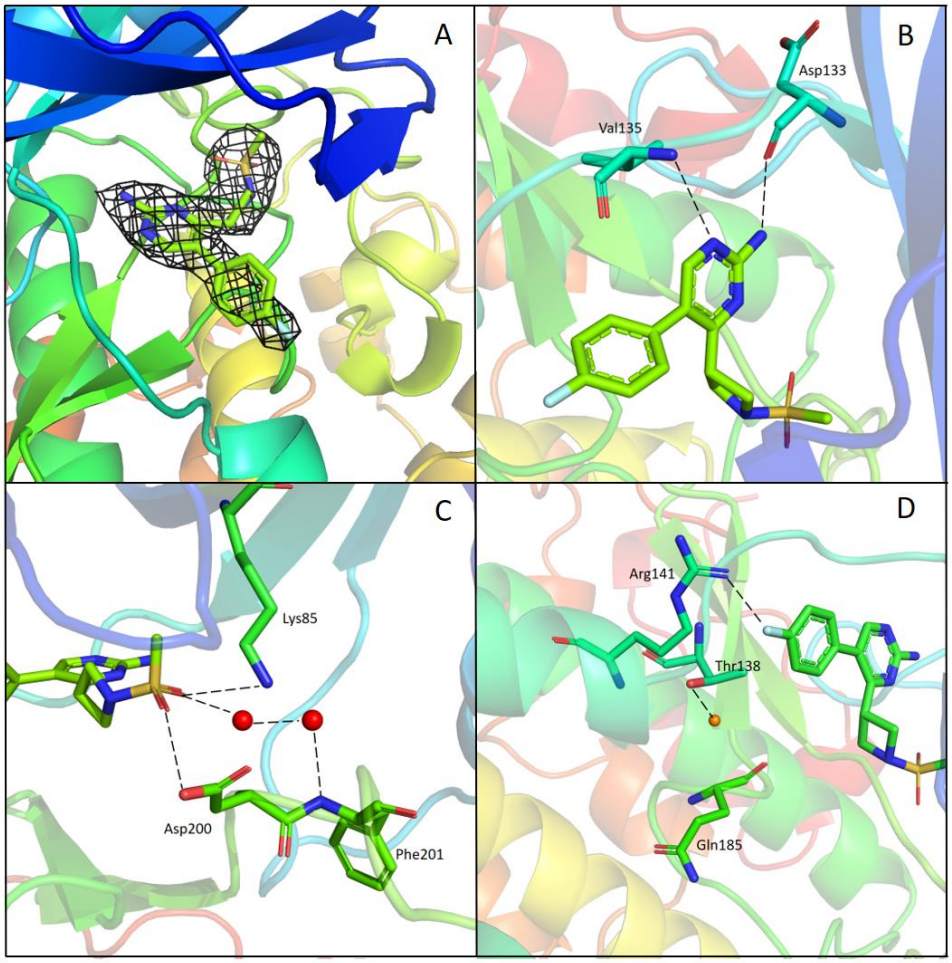


Fig. 8) A) Electron density corresponding to G5 contoured at 1.2 sigma; B) Direct hydrogen bonding interaction between the core pyrimidine 2-amine of G5 with the backbone oxygen of Asp133 and nitrogen of Val135; C) Hydrogen bonding interactions of methyl sulfonyl group of G5 with Lys85 and Asp200. Water (red spheres) mediated hydrogen binding network is also observed; D) Conserved water (red sphere) is observed hydrogen-bonded to Thr138. Fluorine of fluoro-phenyl is hydrogen bonded to Arg141.

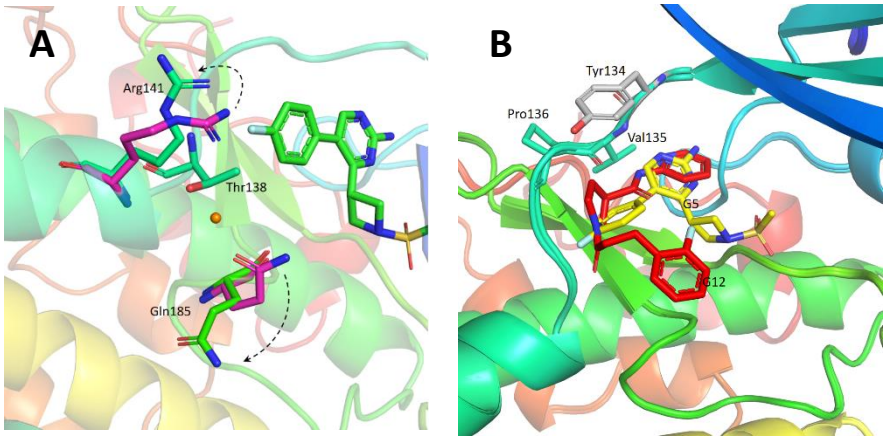


Fig. 9) A) In GSK-3β-G5 crystal complex (green), residues Arg141 and Gln185 are flipped away from the binding pocket in comparison to PDB 1Q4L (purple); B) Overlapped structures of G12 (red) and G5 (yellow).

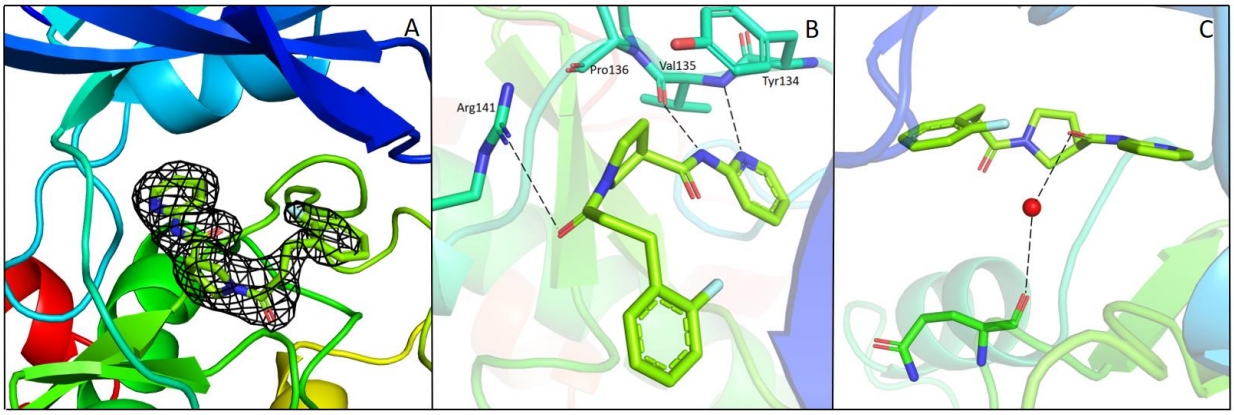


Fig. 10) A) Electron density corresponding to G12 contoured at 1.2 Sigma; B) Hydrogen bonding interaction of G12 with the residues of hinge region; C) Water mediated hydrogen bond between carboxamide oxygen of G12 and Gln185.

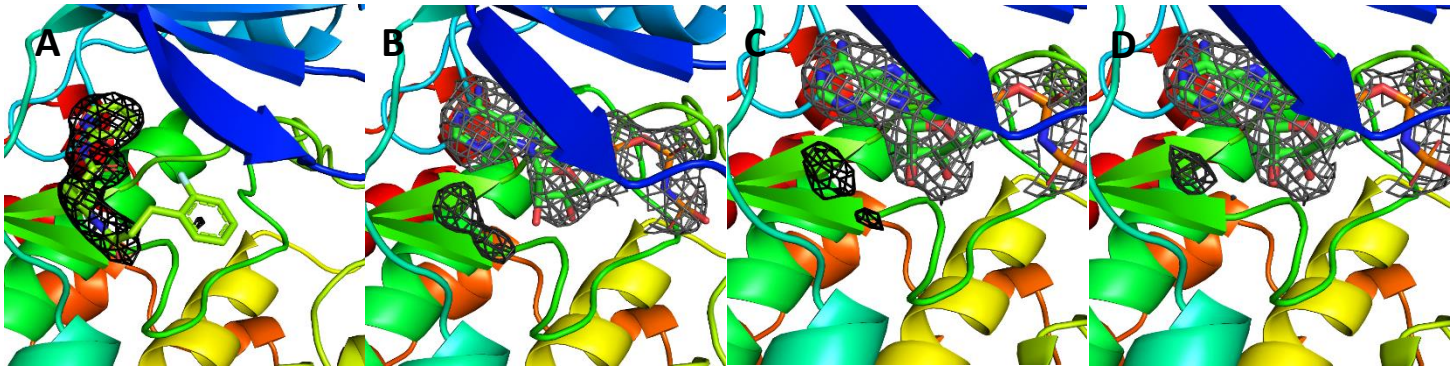


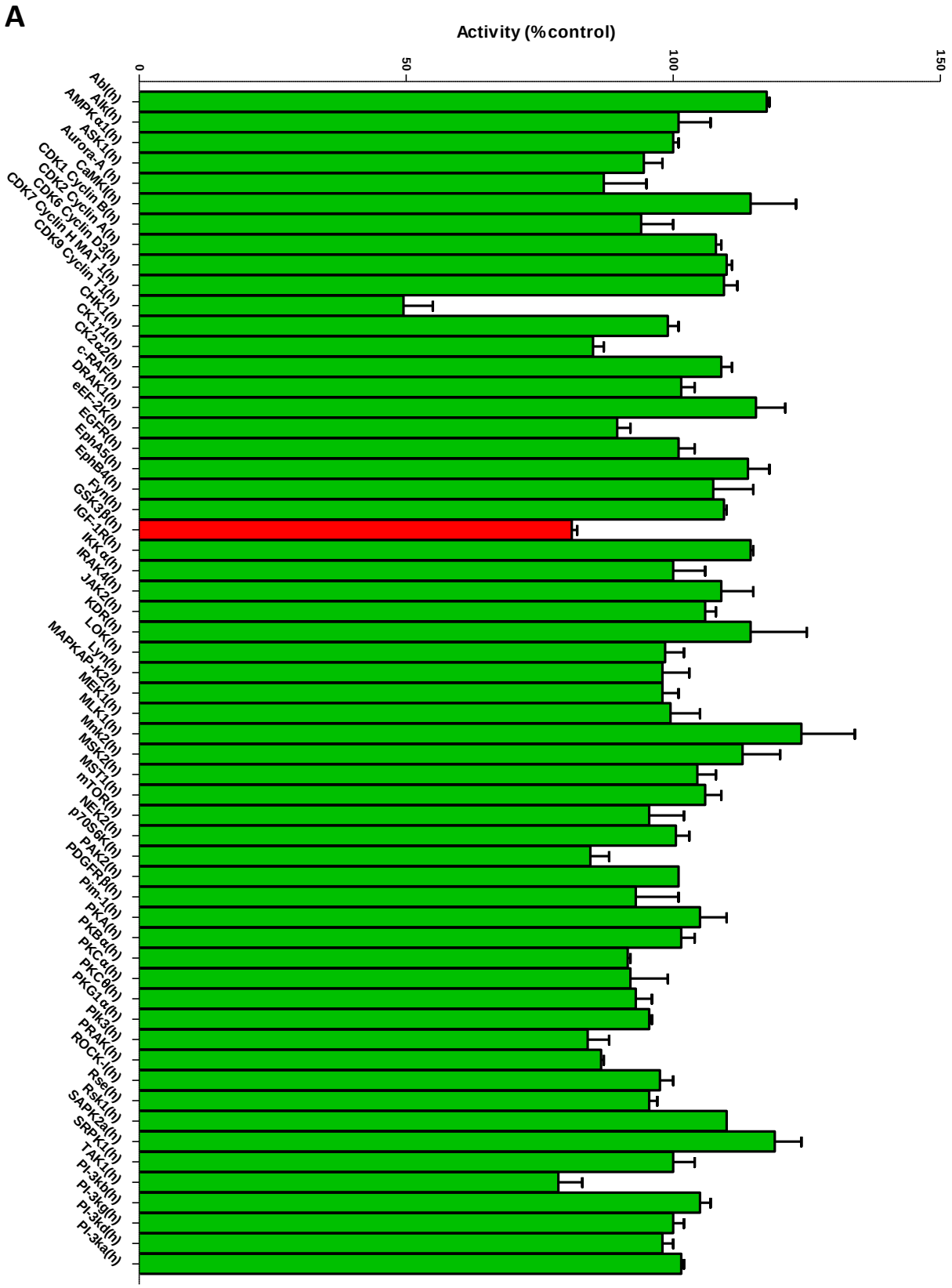
Fig. 11) 3D structures GSK-3β in complex with G12 in the presence of increasing concentrations of AMP-PNP (0 mM AMP-PNP (A); 1 mM AMP-PNP (B); 2 mM AMP-PNP (C); 4 mM AMP-PNP (D)).

G12 and G5 selectivity evaluation through a Kinase Panel

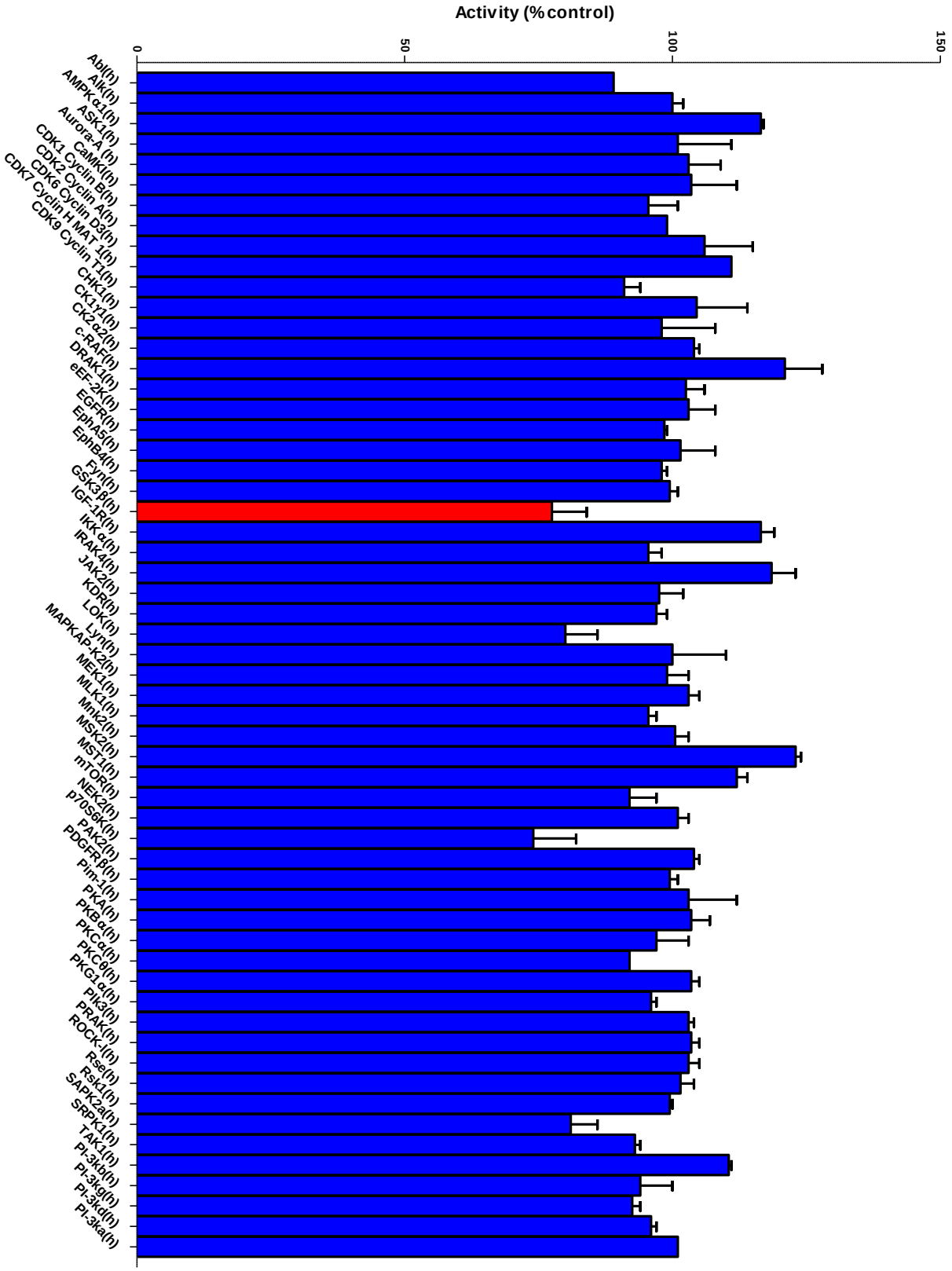
Using the *Kinase screening and profiling* service of *Eurofins* (Kinase Diversity Panel) the selectivity of the identified hit compounds G12 and G5 towards a panel of 58 kinase of the humane proteome was evaluated. Inhibitory activity assays were performed at constant compounds concentrations (20 µM) and ATP concentrations specific for each kinase (± 15 µM K_m). Results are reported in Fig. 12A and Fig. 12B.

From these analyses both G12 and G5 showed high selectivity towards GSK-3β. Notably, G12 showed also a significant inhibitory effect on CDK9 (50%), TAK1 (20%), p70S6k (15%) and G5 showed a significant inhibitory effect also on p70S6K (26%), SAPK2a (19%), LOK1 (21%). These cross-reactivity results provided new information on inhibitors binding ability, and suggested the existence of common structural and binding features among the kinases-inhibitors complex structures. These evidences will be further discussed below.

Fig. 12) Bar Graphs showing G12 (A) and G5 (B) inhibitory effect on different kinases of the commercial panel (Eurofins service); All the experiments were performed using 20 µM of compound and ATP concentrations ± 15 µM of each $K_{m.}$


A
Activity (% control)
0
50
100
150
Abl(h)
Alk(h)
AMPKα1(h)
ASK1(h)
Aurora-A (h)
CaMKI(h)
CDK1 Cyclin B(h)
CDK2 Cyclin A(h)
CDK6 Cyclin D3(h)
CDK7 Cyclin H MAT 1(h)
CDK9 Cyclin T1(h)
CHK1(h)
CK1γ1(h)
CK2α2(h)
c-RAF(h)
DRAK1(h)
eEF-2K(h)
EGFR(h)
EphA5(h)
EphB4(h)
Fyn(h)
GSK3β(h)
IGF-1R(h)
IKKα(h)
IRAK4(h)
JAK2(h)
KDR(h)
LOK(h)
Lyn(h)
MAPKAP-K2(h)
MEK1(h)
MLK1(h)
Mnk2(h)
MSK2(h)
MST1(h)
mTOR(h)
NEK2(h)
p70S6K(h)
PAK2(h)
PDGFRβ(h)
Pim-1(h)
PKA(h)
PKBα(h)
PKCα(h)
PKCθ(h)
PKG1α(h)
Plk3(h)
PRAK(h)
ROCK-I(h)
Rse(h)
Rsk1(h)
SAPK2a(h)
SRPK1(h)
TAK1(h)
PI-3kb(h)
PI-3kg(h)
PI-3kd(h)
PI-3ka(h)

B


Activity (% control)
0
50
100
150
Abl(h)
Alk(h)
AMPKα1(h)
ASK1(h)
Aurora-A (h)
CaMKI(h)
CDK1 Cyclin B(h)
CDK2 Cyclin A(h)
CDK6 Cyclin D3(h)
CDK7 Cyclin H MAT 1(h)
CDK9 Cyclin T1(h)
CHK1(h)
CK1γ1(h)
CK2α2(h)
c-RAF(h)
DRAK1(h)
eEF-2K(h)
EGFR(h)
EphA5(h)
EphB4(h)
Fyn(h)
GSK3β(h)
IGF-1R(h)
IKKα(h)
IRAK4(h)
JAK2(h)
KDR(h)
LOK(h)
Lyn(h)
MAPKAP-K2(h)
MEK1(h)
MLK1(h)
Mnk2(h)
MSK2(h)
MST1(h)
mTOR(h)
NEK2(h)
p70S6K(h)
PAK2(h)
PDGFRβ(h)
Pim-1(h)
PKA(h)
PKBα(h)
PKCζ(h)
PKCθ(h)
PKG1α(h)
Plk3(h)
PRAK(h)
ROCK-I(h)
Rse(h)
Rsk1(h)
SAPK2a(h)
SRPK1(h)
TAK1(h)
PI-3kb(h)
PI-3kg(h)
PI-3kd(h)
PI-3ka(h)

Discussion

In this work, an initial NMR screening campaign of the IIT internal fluorinated fragments library allowed the identification of hit compounds able to bind GSK-3β in the presence of saturating AMP-PNP concentrations. This starting approach led to the selection of compounds able to bind GSK-3β in an allosteric site or in the ATP binding site but with an affinity at least comparable with the one of AMP-PNP. It is important to remember that different biophysical techniques have different sensitivities for protein binding detection and therefore hits, which are identified with one technique, can result as no binders with other methods. This is especially true for ^{19}F NMR that, as demonstrated by Dalvit 2006, is the most sensitive technique for the identification of very weak binders, which many times cannot be identified by other biophysical methods[53].
On the one hand, NMR is the best technique for screening fragments or small molecules that even if weak binders may be the critical puzzle pieces fitting in a protein hotspot with the potential of being evolved into an efficient lead compound[37,54]. On the other hand, among the weak binders, our goal was to set an initial cut-off (AMP-PNP affinity) to avoid too weak binders and narrow down the initial screening campaign. Notably, a similar initial step of the drug discovery screening cascade has been very recently published by *AstraZeneca* as a tool in the search of allosteric inhibitors of MEK1 kinase[40].
From our initial NMR screening on GSK-3β, the identified compounds were tested with MST to validate the binders and further narrow down the selected hits: indeed, only four compounds among the 32 molecules selected by NMR were showing binding to GSK-3β both in absence of AMP-PNP and under AMP-PNP saturating concentrations. MST allowed also to determine the affinity k_d of one of the four compounds for GSK-3β. G12 and G5 affinities for GSK-3β were studied also using SPR, which, provided preliminary K_D values of 11.2 μM and 7.8 μM for G12 and G5, respectively.

After identifying the first set of potential GSK-3β binders, activity assays were performed to better characterize the kinetics properties of the binding and to get insights on the mechanism of inhibition. G12 and G5 showed a good GSK-3β inhibitory activity (more than 80% at 100 μM concentration) and were selected for further investigations. Indeed, from displacement assays, G12 emerged to be an ATP-competitive inhibitor not substrate competitive, while G5 to be an ATP-competitive inhibitor of GSK-3β with a mixed inhibition for the substrate.

Additionally, binding effects of G12 and G5 of GSK-3β structure was studied using thermal shift assays. The reported increase of the protein melting point (T_m) in the presence of each of the two compounds allowed to confirm the binding event as a stabilization effect induced by the compounds on the protein structure.

A kinase inhibitory activity panel study, (Fig. 12), showed that both G5 and G12 were relatively more specific towards GSK-3β compared to other kinases. Nevertheless, both compounds displayed interactions with the protein that could be traced to known ATP- competitive GSK-3β inhibitors.

In the attempt of explaining the observed specificity of the identified compounds for GSK-3β, the structures of few kinases, present in the selectivity panel study, were compared and overlapped to the structure of GSK-3β in complex with G5 and G12. For G5 compound, we initially ascribed the observed specificity towards GSK-3β to its interactions with Lys85, Asp200 and Arg141 based on structural comparisons and previous study of Kramer and colleagues, which claimed that the residues Lys85, Asp133, Val135, Glu137, Arg141, Gln185, Asp200, Arg220 are crucial amino acids for interactions with the binding pocket of GSK-3β[55]. In fact, kinases reporting G5 inhibitory effects similar to the one of GSK-3β (i.e. p70S6K, SAPK2a, LOK1) have conserved Lys85 and Asp200 residues (Fig. 13C) Surprisingly, also kinases whose activity was only marginally affected by G5 (such as AMPK1α and MEK1) conserve Lys85 and Asp200 in the same positions. This analysis suggests that the presence of these residues might not be the explanation for G5 selectivity. The structure of GSK-3β-G5 was also overlapped with the available GSK-3β-AMP-PNP structure (PDB 1PYX). Here, a relevant proximity of the methyl sulphonyl group of G5 to the first phosphate of AMP-PNP is visible (Fig 13A, black arrows), suggesting that G5 inhibits GSK-3β acting as ATP. Moreover, from the kinase panel, G5 did not show any evident inhibitory activity on CDKs (Fig. 12B), even though these kinases have a high structural homology with GSK-3β (86% structure homology[56]). Leu132 in GSK-3β is replaced by a bulkier phenylalanine in all the CDKs (Fig. 13B). In GSK-3β, water molecules next to Leu132 are involved in a hydrogen bonding network with methyl sulphonyl group of G5 and Lys85, Asp200. When in CDKs Leu132 is replaced by a bulkier phenylalanine, this network is disrupted (Fig. 13B). Indeed, the presence of Leu132 and the related water-mediated hydrogen bonding network could be the reason of the specificity of G5 inhibitory activity on GSK-3β.

As for G5, we compared GSK-3β-G12 structure with the structures of some of the kinases present in the kinase panel. For some kinases, G12 (i.e. CDK9, TAK1) was showing similar or better inhibition compared to GSK-3β, whether others kinases (i.e. MLK1, SAPK2a, ABL1) were almost not affected by the presence of the small molecule. CDK9 and GSK-3β have a sequence homology of 48% and very similar structures (Fig. 14A). However, according to structures overlap, we may hypothesize that the presence of Phe103 in CDK9 leads to a shift of G12 along the hinge and to the formation of a new π-stacking interaction between the pyrimidine ring of G12 and the phenylalanine ring moiety. This new interaction could be the explanation of G12 inhibitory activity on GSK-3β at variance with CDK-9.

Other considerations were made comparing MLK1 (i.e. MAP3K9) and GSK-3β-G12 structure. MLK1 shares 47% of sequence homology with GSK-3β. Arg141 in GSK-3β is retained as Arg230 in

MLK1 (Fig. 14B), even though no inhibition was reported. In addition to this, in MLK1 structure, Pro136 of GSK-3β is replaced by Arg224. These two structural evidences confirm that Arg141 is not a key residue for GSK-3β-G12-interaction and, show that the relatively restrained backbone of hinge region offered by Pro136 is critical for G12 binding to GSK-3β. In line with these data, Pro136 is replaced by Gly110 in SAPK2a (Fig 14C) and by Thr318 in ABL (Fig 14D) and in these kinases no inhibitory effect was observed in the presence of G12. Interestingly, Pro136 is replaced by Glu107 also in CDK9. However, in CDK9, since G12 might be shifting upwards to have stacking interaction with Phe103, its interaction with Pro136 could not be necessary, as a new backbone hydrogen bonds might be provided by Cys106, Phe105, or Asp104. Overall, backbone interactions offered by Pro136 seem more helpful in G12 binding to GSK-3β compared to Arg141 side chain interactions. G12 might therefore be not only a good initial hit to be developed as a GSK-3β inhibitor, but also a good starting point for the development of specific inhibitors for CDK9/Cyclin T1, which is already known as a pharmacologically important target[57].

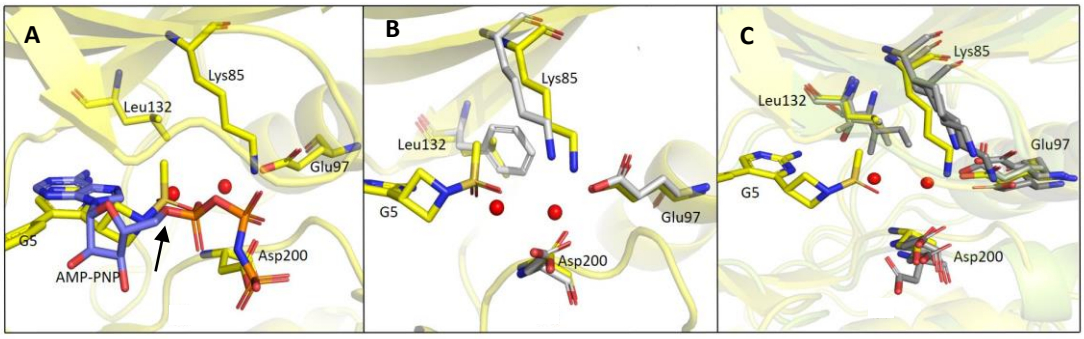


Fig. 13) A) Overlapped structure of GSK3β-G5 complex with GSK3β-AMP complex (PDB 1PYX). B) Overlapped structure of GSK3β-G5 with CDK9 structure (PDB 3BLQ). C) Overlapped structure of GSK3β-G5 with Kinases showing inhibition with G5 (P70S6K/PDB 3A61, SAPK2a/PDB 1OZ1 and LOK1/PDB 6HXF). GSK3β-G5 complex is in yellow, and comparison structures are represented in grey (and blue).

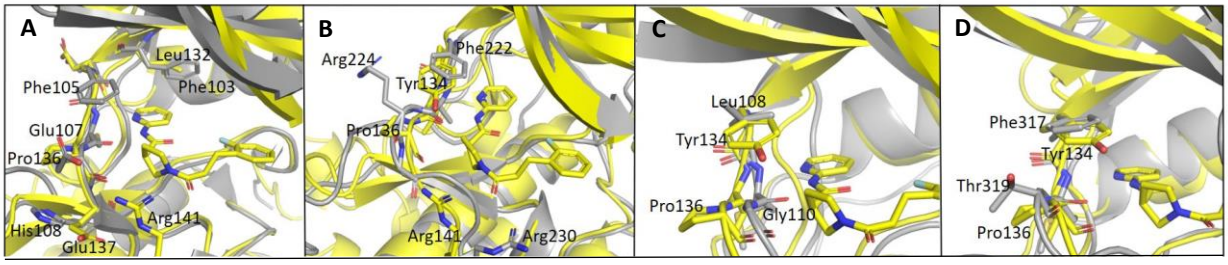


Fig. 14) A) Overlapped structure of CDK9 (PDB 3BLQ) and GSK3ϐ-G12 complex; B) Overlapped structure of GSK3ϐ-G12 and MLK1 (PDB 4UY9) C) Overlapped structure of GSK3ϐ-G12 and SAPK2a (PDB 4EWQ) D) Overlapped structure of GSK3ϐ-G12 and ABL (PDB 4WA9).

To corroborate our structural data we performed an additional computational homology search of our internal IIT database, to identify G12 and G5 homologue compounds. The experimental analysis of the homologues was performed in the attempt of gathering more details about binding affinities and residues contributes and, eventually, identifying more potent compounds. Out of seven identified homologues, two compounds, ARN1484 and ARN9133, showed a similar inhibitory activity compared to G12 and G5 (27.2 µM and 38.1 µM, respectively; Fig. 15). The 3D crystal structures of these compounds in complex with GSK-3β were solved (Fig. 16). Crystal structures of ARN1484/ARN9133-GSK-3β complexes showed a network of interactions for these two compounds with the protein very similar to the ones observed for G12 and G5, corroborating their similar IC_{50} values.

In conclusion, with our study we propose a drug discovery screening pipeline that allowed the selection of promising fragments, inhibitors of GSK-3β, with a binding affinity not higher than a reference binder compound (here, AMP-PNP) (Fig. 17).

The thorough biophysical characterization of the protein-hit interactions, allowed a comprehensive interpretation of compounds inhibitory activity data. The identified fragments are promising hits to be developed into novel more potent and selective GSK-3β inhibitors. Furthermore, the proposed drug discovery screening pipeline has the potential to be generalized to the investigations of other pharmacological target proteins.

The final version of this manuscript was published online on March 31st 2022, on IJMC journal[58].

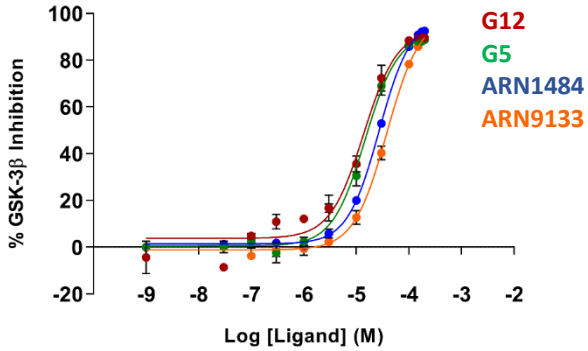


Fig. 15) Activity assay IC_{50} data (LANCE TR-FRET assay) for the two compound analogues (ARN1484 and ARN9133) are close to the data reported for G5 and G12 fragments.

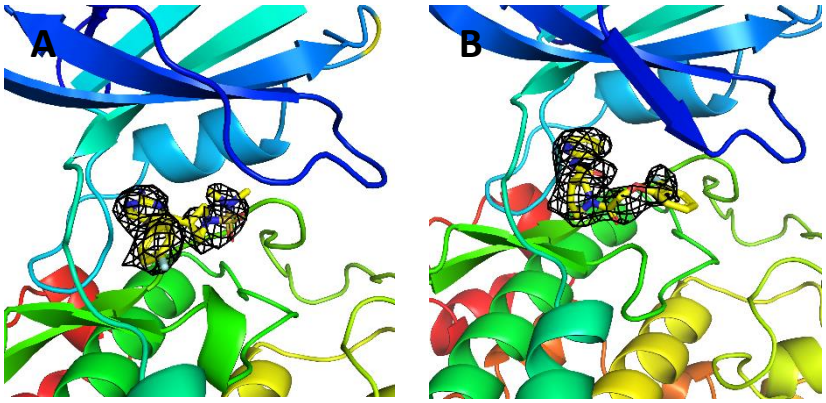


Fig. 16) A) 3D structure of GSK-3β in complex with ARN9133; B) 3D structure of GSK-3β in complex with ARN1484.

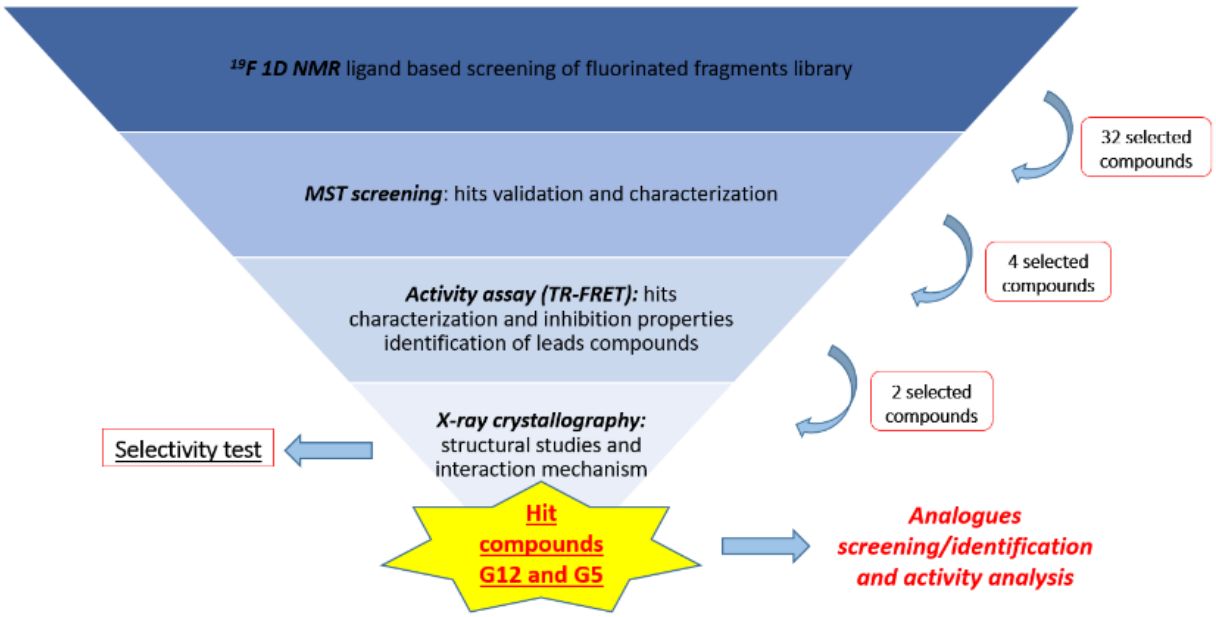


Fig. 17) Drug discovery pipeline schematic picture. The project started with a NMR ligand-based screening of the IIT internal fluorinated fragments library. This first step allowed to make a first compounds skimming. After that, selected hits went through biophysical MST analyses to proceed with a most strict selection, to identify the best GSK-3β binders. Four compounds emerged from this second "cut-off" and went through following activity and structural characterizations. Two out of these four compounds, came out as promising GSK-3β inhibitors, with a good level of selectivity as well. Analogues were analyzed and detected as well.

REFERENCES GSK-3β:

1. Grimes, C. A. & Jope, R. S. The multifaceted roles of glycogen synthase kinase 3β in cellular signaling. *Progress in Neurobiology* (2001) doi:10.1016/S0301-0082(01)00011-9.

2. Liang, M.-H. & Chuang, D.-M. Differential Roles of Glycogen Synthase Kinase-3 Isoforms in the Regulation of Transcriptional Activation. *J Biol Chem* **281**, 30479–30484 (2006).

3. Beurel, E., Grieco, S. F. & Jope, R. S. Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. *Pharmacology and Therapeutics* (2015) doi:10.1016/j.pharmthera.2014.11.016.

4. Pecoraro, C. *et al.* GSK3β as a novel promising target to overcome chemoresistance in pancreatic cancer. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* **58**, 100779 (2021).

5. ter Haar, E. *et al.* Structure of GSK3β reveals a primed phosphorylation mechanism. *Nature Structural Biology* **8**, 593–596 (2001).

6. Dajani, R. *et al.* Crystal structure of glycogen synthase kinase 3β: Structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* (2001) doi:10.1016/S0092-8674(01)00374-9.

7. Wang, Q. M., Park, I. K., Fiol, C. J., Roach, P. J. & DePaoli-Roach, A. A. Isoform Differences in Substrate Recognition by Glycogen Synthase Kinases 3α and 3β in the Phosphorylation of Phosphatase Inhibitor 2. *Biochemistry* (1994) doi:10.1021/bi00167a018.

8. Sutherland, C. What Are the bona fide GSK3 Substrates? *Int J Alzheimers Dis* **2011**, 505607 (2011).

9. Hughes, K., Nikolakaki, E., Plyte, S. E., Totty, N. F. & Woodgett, J. R. Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. *EMBO J* **12**, 803–808 (1993).

10. Kaidanovich-Beilin, O. & Woodgett, J. R. GSK-3: Functional Insights from Cell Biology and Animal Models. *Front Mol Neurosci* **4**, 40 (2011).

11. Linding, R. *et al.* Systematic Discovery of In Vivo Phosphorylation Networks. *Cell* (2007) doi:10.1016/j.cell.2007.05.052.

12. La Pietra, V. *et al.* Design, synthesis, and biological evaluation of 1-phenylpyrazolo[3,4- e]pyrrolo[3,4- G]indolizine-4,6(1 H,5 h)-diones as new glycogen synthase kinase-3β inhibitors. *Journal of Medicinal Chemistry* (2013) doi:10.1021/jm401466v.

13. Arfeen, M. *et al.* Design, synthesis and biological evaluation of 5-benzylidene-2-iminothiazolidin-4-ones as selective GSK-3β inhibitors. *European Journal of Medicinal Chemistry* (2016) doi:10.1016/j.ejmech.2016.04.075.

14. Ye, Q. *et al.* Synthesis and biological evaluation of novel 4-azaindolyl-indolyl-maleimides as glycogen synthase kinase-3β (GSK-3β) inhibitors. *Bioorganic and Medicinal Chemistry* (2009) doi:10.1016/j.bmc.2009.05.031.

15. Noori, M. S. *et al.* Identification of a Novel Selective and Potent Inhibitor of Glycogen Synthase Kinase-3. *American Journal of Physiology-Cell Physiology* (2019) doi:10.1152/ajpcell.00061.2019.

16. Khan, I., Tantray, M. A., Alam, M. S. & Hamid, H. Natural and synthetic bioactive inhibitors of glycogen synthase kinase. *European Journal of Medicinal Chemistry* (2017) doi:10.1016/j.ejmech.2016.09.058.

17. Li, X. *et al.* Lithium Regulates Glycogen Synthase Kinase-3β in Human Peripheral Blood Mononuclear Cells: Implication in the Treatment of Bipolar Disorder. *Biological Psychiatry* (2007) doi:10.1016/j.biopsych.2006.02.027.

18. Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc Natl Acad Sci U S A* (1996) doi:10.1073/pnas.93.16.8455.

19. Jope, R. S., Mines, M. A. & Beurel, E. Regulation of cell survival mechanisms in alzheimer's disease by glycogen synthase kinase-3. *International Journal of Alzheimer's Disease* (2011) doi:10.4061/2011/861072.

20. Jope, R. S., Yuskaitis, C. J. & Beurel, E. Glycogen synthase kinase-3 (GSK3): Inflammation, diseases, and therapeutics. *Neurochemical Research* (2007) doi:10.1007/s11064-006-9128-5.

21. Dewhurst, S., Maggirwar, S. B., Schifitto, G., Gendelman, H. E. & Gelbard, H. A. Glycogen synthase kinase 3 beta (GSK-3β) as a therapeutic target in neuroAIDS. *Journal of Neuroimmune Pharmacology* (2007) doi:10.1007/s11481-006-9051-1.

22. Frame, S. & Zheleva, D. Targeting glycogen synthase kinase-3 in insulin signalling. *Expert Opinion on Therapeutic Targets* (2006) doi:10.1517/14728222.10.3.429.

23. S., A., R.H., B. & G., A. The Possible Involvement of Glycogen Synthase Kinase-3 (GSK-3) in Diabetes, Cancer and Central Nervous System Diseases. *Current Pharmaceutical Design* (2011) doi:10.2174/138161211797052484.

24. Mccubrey, J. A. *et al.* GSK-3 as potential target for therapeutic irvention in cancer. *Oncotarget* (2014) doi:10.18632/oncotarget.2037.

25. *Safety Study of a Glycogen Synthase Kinase 3 (GSK3) Inhibitor in Patients With Alzheimer´s Disease*. (2013) doi:ClinicalTrials.gov Identifier: NCT00948259.

26. *Evaluation of Lithium as a Glycogen-Synthase-Kinase-3 (GSK-3) Inhibitor in Mild Cognitive Impairment*. (2019) doi:ClinicalTrials.gov Identifier: NCT02601859.

27. Smith, D. G. *et al.* 3-Anilino-4-arylmaleimides: Potent and selective inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorganic and Medicinal Chemistry Letters* (2001) doi:10.1016/S0960-894X(00)00721-6.

28. Cross, D. A. E. *et al.* Selective small-molecule inhibitors of glycogen synthase kinase-3 activity protect primary neurones from death. *Journal of Neurochemistry* (2001) doi:10.1046/j.1471-4159.2001.t01-1-00251.x.

29. Leclerc, S. *et al.* Indirubins Inhibit Glycogen Synthase Kinase-3β and CDK5/P25, Two Protein Kinases Involved in Abnormal Tau Phosphorylation in Alzheimer's Disease. *Journal of Biological Chemistry* (2001) doi:10.1074/jbc.m002466200.

30. Berg, S. *et al.* Discovery of novel potent and highly selective glycogen synthase kinase-3β (GSK3β) inhibitors for Alzheimers Disease: Design, synthesis, and characterization of pyrazines. *Journal of Medicinal Chemistry* (2012) doi:10.1021/jm201724m.

31. Pandey, M. K. & DeGrado, T. R. Glycogen synthase kinase-3 (GSK-3)-targeted therapy and imaging. *Theranostics* (2016) doi:10.7150/thno.14334.

32. Bertrand, J. A. *et al.* Structural characterization of the GSK-3β active site using selective and non-selective ATP-mimetic inhibitors. *Journal of Molecular Biology* (2003) doi:10.1016/j.jmb.2003.08.031.

33. Vulpetti, A., Hommel, U., Landrum, G., Lewis, R. & Dalvit, C. Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *J Am Chem Soc* (2009) doi:10.1021/ja905207t.

34. Dalvit, C. NMR methods in fragment screening: theory and a comparison with other biophysical techniques. *Drug Discov Today* **14**, 1051–1057 (2009).

35. Vulpetti, A. & Dalvit, C. Fluorine local environment: From screening to drug design. *Drug Discovery Today* (2012) doi:10.1016/j.drudis.2012.03.014.

36. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science (1979)* (1996) doi:10.1126/science.274.5292.1531.

37. Jahnke, W. & Erlanson, D. A. *Fragment-based Approaches in Drug Discovery*. *Fragment-based Approaches in Drug Discovery* (2006). doi:10.1002/3527608761.

38. Dalvit, C. *et al.* High-throughput NMR-based screening with competition binding experiments. *J Am Chem Soc* (2002) doi:10.1021/ja020174b.

39. Jahnke, W. Perspectives of biomolecular NMR in drug discovery: The blessing and curse of versatility. *Journal of Biomolecular NMR* (2007) doi:10.1007/s10858-007-9183-5.

40. di Fruscia, P. *et al.* Fragment-Based Discovery of Novel Allosteric MEK1 Binders. *ACS Medicinal Chemistry Letters* **12**, 302–308 (2021).

41. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments* (1958) doi:10.1063/1.1716296.

42. Carr, H. Y. & Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review* (1954) doi:10.1103/PhysRev.94.630.

43. Rich, R. L., Errey, J., Marshall, F. & Myszka, D. G. Biacore analysis with stabilized G-protein-coupled receptors. *Anal Biochem* **409**, 267–272 (2011).

44. Huynh, K. & Partch, C. L. Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* (2015) doi:10.1002/0471140864.ps2809s79.

45. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* (2011) doi:10.1107/S0907444910048675.

46. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* (2013) doi:10.1107/S0907444913000061.

47. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* (1997) doi:10.1107/S0907444996012255.

48. Liebschner, D. *et al.* Polder maps: Improving OMIT maps by excluding bulk solvent. *Acta Crystallographica Section D: Structural Biology* **73**, 148–157 (2017).

49. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* (2010) doi:10.1107/S0907444910007493.

50. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* (2002).

51. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology* (2011) doi:10.1038/nbt.1990.

52. Bregman, H., Williams, D. S., Atilla, G. E., Carroll, P. J. & Meggers, E. An organometallic inhibitor for glycogen synthase kinase 3. *J Am Chem Soc* (2004) doi:10.1021/ja046049c.

53. Dalvit, C., Caronni, D., Mongelli, N., Veronesi, M. & Vulpetti, A. NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Current Drug Discovery Technologies* vol. 3 115–124 (2006).

54. Singh, M., Tam, B. & Akabayov, B. NMR-Fragment Based Virtual Screening: A Brief Overview. *Molecules* **23**, (2018).

55. Kramer, T., Schmidt, B. & Lo Monte, F. Small-molecule inhibitors of GSK-3: Structural insights and their application to Alzheimer's disease models. *International Journal of Alzheimer's Disease* (2012) doi:10.1155/2012/381029.

56. Eldar-Finkelman, H. & Martinez, A. GSK-3 Inhibitors: Preclinical and Clinical Focus on CNS. *Frontiers in Molecular Neuroscience* (2011) doi:10.3389/fnmol.2011.00032.

57. Mandal, R., Becker, S. & Strebhardt, K. Targeting CDK9 for Anti-Cancer Therapeutics. *Cancers* vol. 13 (2021).

58. Balboni, B. *et al.* Identification of Novel GSK-3β Hits Using Competitive Biophysical Assays. *International Journal of Molecular Sciences* **23**, 3856 (2022).