

**Computer-Aided Drug Design of DNA Repair Inhibitors Targeting the  
ERCC1-XPF Endonuclease**

by

Francesco Gentile

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics

University of Alberta

© Francesco Gentile, 2019

## Abstract

Computational methods are nowadays essential tools employed in the field of drug design. Indeed, bringing a molecular compound from bench to clinic is a challenging, time-consuming and expensive goal to achieve. Computer-aided drug design (CADD) techniques aim to simplify this process, by *in-silico* identifying compounds active against a critical macromolecular target. With the knowledge of the molecular structure of the compounds and preferably also of the target of interest, virtual screening can produce a high-confidence, limited set of small molecules to be tested experimentally, reducing enormously both the time and the cost otherwise required for the *in-vitro* screening. Computer simulations of ligand-receptor complexes provide also structural insights of the interactions which can rationally drive the hit-to-lead optimization process.

In this thesis, CADD techniques are used to identify and optimize DNA repair inhibitors. Although DNA repair mechanisms are essential to maintain genome integrity in normal cells, they counteract the action of many DNA-damaging therapies used in cancer. Hence a drug able to inhibit DNA repair in cancer cells could be used in combination with these therapies to enhance their effect and reduce drug resistance effect. The nucleotide excision repair (NER) pathway, responsible for repairing bulky DNA damages, is the main focus of this thesis. The excision repair cross-complementation group 1 (ERCC1)-xeroderma pigmentosum, complementation group F (XPF) heterodimer is a critical NER complex involved in repairing platinum-based chemotherapy and radiation-induced DNA damages. Hence, inhibiting the action of ERCC1-XPF, through disrupting essential protein-protein interactions or blocking the

catalytic activity, is one of the most promising strategies to enhance the effects of these therapeutic approaches.

Therefore, the CADD effort discussed in this work has been directed towards three different aspects of ERCC1-XPF: first, the interaction between ERCC1 and the xeroderma pigmentosum, complementation group A (XPA) protein, essential to recruit the repair machinery to the damaged DNA zone. Second, the catalytic activity, performed by the active site present on the XPF surface. Last, the ERCC1-XPF heterodimerization, required to have a functional endonuclease.

The results reported in this thesis are of great importance for the field of combination cancer therapy. On one hand, this work presents a number of new hit or lead compounds able to inhibit DNA repair. Thanks to the collaborative environment of the Alberta DNA Repair Consortium (of which this work is part), many of these small molecules are currently under optimization and experimental testing with the final aim to translate them to clinical use. On the other hand, detailed biophysical and structural characterization of these compounds and their binding to the targets are covered as well, opening new venues for future rational drug design works.

## Preface

The work presented here was completed at the Department of Physics of the University of Alberta. This thesis is based on material that was previously published.

Chapter 2 is part of a book chapter which has been published as Preto, J.; Gentile, F.; Winter, P.; Churchill, C.; Omar, S. I.; Tuszynski, J. A. *Molecular Dynamics and Related Computational Methods with Applications to Drug Discovery*. In *Coupled Mathematical Models for Physical and Biological Nanoscale Systems and Their Applications*; Bonilla L.L., Kaxiras E., Melnik R., Eds.; Springer, Cham, 2018; pp. 267–285. I and Preto J. systematically reviewed the state-of-the-art of computational methods commonly used in drug design and wrote the manuscript. Winter P., Churchill C. and Omar S.I. contributed with the editing of the manuscript. Tuszynski J.A. was the supervisory author and was responsible for manuscript composition and editing. In this thesis, I included only the sections of the book chapter which I primarily wrote by myself.

Chapter 3 is part of a review article published as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. *Modelling DNA Repair Pathways: Recent Advances and Future Directions*. *Curr. Pharm. Des.* 2016, 22, 3527–3546. I entirely conducted the bibliographical review of the current status of modeling techniques for DNA repair pathways and I drafted the manuscript. Tuszynski J.A. and Barakat K.H. were the supervisory authors, they conceived the review idea and were involved in manuscript composition and editing.

Chapter 4 was published as research article as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. *New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy*. *J. Mol. Graph. Model.* 2016, 1, 71–82. I was responsible for the project design, I performed all the simulations, analyzed the results and drafted the manuscript. Tuszynski J.A. and Barakat K.H.

were the supervisory authors, and were involved with project design and manuscript composition and editing. Supplementary materials are available in Appendix A. A commentary of the manuscript has been published as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. Commentary: New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy. *J Cancer Treat Diagnosis* 2017, 1, 1–3. I was responsible for reviewing the original research article and drafting the manuscript. Tuszynski J.A. and Barakat K.H. were the supervisory authors, and were involved with manuscript composition and editing. The commentary is available in Appendix C.

Chapter 5 has been published as research article as Gentile, F.; Barakat, K. H.; Tuszynski, J. A. Computational Characterization of Small Molecules Binding to the Human XPF Active Site and Virtual Screening to Identify Potential New DNA Repair Inhibitors Targeting the ERCC1-XPF Endonuclease. *Int. J. Mol. Sci.* 2018, 19, 1328. I was responsible for the project design, I performed all the simulations, analyzed the results and drafted the manuscript. Tuszynski J.A. and Barakat K.H. were the supervisory authors, and were involved in project design and manuscript composition and editing. Supplementary materials are available in Appendix A.

Chapter 6 is under consideration for publication as Elmenoufy A.H.; Gentile F.; Jay D.; Karimi-Busheri F.; Yang X.; Soueidan O.M.; Weilbeer C.; Mani R.S.; Barakat K.H.; Tuszynski J.A.; Weinfeld M.; West F.G. Targeting DNA Repair in Tumor Cells via Inhibition of ERCC1-XPF in *ACS Cent. Sci.* I was responsible for the whole computational part of the project, I performed all the simulations and analyzed the results. Elmenoufy A.H., Soueidan O.M. and Weilbeer C. worked on the chemical synthesis of new analogues. Jay D., Karimi-Busheri F., Yang X. and Mani R.S. carried out the biological characterization of the compounds. All these authors contributed to draft the manuscript. Barakat K.H., Tuszynski J.A., Weinfeld M. and West F.G.

were the supervisory authors, designed the project and contributed to manuscript editing. Supplementary materials are available in Appendix B.

Appendix D has been published as Gentile, F.; Deriu, M. A.; Barakat, K.; Danani, A.; Tuszynski, J. A Novel Interaction Between the TLR7 and a Colchicine Derivative Revealed Through a Computational and Experimental Study. *Pharmaceuticals* 2018, *11*, 22. I was responsible for the project design, I performed all the simulations, analyzed the results and drafted the manuscript. Deriu M.A. assisted in experimental design, as well as in manuscript composition and editing. Tuszynski J. designed the study. Barakat K., Danani A. and Tuszynski J. were the supervisory authors and were involved in manuscript composition and editing.

*To my family, for their love and support*

## Acknowledgments

My greatest thanks go to my supervisors, Jack Tuszynski and Khaled Barakat, who have been always incredible supportive and helpful during my Ph.D. program. Their research attitude, their knowledge and the respect for their trainees are the best qualities a student seeks in his mentors. I would also like to thank my supervisory committee members, Mariusz Klobukowski and Robert Wolkow, for their guidance in research and help with award applications. I want to thank also all my friends and colleagues of the Tuszynski lab, Jordane Preto, Holly Freeman, Philip Winter, Sara Omar, Sahar Arbabimoghadam, Mahshad Moshari, Aarat Kalra, Douglas Friesen and Niloofar Nayebi. They have supported me during all the four years of the program, and I could not have asked for a more friendly, supportive and enjoyable environment where to conduct my research. I would also like to acknowledge the financial support I have received from Alberta Innovates, Alberta DNA Repair Consortium, Canadian Cancer Society, Biophysical Society, the University of Alberta Graduate Students' Association and Novartis Pharmaceuticals Canada Inc, which gave me the opportunity to conduct my research and travel to several conferences. I would also like to thank the people who have been constantly with me during my Ph.D., during the dark and the bright days: my family; my friends Alberto, Marco, Davide and Federico; and my girlfriend Pei and her wonderful family. Thanks for your unconditioned and persistent encouragement to pursue my life dreams.

## Table of Contents

<b>Chapter 1 - Introduction .....</b>	<b>1</b>
<b>Chapter 2 - Molecular Dynamics and Related Computational Methods with Applications to Drug Discovery .....</b>	<b>4</b>
<b>2.1 Abstract.....</b>	<b>4</b>
<b>2.2 Introduction.....</b>	<b>4</b>
<b>2.3 Molecular Dynamics .....</b>	<b>5</b>
<i>2.3.1 General .....</i>	<i>5</i>
<i>2.3.2 Molecular Dynamics and Drug Discovery .....</i>	<i>9</i>
<b>2.4 Other Computational Drug-Discovery Methods.....</b>	<b>11</b>
<i>2.4.1 Binding Pocket Prediction .....</i>	<i>11</i>
<i>2.4.2 Ligand-Receptor Docking.....</i>	<i>15</i>
<i>2.4.3 Virtual Screening .....</i>	<i>20</i>
<b>2.5 Conclusions.....</b>	<b>23</b>
<b>Chapter 3 - Modeling NER Repair Pathway: Recent Advances and Future Directions .....</b>	<b>25</b>

<b>3.1 Abstract.....</b>	<b>25</b>
<b>3.2 Introduction.....</b>	<b>26</b>
<b>3.3 DNA Repair Pathways and Resistance to Cancer Treatment .....</b>	<b>29</b>
<i>3.3.1 Nucleotide Excision Repair .....</i>	<i>31</i>
<b>3.4 Modeling NER Repair Mechanisms at the Atomistic Level .....</b>	<b>33</b>
<b>3.5 Computer-Aided Drug Design for NER Repair Modulators.....</b>	<b>41</b>
<b>3.6 Conclusions.....</b>	<b>45</b>
 <b>Chapter 4 - New Design of NER Inhibitors for Combination Cancer Therapy</b>	
<b>.....</b>	<b>48</b>
<b>4.1 Abstract.....</b>	<b>48</b>
<b>4.2 Introduction.....</b>	<b>49</b>
<b>4.3 Results and Discussion.....</b>	<b>51</b>
<i>4.3.1 Similarity Search-Based Virtual Screening .....</i>	<i>51</i>
<i>4.3.2 Pharmacophore-Based Virtual Screening .....</i>	<i>52</i>
<i>4.3.3 Molecular Docking .....</i>	<i>54</i>
<i>4.3.4 Molecular Dynamics Simulations and MM/PBSA Rescoring.....</i>	<i>57</i>
<b>4.4 Conclusion .....</b>	<b>66</b>

<b>4.5 Materials and Method .....</b>	<b>68</b>
4.5.1 <i>Similarity Search.....</i>	68
4.5.2 <i>Ligand Preparation.....</i>	69
4.5.3 <i>Pharmacophore-Based Virtual Screening.....</i>	69
4.5.4 <i>Target Preparation .....</i>	70
4.5.5 <i>Molecular Docking Simulations .....</i>	71
4.5.6 <i>Molecular Dynamics Simulations .....</i>	72
4.5.7 <i>MM/PBSA .....</i>	73
4.5.8 <i>Visual Analysis.....</i>	75
<b>Chapter 5 - Computational Characterization of Small Molecules Binding to the Human XPF Active Site and Virtual Screening to Identify Potential New DNA Repair Inhibitors Targeting the ERCC1-XPF Endonuclease .....</b>	<b>76</b>
<b>5.1 Abstract.....</b>	<b>76</b>
<b>5.2 Introduction.....</b>	<b>77</b>
<b>5.3 Results and Discussion.....</b>	<b>80</b>
5.3.1 <i>Homology Modeling and Molecular Dynamics Simulation .....</i>	80
5.3.2 <i>Modeling of Small Molecules Binding to the Human XPF Active Site.....</i>	84

5.3.3 Virtual Screening .....	86
<b>5.4 Materials and Methods.....</b>	<b>88</b>
5.4.1 Homology Modeling of the Human XPF Nuclease Domain.....	88
5.4.2 Molecular Dynamics Simulation and Clustering of the Trajectory.....	89
5.4.3 Molecular Docking of Known Inhibitors and Pharmacophore Modeling.....	90
5.4.4 Virtual Screening of the ZINC15 Database.....	93
<b>5.5 Conclusions.....</b>	<b>94</b>
<b>Chapter 6 – Novel ERCC1-XPF Inhibitors for Resistant Colorectal Cancer Treatment Based on Targeting DNA Repair .....</b>	<b>96</b>
<b>6.1 Abstract.....</b>	<b>96</b>
<b>6.2 Introduction.....</b>	<b>97</b>
<b>6.3 Materials and Methods.....</b>	<b>99</b>
6.3.1 Molecular Docking of F06.....	99
6.3.2 Characterization of the Binding Pocket and Pharmacophore Modeling .....	100
6.3.3 Docking-Based Virtual Screening of F06 Analogues .....	100
6.3.4 Molecular Dynamics Simulations and MM/GBSA Rescoring of the Analogues .....	101
6.3.5 General Experimental Procedures for Top F06-Based Analogues Synthesis .....	102

6.3.6 ERCC1-XPF Protein Preparation .....	102
6.3.7 Microplate Fluorescence Incision Assay .....	103
6.3.8 Steady-State Fluorescence Assays .....	103
6.3.9 Cell Culture.....	104
6.3.10 Cellular Repair of Cyclobutane Pyrimidine Dimers .....	104
6.3.11 Clonogenic Survival Assay .....	105
<b>6.4 Results and Discussion.....</b>	<b>106</b>
6.4.1 Molecular Docking of F06.....	106
6.4.2 Virtual Screening of F06 Analogues.....	107
6.4.3 Synthesis of F06-based analogues .....	108
6.4.4 Inhibition of ERCC1-XPF Endonuclease Activity.....	111
6.4.5 Inhibitor Binding to ERCC1-XPF.....	113
6.4.6 Inhibition of Cellular Repair of Cyclobutane Pyrimidine Dimers .....	115
6.4.7 Increased Sensitization to UV Radiation and Cyclophosphamide .....	116
<b>6.5 Conclusion .....</b>	<b>117</b>
<b>Chapter 7 - Conclusions and Future Works .....</b>	<b>119</b>

<b>7.1 Identification of XPA-ERCC1 Inhibitors Using Virtual Screening .....</b>	<b>119</b>
<b>7.2 Exploring the ERCC1-XPF Active Site as Potential Target for Inhibitors.....</b>	<b>121</b>
<b>7.3 Targeting the ERCC1-XPF Heterodimerization Interface with Small Molecules to Disrupt the Functional Complex .....</b>	<b>122</b>
<b>Bibliography .....</b>	<b>124</b>
<b>Appendix A - Supplementary Materials of Chapter 4 and 5.....</b>	<b>158</b>
<b>A.1 Chapter 4 .....</b>	<b>158</b>
<b>A.2 Chapter 5 .....</b>	<b>159</b>
<b>Appendix B – Supplementary Materials to Chapter 6.....</b>	<b>168</b>
<b>B.1 Materials and Methods .....</b>	<b>168</b>
<i>B.1.1 Molecular Docking of F06 .....</i>	<i>168</i>
<i>B.1.2 MD Simulations and MM/GBSA Rescoring of the Analogues .....</i>	<i>169</i>
<i>B.1.3 Synthesis of F06-Based Analogues.....</i>	<i>171</i>
<i>B.1.4 Cell Proliferation Assay .....</i>	<i>171</i>
<b>B.2 Results .....</b>	<b>172</b>
<i>B.2.1 Virtual Screening of F06 Analogues .....</i>	<i>172</i>

<i>B.2.2 Synthesis and Characterization of Inhibitors (1-8)</i> .....	175
<i>B.2.3 Cytotoxicity Profile of Compound 4</i> .....	181
<b>Appendix C - Commentary of Chapter 4</b> .....	<b>183</b>
<b>Appendix D - A Novel Interaction Between the TLR7 and a Colchicine Derivative Revealed Through a Computational and Experimental Study ....</b>	<b>189</b>
<b>D.1 Abstract</b> .....	<b>189</b>
<b>D.2 Introduction</b> .....	<b>189</b>
<b>D.3 Results and Discussion</b> .....	<b>193</b>
<i>D.3.1 Homology Modeling</i> .....	193
<i>D.3.2 Molecular Docking</i> .....	194
<i>D.3.3 Molecular Dynamics Simulations and MM/GBSA Calculations</i> .....	197
<i>D.3.4 Experiments</i> .....	202
<b>D.4 Materials and Methods</b> .....	<b>202</b>
<i>D.4.1 Model Building and Validation</i> .....	203
<i>D.4.2 Ligand Preparation and Molecular Docking</i> .....	204
<i>D.4.3 Molecular Dynamics Simulations</i> .....	204

<i>D.4.4 MM/GBSA Binding Energy Calculations</i> .....	205
<i>D.4.5 Cell Preparation</i> .....	207
<i>D.4.6 Compound Preparation</i> .....	207
<i>D.4.7 Cell Proliferation Assay</i> .....	207
<i>D.4.8 Data Analysis</i> .....	208
<i>D.4.9 Curve Fitting</i> .....	208
<i>D.4.10 Cell Genetics</i> .....	209
<i>D.4.11 Drug Sensitivity</i> .....	209
<b>D.5 Conclusions</b> .....	<b>210</b>
<b>D.6 Supplementary Materials</b> .....	<b>212</b>

## List of Tables

Table 3.1. <i>In-silico</i> discovered NER protein inhibitors. The details about the experimentally measured activities are reported as well. If not indicated, IC <sub>50</sub> values refer to the best obtained in the related experiment. See text and referred papers for more details.....	44
Table 4.1 MM/PBSA binding energies of the top fifteen hits.....	62
Table 5.1. Results from different substitution matrices available in MOE 2013 for the detection of homologs of the human XPF nuclease domain. Proteins are reported with their PDB ID. Top templates were identified as sequences with a E-value lower than the E-value acceptance (1e-12). Others templates showed a E-value between E-Value acceptance and E-value cutoff (10), and a Z-score of at least 6.....	80
Table 6.1. F06-based analogues functionalized with different R substituents.....	110
Table A.1. Conformational entropy changes (TΔS), calculated using normal mode analysis for a set of diverse hit structures. The ligands were selected to possess different features (for example the number of rings and flexible bonds). The unfavorable contributions oscillate between ~20 and 24 kcal/mol and they overlap within the standard deviations. The calculations were run on 100 snapshots extracted from the last 4 ns of 10 ns-long molecular dynamics simulations of each complex.....	158
Table A.2. Composition of the 10 clusters resulting from clustering of the last 106 ns of MD trajectory. The ratio is calculated as number of frames in the cluster divided by the total number of frames (10601). Selected refers to the representative structure of a cluster being included in the molecular docking simulation as target or not.....	161
Table B.1. Pharmacophore features built from the docking pose of F06 to the XPF site.....	174
Table B.2. Computational results for the subset of analogues chosen for synthesis. The ligand efficiency was calculated using the GBVI/WSA MOE score divided for the number of heavy atoms of each molecule. The analogues are ranked based on their MM/GBSA score which includes the entropy contributions. The lead compound F06 is reported as the last entry.....	174
Table B.3. Half-maximum inhibitory concentrations for the compounds with the highest inhibitory potentials. The data was obtained from at least three different experiments of V <sub>o</sub> versus compound concentration.....	180
Table C.1. Top eighteen hits of the VS experiment, scored according to their MM/PBSA binding energy. All the compounds showed better values of the energy than the lead structures (rank 19 and 20). In total, seventy-two molecules showed lower binding energy values. The ID column indicates the ID associated to the compound in its original database. The MM/PBSA energies do not include the entropy contribution. The VS method column indicates the ligand-based screening technique used to include the compound in the target-based VS step from the indicated database (last column).....	187

Table D.1. Results from the MM/GBSA calculations over the last 5 ns of MD production trajectories of the R-837-TLR7 and CR42-24-TLR7 complexes. The total binding energies of the two compounds are in a comparable range, with the more favorable van der Waals and electrostatic contributions of the CR42-24-TLR7 complex being partially compensated by a larger, unfavorable polar desolvation term, when compared to R-837. The total values reported here do not include the entropic terms. The value of the standard deviation is reported in brackets for each term. .... 199

Table D.2. Cell lines in the Oncolines™ panel used in this study and their genetic status. Indicated in the second column are the genes that are either mutated, part of a translocation or that have altered copy numbers (*italic*) in that cell line. .... 215

## List of Figures

- Figure 3.1. Steps of human global-genome and transcription-coupled NER pathways. The recognition step is different for the two pathways. In transcription-coupled NER, a stalled polymerase at the damaged site is the trigger for the repair activity. The polymerase recruits then the CSA and CSB proteins. In global-genome NER the XPC-RAD23B complex recognizes instead the lesion and starts the repair action. The subsequent steps are common to both the pathways. See text for more details. Adapted from KEGG database (<http://www.genome.jp/kegg/>) [125,126]. ..... 32
- Figure 3.2. Simulation of binding mechanism between ssDNA and RPA. The CG model for RPA-ssDNA interaction not only permits to reproduce the crystallographic complex, but also to simulate the binding process, where DNA gradually interacts with four sites through a sliding process. The approaching of the DNA with each binding site is represented on the left, where the purple line indicates the interacting nucleotide index and the black one the difference between the site-nucleotide minimal distance in the model and the crystal structure. A-D refers to binding site I-IV. The four snapshots occurred at the simulation times indicated by the respective arrow. Adapted with permission from [141]. ..... 37
- Figure 3.3. Superposition of predicted XPA binding motif and NMR structure of bound XPA peptide. The representative structures of the most populated clusters of the XPA peptide, obtained over 10  $\mu$ s seconds of MD simulations of the XPA free form, are represented in cyan and superposed to the NMR bound structure of XPA peptide (red). ERCC1 central domain is represented in grey. The similarities between the hairpin motif and the experimental bound one provided evidences of this structure being responsible for XPA recognition by ERCC1 in Fadda's work. Adapted with permission from [147]. ..... 40
- Figure 3.4. Binding mode and chemical structure of NSC-130813. This small molecule binds to XPF, inhibiting the XPF-ERCC1 protein-protein interaction. This compound has been discovered with a structure-based VS protocol targeting the XPF C-terminal by Jordheim et al. Adapted with permission from [84]. ..... 45
- Figure 4.1. Chemical structures of NERI01 (left) and compound 10 (right). The two compounds share similar chemical features, in particular a hydrophobic backbone with two nitro groups at the extremities. .... 52
- Figure 4.2. The ligand/structure-based pharmacophore model built in MOE. ACC means acceptor feature, HYD means hydrophobic feature and ARO means aromatic feature. The radius and the tolerance of the features have been adjusted in order to satisfy the receptor-based electrostatic maps, calculated in MOE (see text for more details). ..... 54
- Figure 4.3. Superposition of the NMR configurations of ERCC1 key residues. This set of ten residues is mainly responsible of the interaction with the XPA peptide. The residues are colored according to their names. .... 55

Figure 4.4. Binding poses of NERI01 and compound 10 in the binding site of structure #7. The polar extremities of the two compounds are interacting with the Arg residues delimiting the binding site (Arg106 and Arg144). The hydrophobic backbone is stacked between the aromatic layers constituted by the hydrophobic residues of the binding pocket and Tyr145, 152. The surface of the binding pocket is colored depending on the type of residues, in white where hydrophobic, in green where polar, in blue where basic and in red where acidic. .... 57

Figure 4.5. Predicted binding poses of A) NERI01 and B) 6971912. Hydrogen bond networks are also represented. The nitro/positive groups can interact with the positive charged parts of the Arg residues. The hydrogen bond acceptors of the ligands interact with Ser142 or Gln107. .... 59

Figure 4.6. Chemical structures of the top fifteen hits derived from the turbo similarity search. The compounds have been ranked using the MM-PBSA binding energies. The PubChem Compound ID is reported below each structure. .... 61

Figure 4.7. Heat maps of the A) total B) van der Waals and C) electrostatic contributions of the ERCC1 residues to the binding energies. NERI01 and c10 are the two lead structures. Compound names followed by (ph) are derived from the pharmacophore-based virtual screening. The other ligands are indicated with the PubChem Compound ID. A clear binding pattern can be observed for the total energy, with Tyr145, Arg106, Gln107 and Phe140 being the highest and most conserved contributors. The pattern is very similar for the van der Waals contributions, indicating the dominating role of hydrophobic interactions due to the stacking of ligand rings. Consistently with the instabilities observed during the molecular dynamics simulations, it is not possible to observe a well-defined pattern for the electrostatic interactions, although Arg106 is the main partner for the major part of the compounds. Asp129 contribution is often unfavorable, due to the electrostatic interference with the negatively charged atoms of the ligands. .... 63

Figure 4.8. RMSD trend of NERI01 and 6971912 during the 2 ns of production simulation. Only the heavy atoms of the compounds have been considered. 6971912 shows a smaller displacement from the original docking pose during the simulation, but also higher, still stable fluctuations. . 65

Figure 4.9. Binding poses and hydrogen bond networks of the new scaffolds from the NCI sets, A) ligand 1, B) ligand 80 and C) ligand 63. These three compounds are the top hits identified through the pharmacophore-based screening and represent promising scaffolds to be optimized. The hydrophobic backbones are well-stacked within the hydrophobic cleft of the ERCC1 central domain. The polar extremities can form hydrogen bonds with the Arg residues, although these interactions were generally unstable during the simulations. .... 66

Figure 5.1. Structural superposition of the four top templates for the human XPF nuclease, 2BGW (blue, XPF from *Aeropyrum pernix*), 2BHN (yellow, XPF from *Aeropyrum pernix*), 1J22 (purple, HeF from *Pyrococcus furiosus*) and 1J25 (green, HeF from *Pyrococcus furiosus*). .... 82

Figure 5.2. RMSD trends for the backbone atoms of the human XPF nuclease domain (red) and the active site (blue), defined as residues D687, R689, E690, D715, E725, R726, K727 and D731. .... 83

Figure 5.3. Superposition of the seven conformations of the human XPF active site obtained by clustering the MD trajectory and including the lowest potential energy conformation as well. The

set of conformations was then used as target for the molecular docking simulations. Red spheres indicate the positions of the  $Mg^{2+}$  ion present in the active site and coordinated by D715, E725 and R726. .... 84

Figure 5.4. Pharmacophore model designed by considering the predicted binding poses of the known XPF inhibitors. Three features were present for all the ligand-protein complexes, namely the anionic one (Ani, in green) in proximity of the  $Mg^{2+}$  ion, the donor projection (Don2, in blue) close to E712, and the acceptor projection (Acc2, in gold) close to K727. Two docked ligands satisfying the pharmacophore model are also reported (13 [205] in dark grey, 33 [204] in light grey). All the conformations of the active site extracted from clustering the MD trajectory are represented. .... 86

Figure 5.5. Binding modes of two promising hits derived from the VS. A) ZINC000049131978, hit #5 B) ZINC000038550857, hit #13. Hydrogen bonds are represented in purple dotted lines. The surface of the active site is colored depending on the electrostatic potential, ranging from -40 (intense red) to +40 (intense blue). See text for a detailed description of the electrostatic interactions. .... 87

Figure 5.6. Chemical structures of the ERCC1-XPF inhibitors used to build the pharmacophore model after molecular docking simulations targeting the human XPF nuclease domain were run. All the small molecules investigated in this step included at least one metal-binding motif. .... 92

Figure 6.1. Analysis of the binding features of the top hits. A) Per-residue decomposition of ligand-receptor binding energy. Compound 4 showed a slightly different pattern of total interaction energy, while the lead compound and compound 3 had more similar patterns. Blue cells indicate favorable interactions, while red cells indicate unfavorable interactions. B) Lowest potential energy snapshot extracted from the MD simulation of compound 4 bound to XPF. Hydrogen bonds are colored in purple. The highest pocket contributors in terms of total interaction energy are labelled. C) Interaction diagram for compound 4 and surrounding residues and water molecules. K860 was a strong contributor to the binding energy due to arene-H interaction with one of the aromatic rings of compound 4. The diagram refers to the lowest potential energy snapshot extracted from the MD simulation. .... 109

Figure 6.2. A microplate fluorescent assay to measure inhibition of ERCC1-XPF endonuclease activity by the different compounds. Incubation of the stem-loop substrate with ERCC1-XPF resulted in the release of a fluorescent 8-base fragment. A) Shows the increase in fluorescence (RFU) with time. A representative tracing of the effect of the different compounds (8  $\mu$ M each) on the incision activity is shown. B) (Inset) shows a representative plot of enzyme rate ( $\Delta$ RFU/time) vs compound (4) concentration. .... 112

Figure 6.3. Fluorescence spectroscopy. A) Fluorescence of ERCC1-XPF (13 nM and 20 nM) incubated with compound 5 (negative control), and B) compound 4 (active inhibitor) respectively. C) Unimodal binding pattern and the binding affinity of compound 4 with ERCC1-XPF (70 nM). .... 114

Figure 6.4. A) Immunofluorescence images of the UV-based assay for detecting CPDs in cells treated with compounds 4 and 5. B) Normalized fluorescence intensity of the treated cells. .... 115

Figure 6.5. A) Survival of HCT-116 cells exposed to increasing doses of 254 nm UV radiation and treated with 1 and 2  $\mu$ M compound 4 determined by the clonogenic survival assay. D) Survival of HCT-116 cells exposed to increasing doses of cyclophosphamide and treated with 1 and 2  $\mu$ M compound 4 determined by the clonogenic survival assay. .... 116

Figure A.1. Alignment of the human XPF nuclease domain sequence and all the sequences identified as templates by the template search step in MOE, using the BLOSUM62 matrix. Conserved residues are highlighted in green. .... 159

Figure A.2. Alignment of the amino acid sequences of the human XPF nuclease domain, 2BGW and 1J25. Fully conserved residues are reported in the Consensus line in bold. Partially conserved residues are reported in the Consensus line in normal font. Residue columns are colored based on their degree of conservation when compared with the XPF nuclease sequence, ranging from red (not conserved) to blue (fully conserved). .... 159

Figure A.3. A) DBI (blue) and SSR/SST ratio (purple) and B) pSF (green) trends obtained by varying the number of clusters of human XPF active site conformations. At 10 clusters (black vertical line), we observed a local minimum of the DBI, a kink in the SSR/SST curve and a maximum of the pSF, indicating optimal cluster counting. .... 160

Figure A.4. Chemical structures, ZINC IDs and London dG scores (in kcal/mol) of the top fifty hits obtained by screening the ZINC15 compound database. .... 167

Figure B.1. A) One-pot sequential addition reaction for synthesis of compounds **1-8**. B) Main steps and intermediates included in the one pot sequential addition reaction. .... 171

Figure B.2. Visual analysis of the F06 docking pose. A) Shape complementarity between the docked pose of F06 and the XPF binding site. The compound was docked to the XPF pocket that wraps around residue Phe293 of ERCC1 in the dimerization complex. B) Non-bonded interactions. Observed hydrogen bonds are represented in purple. Residues involved in hydrogen bonding or hydrophobic interactions with the ligand are explicitly represented. See text for more details. C) Affinity mapping of the F06 binding site on the XPF HhH2 domain. Blue color indicates zones of interaction where a donor probe atom shows a potential of at least -2 kcal/mol, therefore where a ligand donor atom will contribute favourably to the free energy of binding. Red indicates acceptor favorable zones, and white hydrophobic favorable zones. D) Pharmacophore model designed for the lead compound binding to the XPF binding site. Aro features are aromatic, Acc are hydrogen bond acceptors and Don hydrogen bond donors. .... 173

Figure B.3. Cell proliferation of HCT-116 cells treated with compounds 4 (active inhibitor) and 5 (non-active inhibitor) using MTS assay. .... 181

Figure B.4. The inactivation experiments were carried out at 25°C in the reaction medium indicated in the “Materials and Methods” section which also included 10% or 5% DMSO final concentration (A or B, respectively). The reaction was started (T0) by the addition of ERCC1-XPF (2.5  $\mu$ g protein/ml) followed by the addition of 3  $\mu$ M compound 4. 10  $\mu$ l aliquots from these media (25 ng protein) were withdrawn at the indicated times and the activity ( $\Delta$ RFU/time) was measured for 6 minutes by adding 200 nM stem-loop substrate and additional compound 4 diluted in DMSO (10% or 5% DMSO final concentration, A and B, respectively) to render final

concentrations of 100 nM stem-loop substrate and 3  $\mu$ M compound 4 in 10  $\mu$ l for a final total reaction volume of 20  $\mu$ l. In T12, the activity was measured by the addition of 200 nM stem-loop substrate which had already been pre-incubated with 3  $\mu$ M compound 4 for 12 minutes to give the same final concentrations of reagents as indicated for T0. .... 182

Figure D.1. Chemical structure of A) R-837, B) colchicine and C) CR42-24. For the latter two, the rings are named according to the convention used in the article (A, B, C). .... 191

Figure D.2. Structure of the homodimeric model of the TLR7 ectodomains obtained in our previous work from the TLR8 template. The a and b monomer surfaces are represented in brown and cyan, respectively. Each monomer secondary structures are colored as follow: helices in red, strands in yellow, loops and turns in white. The R-837 binding site targeted in this study, lying at the interface of the two monomers, is represented as green surface. .... 194

Figure D.3. Analysis of the predicted binding pose of CR42-24 to the R-837 binding site of TLR7. A) Comparison between the docking poses of R-837 (dark green carbons) and CR42-24 (cyan carbons) to the TLR7 binding site. Red surface indicates solvent-exposed zones, white surface indicate hydrophobic zones and purple surface indicates polar zones of the site. Key common interactions are the position of the three-ring moieties in the hydrophobic zone of the binding pocket and the orientation of amine groups in both the compounds towards the polar zone surrounding aD555. In addition, the carbonyl group of the C ring of CR42-24 was oriented in proximity of bK432. B) Hydrophobic contacts of CR42-24 with the TLR7 residues of the binding pocket. From the reported docking pose, we observed two distinct hydrophobic zones interacting with the compound, one constituted by aL557, bY356 and bF408 (represented in purple) and the other one by aI585, aT586, bF351 and bV381 (represented in orange). Two hydrophobic clefts, interacting with the same compound rings (A and C) as in our model, were observed for colchicine bound to the tubulin site. The pose is represented in an orientation similar to Figure S6 from [20] for comparison with the colchicine binding pose to the tubulin site. .... 196

Figure D.4. RMSD trends for the TLR7 backbone atoms (green), the backbone atoms of the binding pocket residues (blue), the side chain atoms of the binding pocket (cyan) and the heavy atoms of the CR42-24 molecule (green), calculated between the initial phase of restraint release and the end of the 10 ns-long production simulation (10.2 ns of total time). .... 198

Figure D.5. Decomposition of total, van der Waals and electrostatic interaction energies between the residues of the TLR7 pocket and the two compounds R-837 (violet) and CR42-24 (green). Similar patterns were observed for the total and van der Waals interactions of the two compounds with the TLR7 dimer, while the electrostatic interaction pattern was different. The unfavorable interaction of CR42-24 with the key residue aD555 was compensated by a favorable, electrostatic-driven interaction with bK432, which was not observed for other TLR7 ligands in our previous study. .... 201

Figure D.6. Volcano plot reporting the Anova analysis performed over thirty-nine cancer genes, including TLR7. The x-axis represents the factor of IC<sub>50</sub> shift, while the y-axis reports the p-values, i.e. the confidence level for genetic association of mutations in a particular gene with a

IC<sub>50</sub> shift. The areas in the circles are proportional to the number of mutants present in the cell panel (each reported gene was present in at least three Oncolines™ cell lines). ..... 202

Figure D.7. The additional four poses of CR42-24 obtained with molecular docking to the R-837 binding site of TLR7. Computed binding energies are also reported. .... 212

Figure D.8. Superposition of the crystallographic pose of DAMA-colchicine (dark green carbons) in tubulin from pdb ID 1SA0 and the CR42-24 docking pose in TLR7 (carbons in light green). The first hydrophobic cleft of the tubulin binding site, interacting with the C ring of the DAMA-colchicine, is reported in blue sticks. The corresponding first hydrophobic zone of TLR7 is reported in purple sticks. The second hydrophobic zone is reported in cyan sticks for tubulin and orange sticks for TLR7. Tubulin backbone is reported in yellow ribbons, TLR7 backbone in grey ribbons. .... 213

Figure D.9. RMSD trends of the MD production simulation for the TLR7 backbone atoms (green), the backbone atoms of the binding pocket (blue), the side chain atoms of the binding pocket (cyan) and the heavy atoms of CR42-24 (red), calculated using the average structure of the last 5 ns as reference. The highest fluctuations are observed for the side chain atoms of the binding pocket, as they adapted to the bound conformation upon ligand binding. .... 214

Figure D.10. Salt bridge (in purple) established between aD555 and bK432, keeping the former far from the amine of the acetamide group of CR42-24. .... 215

## Chapter 1 - Introduction

Developing a new drug is a complicated, expensive and time-consuming process. Serendipity played an important role in the discovery of active compounds for many years. With the recent advent of the omic era, however, pharmacological research has become more and more dependent on rational drug design techniques. A typical rational drug design pipeline is divided in three steps: identification of a biomolecule critically involved in the development or progression of a disease; biological and structural characterization of such target; rational design of compounds able to modulate its action. In this context, computer-aided methods cover today a role at least as important as traditional “wet” experimental techniques. Techniques such as homology modeling and molecular dynamics can, for example, predict the 3D structure and the dynamics of a protein or nucleic acid in a limited amount of time, especially when compared with structural biology methods such as x-ray crystallography. Molecular docking is another computational method which allows to quickly identify the native bound conformation of a ligand to a target at the atomistic scale, as well as to estimate the binding affinity. Structure-based computational drug design includes one or more of the mentioned methods. Indeed, the 3D structure of the target can be obtained from a repository of experimentally-determined structures such as the Protein Data Bank (PDB), or modeled computationally. Molecular docking and scoring can then be employed to target a specific site of the target and *in-silico* screen large compound libraries, retaining just high scoring entries for experimental testing (virtual screening, VS) or to design better analogues of a known lead compound by optimizing its interactions with the target and the surrounding solvent. Ligand-based drug design, on the other hand, uses the information of known active compounds to identify novel small molecules without making use of the target structure. Using these methodologies, drug design is on its way to become a high-

confidence, cheaper and faster process, with the hope to provide optimized novel therapies for many diseases.

The focus of this thesis work is DNA repair. Our genome is constantly under attack by a number of endogenous and exogenous agents that can damage it. To protect against these damages and the deleterious effects they can cause, cells are provided with DNA repair mechanisms that can recognize, remove and replace damaged DNA zones. Indeed, an aberrant or deficient repair pathway in healthy cells can lead to mutations during the DNA replication and consequently to cancer and other diseases. In cancer cells, however, DNA repair can interfere with widely-used DNA-damaging therapies such as platinum-based chemotherapy and radiotherapy, leading to acquired drug resistance. Developing inhibitors that can switch off the DNA repair in tumor cells is therefore a promising strategy to enhance the effect of such DNA-damaging strategies. Hence, the rationale I have adopted in this thesis was to design inhibitors of the action of the DNA excision repair protein ERCC1-DNA repair endonuclease XPF (ERCC1-XPF), a protein complex essential for one of these pathways, the nucleotide excision repair (NER) pathway. NER protein expression and action in cancer cells is linked with the response to platinum-based chemotherapy, one of the gold standards of cancer treatment around the world. The two main questions this thesis is trying to address are: is it possible to rationally design compounds able to optimally bind to one of the three critical binding sites of ERCC1-XPF? Will these compounds be able to inhibit the NER activity and potentiate the effect of DNA-damaging cancer therapies? The following chapters are my attempt to answer these two questions.

This thesis is based on a series of book chapters and articles that were or are being published: Chapter 2 covers the theoretical background of computational methods used in computer-aided drug design (CADD). Chapter 3 presents a systematic review of recent research studies

investigating various aspects of NER through computational approaches. Chapter 4 reports the development and application of a VS pipeline aimed to inhibit the DNA repair protein complementing XP-A cells (XPA)-ERCC1 interaction, essential to recruit ERCC1-XPF to the damaged site. In chapter 5, the focus is switched to the catalytic activity of the complex, where I investigated the binding modes of known inhibitors of the active site of XPF and carried out a VS campaign aimed to identify novel inhibitors. In Chapter 6, a rational drug optimization effort for ERCC1-XPF dimerization inhibitors is reported. Chapter 7 summarizes the reported results and suggests possible future directions. Appendix A and B includes all the published supplementary materials of the chapters before. Appendix C is a commentary of the results reported in Chapter 4. Appendix D is a computational investigation of a possible off-target interaction of a tubulin dimerization inhibitor, which I carried out as side project along with my thesis's main topic. The chapter order follows the chronological order in which the different projects were developed, and each of them is structured as a review or research article in the field of CADD.

## **Chapter 2 - Molecular Dynamics and Related Computational Methods with Applications to Drug Discovery<sup>1</sup>**

### **2.1 Abstract**

The main objective of this review chapter is to give the reader a practical toolbox for applications in quantitative biology and computational drug discovery. The computational technique of molecular dynamics is discussed, with special attention to force fields for protein simulations and methods for the calculation of solvation free energies. Additionally, computational methods aimed at characterizing and identifying ligand binding pockets on protein surfaces are discussed. Practical information about available databases and software of use in drug design and discovery is provided.

### **2.2 Introduction**

Computational drug discovery is a conceptual approach to finding drug-like molecules by rational design, based on the information regarding their intended biomolecular target. A drug target is an important molecule, usually a protein, involved in a particular metabolic or signaling pathway that is specific to a disease condition. Most approaches attempt to inhibit the functioning of an aberrant or over-expressed pathway in the diseased state by interfering with the normal activity of the target. Medicinal compounds as candidate drugs can have their structures rationally designed at a molecular level in such a way as to optimize their binding to the active region of their target biomolecule in order to inhibit its activity and to simultaneously minimize

---

<sup>1</sup> This chapter is part of a book chapter published as Preto, J.; Gentile, F.; Winter, P.; Churchill, C.; Omar, S. I.; Tuszynski, J. A. Molecular Dynamics and Related Computational Methods with Applications to Drug Discovery. In *Coupled Mathematical Models for Physical and Biological Nanoscale Systems and Their Applications*; Bonilla L.L., Kaxiras E., Melnik R., Eds.; Springer, Cham, 2018; pp. 267–285.

their effects on other important biomolecules that may cause undesired side effects. Since many challenges are posed by the large chemical and biological spaces involved in designing drugs with high specificity and selectivity, serendipity has traditionally played an additional important role in finding potential new drugs. Conversely, structure-based drug design requires knowledge of the structure of the biomolecular target, and it utilizes 3D information about biomolecules obtained from techniques such as x-ray crystallography and NMR spectroscopy.

The first step in the rational drug design process is usually the identification and characterization of the biomolecular target, such as a protein or a DNA sequence. From here, computational techniques can be used to model a drug within the binding site of the biomolecular target, and this information can be used to design novel drug panels with enhanced activity. Of the computational techniques available, molecular dynamics (MD) is particularly important in the investigation of target characterization and drug-target interactions. In Section 2.3, an overview of the main aspects of MD simulations—including force field descriptions—and related methods intended to characterize drug-target binding is provided. In Section 2.4, other computational drug-discovery strategies, such as binding pocket prediction and molecular docking, are described. Virtual screening (VS) techniques are also discussed.

## **2.3 Molecular Dynamics**

### *2.3.1 General*

Like most experimentally-measured properties of molecular systems, the binding affinity of a drug to its target is a thermodynamic quantity, i.e., an ensemble average over a representative statistical ensemble of a system. As a result, the knowledge of a single 3D structure of a given protein complex—obtained, e.g., from x-ray crystallography or cryo-electron microscopy—even

if associated with a global energy minimum, is not enough to theoretically predict such macroscopic properties. Instead, it is necessary to generate a representative ensemble of conformations of the same system at a given (typically physiological) temperature. Two popular computational methods may be applied to this end: molecular Monte Carlo simulations (MC) [1] and MD [2]. For the study of dynamic or non-equilibrium properties (e.g., the transport of molecules across biomembranes, chemical reactions, etc.), only the second method may be utilized. Although MC simulations are simpler than MD ones, they usually do not lead to any better statistics in a given amount of time [3]. That is why MD is generally preferred over MC. Popular MD engines include Amber [4], GROMACS [5], LAMMPS [6], and NAMD [7].

MD simulations usually involve the numerical integration of Newton's equations of motion for a system of  $N$  interacting atoms representing the system of interest, possibly including the molecules of the surrounding solvent:

$$m_i \frac{d^2 r_i}{dt^2} = F_i, \quad i = 1 \dots N, \quad (2.1)$$

where  $r_i$  is the position of atom  $i$ ,  $m_i$  its mass and  $F_i$  is the force acting on it, equal to the negative derivative of the molecular potential  $U$ , i.e.,  $F_i = -\partial U / \partial r_i$ .

Using Newton's equations of motion automatically implies the use of classical physics, classical MD having the advantage of being far less computationally demanding than real quantum-dynamical simulations, which require solving the time-dependent Schrödinger equation for the system of interacting particles forming the molecule. However, because of classical approximations, standard MD simulations suffer from several limitations that the reader should be aware of. First, electronic motions are not considered per se. Instead, it is supposed that

electrons are always in their ground state adjusting their dynamics instantly when atoms are moved (Born-Oppenheimer approximation). Secondly, most potential energy functions  $U$  used to model atomic interactions, commonly referred to as *force fields* in chemistry and biology, are empirical thus approximate. They usually consist of a summation of bonded forces and non-bonded pair-additive forces. Such analytical potentials include free parameters (e.g., coupling constants, equilibrium bond lengths, van der Waals radii, etc.), which are estimated by fitting against detailed electronic calculations or experimental properties (e.g., spectroscopy measurements, elastic constants) in order to reproduce observed experimental equilibrium behaviors [8–10]. Typical classical MD force fields adopt the following functional form:

$$\begin{aligned}
 U = & \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \\
 & + \sum_{\text{torsions}} K_\phi [\cos(n\phi + \varphi) + 1] + \sum_{i < j}^N \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (2.2)
 \end{aligned}$$

The first term in Equation 2.2 represents the potential between two chemically-bound atoms, modeled as a simple harmonic potential,  $b$  being the distance between the two atoms and  $b_0$  the equilibrium bond length. The proximity between three atoms, which are connected via chemical bonds can be described with an angle. The second term in Equation 2.2 stands for this angle-dependence involving three atoms and is also modeled by a harmonic potential,  $\theta$  being the angle between the three atoms in the structure and  $\theta_0$  its equilibrium value. The third term represents the dihedral angle (torsion) potential and depends on four atom coordinates. Such a potential is periodic and is represented by a cosine function with  $n$ , the number of maxima and  $\varphi$ , the angular offset. The variable  $\phi$  is obtained from dihedral angles in the structure. Noticeably, an

additional term may be included in Equation 2.2 to model out-of-plane bending motions, i.e., improper dihedral angles. This is usually done through a cosine or a harmonic function. The last two terms in Equation 2.2 account for non-bonded interactions and are calculated pairwise between atoms  $i$  and  $j$ . The fourth term is the van der Waals potential, which is typically represented by a Lennard-Jones 6-12 potential. The  $1/r^6$  term is the attractive component while the  $1/r^{12}$  term approximates Pauli repulsion. Parameters  $A_{ij}$  and  $B_{ij}$  are atom specific while  $r_{ij}$  stands for the distance between atoms  $i$  and  $j$ . The final term corresponds to the electrostatic potential between atoms, and is modeled as a Coulomb potential. Parameters  $q_i$  and  $q_j$  represent (fixed) charges on atoms  $i$  and  $j$ , while the constant  $\epsilon_0$  is the vacuum permittivity. Electrostatic interactions dominate over van der Waals forces for long-range intermolecular interactions and they play a significant role in non-chemical binding.

MD simulations result in trajectories, which contain information about the changes of atomic positions over time, which can be analyzed in great detail to extract pertinent information regarding the dynamics of the system. This includes the root-mean-square deviation (RMSD) of ligand and protein atoms, supramolecular (non-covalent) interactions, changes in the potential energy of the system, short-lived reaction intermediates [11], conformational changes, flexibility, and optimum binding modes [12] among many various properties of the biomolecule and its environment. In a computer-aided drug design process, the mobility of crystal water molecules near proteins observed in MD simulations can help identify the amino acid residues that play an important role in ligand binding. MD simulations can also be used for studying ionic conductivity [13,14], where the simulations provide atomic level insights into ionic mobility. In terms of particular applications, MD has been successfully used to study clinically important

proteins such as HIV-1 gp120 [15], binding sites [16], drug resistance mechanisms [17], and protein folding [18,19] to name but a few.

### 2.3.2 Molecular Dynamics and Drug Discovery

Molecular dynamics can be used together with other methods to solve a host of problems in biomolecular modeling [20,21]. In the case of VS methods that involve large libraries of chemical compounds in order to identify a high-affinity small molecule that is expected to act as an enzyme inhibitor, or a protein-protein interaction blocker, the calculation of the binding energy of potential hits may help prioritize compounds for experimental testing.

While docking and scoring remain the most widely used computational techniques to predict the binding mode and affinity of a drug to its target due their low computational cost, these methods are not particularly accurate. More precise approaches utilize appropriate sampling of the molecular system generated beforehand with MD simulations as is required when estimating ensemble-averaged quantities like binding free energies. *End-point methods* such as linear interaction energy (LIE) and the molecular mechanics Poisson-Boltzmann Surface Area (MM/PBSA) technique, which rely only on appropriate samplings of the end states, i.e., the complex and possibly the free receptor and ligand, have intermediate efficiencies. The LIE method, originally introduced by Aqvist et al. [22], assumes that the binding free energy can be written as a linear combination of average interaction energies between the ligand and the rest of the system (protein, water and ions). More explicitly, the binding free energy of the ligand is expressed as [23]:

$$\Delta G_{bind} = \alpha \langle \Delta E_{vdW}^{L-S} \rangle + \beta \langle \Delta E_{el}^{L-S} \rangle, \quad (2.3)$$

where  $\langle \Delta E_{vdW}^{L-S} \rangle = \langle E_{vdW}^{L-S} \rangle_{bound} - \langle E_{vdW}^{L-S} \rangle_{unbound}$  refers to the change in van der Waals interactions between the bound and unbound states of the ligand. The averages stand for ensemble averages obtained from MD simulations whereas the L-S label indicates that the interaction energies are computed only between the ligand and the surroundings. Similarly,  $\langle \Delta E_{el}^{L-S} \rangle = \langle E_{el}^{L-S} \rangle_{bound} - \langle E_{el}^{L-S} \rangle_{unbound}$  corresponds to the change in intermolecular electrostatic interactions between the bound and unbound states. Parameters  $\alpha$  and  $\beta$  are generally obtained empirically using an appropriate fitting procedure.

Alternatively, MM/PBSA [24], which is arguably the most popular end-point method, turned out to be successful in a number of drug-design case studies [25–27]. The method basically provides an estimate of the binding free energy as:

$$\Delta G_{bind} = \langle \Delta E_{MM} \rangle - T\Delta S + \Delta G_{solv}, \quad (2.4)$$

where  $\langle \Delta E_{MM} \rangle - T\Delta S$  can be regarded as the change in the free energy of the system in vacuum (gas phase); it includes the change in the molecular mechanics energy due to the binding  $\langle \Delta E_{MM} \rangle = \langle E_{MM} \rangle_{bound} - \langle E_{MM} \rangle_{unbound}$  and the change in the conformational entropy  $\Delta S$ , usually estimated from normal mode analysis (NMA) performed on the complex structure and on the free ligand and protein structures. As in the LIE method, every average quantity corresponds to an ensemble average obtained from output MD trajectories. The entropy contribution, which is relatively time-consuming and inaccurate to compute using NMA, can be neglected if a comparison of states of similar entropy is desired such as in the case of comparing two or more ligands binding to the same protein binding site. Finally,  $\Delta G_{solv}$  stands for the difference of solvation free energies due to the binding, it is given as  $\Delta G_{solv} = \Delta G_{solv}^{complex} - \Delta G_{solv}^{lig} - \Delta G_{solv}^{prot}$  where every term on the right-hand side is given as the sum of polar and nonpolar contributions.

The polar parts are obtained by solving the Poisson-Boltzmann (PB) equation or by using the Generalized-Born (GB) model (as in the MM/GBSA method) whereas the nonpolar terms are estimated from a linear relation to the solvent accessible surface area (SASA). Despite the fact that MM/PBSA and MM/GBSA are computationally-inexpensive methods, they contain several crude and questionable approximations, e.g., due to the use of implicit solvent models to compute the solvation energies [28]. The capability of the MM/PBSA method to predict the correct binding free energy turns out to be more sensitive to the investigated system compared to the MM/GBSA method, the latter being more useful in multi-target comparisons [29]. Noticeably, the MM/PBSA and GBSA techniques can be used to perform per-residue-free-energy decompositions. The benefit of such decompositions is twofold: providing important information about residues which significantly contribute to the binding energy (hot spots) and giving insights into the changes in binding free energies due to mutations, especially single point mutation.

## **2.4 Other Computational Drug-Discovery Methods**

This section is intended to provide an account of other popular useful techniques for drug design and VS studies that can be used instead or in conjunction with MD-based methods discussed in the previous section.

### *2.4.1 Binding Pocket Prediction*

Identifying and characterizing a suitable binding pocket in a 3D protein structure is a central aspect of any drug discovery study. This step is also relevant to shed light on biomolecular functions as many proteins are biologically functional only after interacting with cofactors or other biological molecules.

A common way to define a binding pocket, if a ligand is already bound to it, is to introduce a distance cut-off. Binding pocket atoms are typically defined whenever their distance to the ligand is below 4–8 Å. Following is a list of physicochemical key properties of binding pockets:

- The solvent-accessible surface area (SASA) which is usually computed as the atoms of the pocket reachable by a solvent probe sphere rolling over the protein surface.
- The volume of the pocket and its depth which corresponds to the average distance of the pocket atoms to their nearest water molecules from bulk solvent [30].
- The pocket hydrophobicity which depends on polar and non-polar residues involved in the binding site.
- The number of hydrogen bond donors and acceptors on the pocket surface.
- The conservation of residues over similar binding pockets of other proteins, which is particularly relevant for functional sites.

In addition to experimental binding site detection techniques such as NMR-based methods [31], a number of computational methods can be found which may be helpful in the process of binding pocket identification of a molecular target. Pocket finder algorithms are usually tested and validated on protein and ligand datasets. Such tests are intended to estimate the reliability of identifying the correct binding pocket within the first one to three hits provided by an algorithm. Two popular publicly-available databases are the Protein Data Bank (PDB) [32], which provides 3D protein structures for input into the pocket finding algorithms, and the PDBbind database [33], which contains bound protein structures filtered from the PDB database. The current version of PDBbind has around 3100 protein-ligand complexes, 1300 of these having been

manually selected to form the refined set with the focus on the quality of structures and binding data. Due to its large size and manual curation, the refined set of the PDBbind database provides a suitable benchmark for most case studies. Further reduced from this, is the core set of 210 complexes. Optimal databases for pocket prediction testing should include high-resolution, diverse and non-redundant protein-ligand complexes. Pocket finder algorithms are generally split into two classes, namely, geometric-based and energetic-based approaches.

*Geometry-based algorithms* have the advantage of a low computational cost. The underlying assumption behind such methods is that the ligand binding pocket corresponds to the larger cleft within the protein structure [34,35]. Therefore, geometrical criteria may be sufficient to identify the correct binding location on a protein. One such example, SURFNET [36] is an early-developed program which fits spheres between pairs of atoms so that they do not contain more than one atom. The binding pocket is defined as the volume containing the largest number of adjacent spheres. An improvement of the program, called SURFNET-ConSurf [37], refines the binding pocket prediction also considering the residue conservation within the binding site. The SURFNET-ConSurf algorithm was tested on a set of 244 non-redundant, diverse and representative ligand-protein complexes, obtained by a filtered version of the PDB database. A 75% rate of successfully recognized native ligand pockets is reported in the original paper about this method [37].

Another algorithm called VisGrid [38] is based on geometrical hashing and identifies cavities by considering the visibility of each point in a 3D grid, that is, the fraction of directions that are not blocked by protein atoms. In this way, a cluster of closely-located grid points with limited visibility indicates a pocket. VisGrid was compared with other pocket prediction methods,

including SURFNET and LIGSITE, and the observed success rates on a set of bound and unbound structures were comparable with existing methods.

LIGSITE [39] uses a grid-based method in which points are either assigned to the solvent or protein category, and cavities are defined as groups of points in which solvent points are surrounded by protein points. Although the LIGSITE original validation identified the correct binding pocket in all the testing cases, a big limitation of this study was the reduced size of the dataset, with only ten ligand-receptor complexes. Its extension, LIGSITE<sup>csc</sup> [40], improves the original algorithm by calculating more accurately the contact between protein surface and solvent using the Connolly surface, and by re-ranking the identified pockets by their degree of residue conservation in homolog proteins. The LIGSITE<sup>csc</sup> testing process is more significant than the LIGSITE one, and a comparison with other geometry-based methods is also provided. The success rates calculated on a set of 210 non-redundant bound structures were 75% for LIGSITE<sup>csc</sup>, 65% for LIGSITE and 42% for SURFNET. The algorithm also showed good performances in recognizing the correct binding pocket in unbound structures.

Another class of methods dedicated at identifying binding pockets are *energy-based methods* which rely on the energetic properties of a binding site. A common approach of these methods is to use molecular probes to search for favorable interaction sites on a protein, and cluster them together to identify putative pockets. An early effort resulted in the GRID Fortran code [41]. The probes employed by this algorithm include water, methyl group, the hydroxyl, amine nitrogen and carboxy oxygen. Energetic contours are calculated with a function considering a 12-6 Lennard-Jones term, an electrostatic term and a hydrogen bond term, and negative energy levels indicate promising interaction sites for each probe.

Laurie and Jackson's Q-SiteFinder method [42] calculates the interaction energy of a methyl probe and the grid points generated on the protein structure. A clustering analysis step links favorable interaction sites to rank putative binding pockets based on their total interaction energy. Q-SiteFinder was tested on a diverse set of bound and unbound protein conformations, resulting in success rates of 74% and 71%, respectively.

EasyMIFs and SITEHOUND [43] are two complementary energy-based tools developed at the Sanchez lab. The first algorithm calculates the interaction energy between grid points and molecular probes using the GROMOS force field, while SITEHOUND recognizes putative binding sites by filtering and clustering the spatial variation of the interaction energy fields calculated by EasyMIFs or any other grid-based program. Multiple probes are used, as well as different site clustering algorithms. SITEHOUND's success rate was evaluated on a set of 77 complexes and it was reported as 95% (bound structures) and 79% (unbound structures) considering the binding pocket identified when present in the top three ranked sites.

Another similar energy-based algorithm, AutoLigand [44], was created by the developers of the popular molecular docking software Autodock. AutoLigand was reported to have a success rate of 73% when tested on a set of 187 bound structures and 80% when tested on 96 unbound structures.

#### *2.4.2 Ligand-Receptor Docking*

Molecular docking methods have been developed to predict how a given compound naturally binds to its biomolecular target, i.e., its binding mode, and to provide an estimate of its binding affinity. Docking software usually rely on optimization algorithms which include both a search algorithm and a scoring function. Such methods require at least one ligand structure and one

target structure as inputs. The location of the targeted site should be provided although blind docking approaches [45,46] can help deal with unknown binding locations in addition to the pocket prediction methods discussed in the previous section.

The search algorithms are dedicated to exhaustively explore the conformational space of the ligand within the targeted pocket. Three groups of such functions have been described, namely matching, systematic and stochastic methods. *Matching algorithms* are based on shape complementarity between the ligand and the receptor site, and, possibly, chemical complementarity. *Systematic search algorithms* explore the degrees of freedom of the ligand in a progressive way. This class of methods can be divided in three subgroups:

- Exhaustive algorithms systematically rotate all ligand dihedral angles until an optimal solution is reached.
- Fragment-based methods break down the ligand into different fragments which are separately placed within the binding site and re-connected in the last step of the process.
- Ensemble-based methods pre-generate a large number of ligand conformations, which are then rigidly placed within the binding site.

The last class of search algorithms includes *stochastic methods* such as MC and evolutionary algorithms, which introduce random changes of the degrees of freedom of the ligands to rapidly reach an optimal solution.

In molecular docking, the binding free energy is calculated using a position-dependent scoring function. This is required not only to identify the correct binding pose corresponding to the lowest binding energy, but also to rank a set of tested compounds according to their affinity to a

target. *Force-field-based scoring functions* used for docking are similar to MD force fields discussed in Section 2.3.1. In *empirical scoring functions*, the different contributions in the binding energies are weighted with coefficients, set beforehand to reproduce experimental dissociation constants of known ligand-receptor complexes. The Autodock4 scoring function [47] is an example of an empirical scoring function where a non-bonded interaction potential is calculated as

$$V = W_{vdW} \sum_{ij} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hb} \sum_{ij} E(\theta) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\frac{-r_{ij}^2}{2\sigma^2}}. \quad (2.5)$$

The first term represents van der Waals contributions, the second is a hydrogen bond term, the third is a Coulombic term for electrostatic interactions, and the last is a desolvation potential.  $W$  represents the empirical coefficients, obtained from the training over 188 bound complexes from a PDB calibration subset. The coefficients  $A$  and  $B$  derive from the AMBER force field, while  $C$  and  $D$  are Autodock internal parameters.  $E(\theta)$  depends on the angle of deviation  $\theta$  from the ideal hydrogen bond geometry.  $S$  and  $V$  are a solvation parameter and the volume of the atoms surrounding one atom, respectively.  $\sigma$  is distance-weighting factor of Autodock. The total Autodock score of a binding pose is calculated by summing the difference of intra-molecular energies between the bound and unbound form of the ligand and the protein, then subtracting the difference of inter-molecular energies. A simple entropic term is also included in the final score to model the variation of the system entropy upon binding. *Knowledge-based scoring functions*

use the potential of mean force (PMF), derived from protein-ligand structures and calculated for each  $ij$  ligand-receptor atom pair type as:

$$w_{ij}(r) = -k_B T \ln \left( \frac{\rho(r_{ij})}{\rho^*(r_{ij})} \right) \quad (2.6)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature in Kelvin degrees,  $\rho(r_{ij})$  is the number density of the  $ij$  atom pair derived from the structural training set and  $\rho^*(r_{ij})$  is the number density in a reference state. Although knowledge-based scoring functions do not provide a precise interaction energy potential due to difficulties arising from the reference state calculation, they directly connect the atomic interactions between ligand and protein to structural data instead of to kinetics as is the case for empirical methods. Knowledge-based scoring functions turn out also to be more computationally efficient than force-field-based methods [48,49]. Finally, *consensus scoring methodologies* combine different scoring function outputs of the same ligand to obtain a single, consensual score. Different combination strategies can be employed, such as weighting and summing up the ranks or performing a regression analysis [50].

Accounting for the flexibility of the binding site is an important issue in molecular docking. As a result, different approaches have been proposed to address this problem. Soft docking algorithms use modified short-range repulsion parameters for the binding site atoms, which allow the ligand to slightly penetrate through the surface of the pocket to mimic the induced fit of the binding. Many algorithms also include the possibility to treat the side chains of pocket residues as flexible, although such methods still ignore backbone dynamics, while increasing noticeably the computational cost. Using multiple receptor conformations when performing molecular docking is a popular way to take into account the backbone flexibility [51]. This approach, called Relaxed

Complex Scheme (RCS), relies on NMR or MD-derived conformational ensembles, which are used as molecular docking targets.

Testing of docking software is usually performed by evaluating the percentage of docking poses with small enough RMSD (typically 2 Å) compared to the co-crystallized poses extracted from a high-resolution structural database.

The original implementation of DOCK [52] is an example of geometry-matching algorithm, where the binding site and the ligand atoms are represented as spheres that are systematically matched using a shape-based routine. DOCK 6 [53], the latest version of the program, applies a fragment-based algorithm and a set of different force-field-based scoring functions which can be selected, such as PBSA, GBSA and Amber scoring methods. In addition, a minimization step is performed for the ligand in order to remove minor protein-ligand clashes and relax its internal geometry. The success rate of the latest DOCK release was estimated around 73% in reproducing crystallographic poses. The authors tested the algorithm on 1043 structures obtained from a ligand-receptor database designed as a benchmark for assessing docking software performances [54].

Autodock [55], probably the most popular docking software, uses a Lamarckian genetic algorithm to independently generate a large number of binding poses, scored with the empirical scoring function described in Equation 2.5. A clustering algorithm can be optionally used to identify the most populated portion of the conformational space of the ligand, from which the lowest energy pose should match the native one. The success rate of the latest version, Autodock4 [56], was around 53% when the software was tested on the calibration structural set.

Autodock Vina [57] utilizes an iterated local search global optimizer searching method. The Vina scoring function combines aspects from knowledge-based and empirical potentials. Tested on the same set used for Autodock4, Vina was able to identify the correct binding pose in 78% of the cases. Noteworthy, Vina scoring function was trained with the PDBbind refined set, much bigger than the training set used for the Autodock scoring function.

Glide [58] is a docking software included within the Schrödinger molecular modeling package. It is based on an exhaustive systematic search algorithm used to sample the ligand conformational space, followed by a minimization step. An optimal choice for the scoring method [58] is given as a combination of a force-field-based function, an empirical function (GlideScore) and the strain energy of the ligand conformation. The pose success rate was reported around 66% when tested on 282 ligand-receptor complexes selected from the PDB database.

GOLD [59] is another popular docking program. The software maps together complementary chemical features of the ligand and the receptor within the binding site. A genetic algorithm is then used to explore different binding modes. Three main scoring functions are available, namely, Goldscore (force-field-based), Chemscore (empirical) and Astex Statistical Potential [60] (knowledge-based). Testing of GOLD on the CCDC/Astex database [61], a PDB subset designed to test docking software, resulted in success rates up to 87% depending on the scoring function used. The correlation coefficients ( $R^2$ ) between experimentally-measured and GOLD binding affinities were reported between 0.51 and 0.55.

### *2.4.3 Virtual Screening*

The discovery of a new drug is an expensive and long process. It is estimated that up to two and half billion dollars and twenty years are required to bring a new product from the bench to the

clinic [62,63]. Consequently, efforts are made to shorten the process and reduce the cost. Some of the time and funding savings are expected to result from a wider use of computational techniques applied to drug discovery. One of such techniques is called virtual screening (VS). VS refers to an *in-silico* active compound search against biomolecular targets [64]. It has the advantage of being fast and inexpensive compared with traditional high throughput screening. Nowadays, libraries including billions of compounds can be virtually screened depending on the available computational resources.

A typical VS workflow consists of sequential series of filtering and scoring steps aimed at providing a set of promising compounds for experimental validation. VS methods can be divided in ligand-based VS (LBVS) and structure-based VS (SBVS). LBVS approaches are computationally faster but they do not provide any estimate of the binding energy of the ligand. On the other hand, SBVS methods are more computationally expensive but they enable to rank potential hits based on their predicted binding affinity. Regardless of the chosen approach, a compound database is always required as starting point for VS. Examples of extensive small molecule repositories are PubChem [65,66], ZINC [67] and the National Cancer Institute databases [68]. Such collections usually include millions of compounds that can be downloaded for screening purposes [69].

LBVS techniques rely solely on the 2D or 3D structure of ligands, ignoring the biological target. The main assumption behind these methods is that structurally related compounds share similar activities [70]. Therefore, the structure of at least one known active compound should be available as template for the computational search, and a measure of the distance between structures needs to be computed. Simple ways to represent the chemical structure of a compound in a computer-readable format are chemical fingerprints or pharmacophore representations.

Chemical fingerprints [71] are binary strings in which each bit codes for the presence or the absence of particular chemical groups. A widely-used way to compare two fingerprints is to use the Tanimoto index, given by

$$T_{A,B} = \frac{c}{a + b - c}, \quad (2.7)$$

where A and B are the two fingerprints, c is the number of bits set to 1 at the same position in both the fingerprints, and, a and b are the total number of bits set to 1 in A and B, respectively. A publicly-available package which can be used for fingerprint-based VS is chemfp [72].

Another way to perform LBVS is to use a pharmacophore model of the active compounds [73], which provides a representation of the ligand from its spatially-distributed chemical features (hydrogen bond acceptor, hydrogen bond donor, hydrophobic moiety, ring structure, polar or charged residue) including the distances between centers forming a chemical structure. In pharmacophore-based VS, the distance between two ligand structures is usually calculated as the RMSD between the superposed pharmacophore points. The main benefit of this approach is to identify molecules with different chemical groups but similar generic features, providing novel scaffolds to medicinal chemistry. Contrary to most chemical fingerprints, pharmacophore models also include 3D properties of the ligand.

Data mining and machine learning methods including support vector machines, neural networks, Bayesian networks and decision trees, are also utilized for LBVS [74]. LBVS methods are useful in case a 3D structure of the target is not available, but they can also be used to clean up large databases in order to generate focused libraries [75]. Indeed, these structurally-related subsets are designed to interact with a specific target and they are built by screening larger and diverse

databases. Focused libraries have a limited size compared with the parent databases. Therefore, they can be rapidly and efficiently screened with SBVS or experimental techniques. LBVS methods have led to the discovery of novel and promising compounds with low-range potency [76]. Examples of such successes are the discovery of anti-cancer tubulin dimerization inhibitors [77], inhibitors of the 17 $\beta$ -HSD2 enzyme for osteoporosis treatment [78] and novel scaffolds for the inhibition of the HIV-1 integrase [79].

SBVS methods provides a ranking of the screened compounds based on their computed binding affinities. Therefore, one or multiple structures of the target are required. SBVS always relies on molecular docking methods, which are used to place the compounds within the targeted pocket, and to estimate the binding affinity from the resulting binding poses. We have already discussed all the docking-related aspects in Section 2.4.2. The docking scoring functions are designed to quickly estimate the binding energy from a ligand pose, therefore they often do not lead to very accurate results. Several strategies have been developed to deal with this, including the already mentioned consensus scoring, MD simulations of the complex structures and/or more accurate scoring functions (e.g., MM/PBSA or GBSA) [80,81]. Recently, SBVS techniques were successfully applied to the discovery of DNA repair inhibitors [80,82–85], anti-malarian compounds [86], kinase inhibitors [87] and HIV-1 inhibitors [88,89].

## **2.5 Conclusions**

This review chapter provides introductory information regarding the computational tools currently used in the drug design and discovery process. We have given an overview of molecular dynamics methods that are very useful in biomolecular target characterization for drug action. We have also given practical information regarding the identification of binding pockets

for putative inhibitors of proteins, as well as an overview of molecular docking techniques that are based on the protein-ligand interactions. These interactions and their ranking involving the binding free energy of the ligand-target pair are used in massive searches for specific and selective inhibitors of particular protein, a methodology referred to as virtual screening. The latter methodology relies on large and diverse databases of pharmacologically-acceptable compounds. Lists of databases and software packages used in all stages of computational drug design have been presented in this chapter to assist in practical aspects of research in this area.

## Chapter 3 - Modeling NER Repair Pathway: Recent Advances and Future Directions<sup>2</sup>

### 3.1 Abstract

Background: A major class of chemotherapy drugs targets the genome of cancer cells. These DNA damaging agents induce damage to the DNA helix, resulting in the programmed death of cancer cells. An over-activated DNA repair mechanism in cancer cells can reduce the efficacy of these drugs, thereby eliminating their therapeutic benefit and developing an acquired resistance to these otherwise effective drugs. A promising approach to enhance the therapeutic window of DNA damaging agents is to target the DNA repair pathways causing this type of resistance.

Methods: Computational approaches have been applied successfully to study many of these DNA repair mechanisms at different scales and focusing on various aspects. The ultimate goal of these studies has been to identify the key players in developing resistance to DNA damaging agents and to design regulators for their activities. This review covers the most important and recent computational efforts toward this goal. This includes modeling the mechanisms involved in DNA repair and identifying novel pharmacological inhibitors for their activities.

Results: We focus here mainly on the pathways associated with an acquired drug resistance to DNA damaging agents, concentrating on the recent advances in modeling the key mechanisms and foreseeing the future directions in this field.

Conclusions: We hope that this short, yet comprehensive review can help in discovering novel strategies to overcome the resistance effects inherent in various cancer treatments.

---

<sup>2</sup> This chapter is part of a review article published as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. Modelling DNA Repair Pathways: Recent Advances and Future Directions. *Curr. Pharm. Des.* 2016, 22, 3527–3546.

## 3.2 Introduction

Our genome is constantly exposed to internal and external damaging agents. Five major DNA repair pathways are working against these continuous environmental attacks on the integrity of DNA in humans. Each of these pathways specifically acts on certain types of lesions. Some of these lesions can be removed by a single enzyme while others can involve multiple and intricate networks of protein-protein interactions. For example, the simplest of these DNA damages is O<sup>6</sup>-methylguanine lesions, which can be directly repaired by a single enzyme, namely the O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) protein. A multi-protein and more complicated example is base excision repair (BER) which removes non-bulky DNA damages such as base modifications. A similar but more intricate pathway is nucleotide excision repair (NER) mechanism, which is involved in identifying and repairing lesions that distort the DNA double helix structure. Mismatch repair (MMR) is active on base-base mismatches and insertion-deletion loops coming from the DNA replication process. Homologous recombination (HR) and nonhomologous end joining (NHEJ) pathways repair double strand breaks that can lead to genomic instability and cell death response. An excellent review of these pathways can be found in Iyama and Wilson's paper [90].

In the context of cancer treatment, DNA repair pathways can be considered as a double-edged sword. On one hand, these mechanisms are invigilating the integrity of the genome, repairing the damages that are constantly generated within the DNA chains. These damages can in fact interfere with the correct base pairing occurring during the replication of DNA, thus leading to dangerous mutations, which can be cytotoxic and lead to the cell death. Impairment of the activity of repair pathways and/or DNA damage accumulation is correlated with different pathological conditions. On the other hand, many chemotherapy approaches target the DNA of

cancer cells to induce their apoptosis leading to tumor shrinkage. The same repair pathways that protect the organism from carcinogenic mutations can act also against cancer treatment, by fixing the induced damages and restoring the native functionality of the tumor cells [91]. In addition, over-expression of DNA repair proteins is often found in certain types of cancer, therefore showing a stronger resistance effect compared with normal cells [92–94]. Exploiting DNA repair mechanisms as a way to improve cancer therapy requires a delicate approach to get the optimal benefit for patients while keeping the balance between these two extremes.

There is ample evidence regarding the relation between DNA damage and the genomic instability of cancer cells [95]. DNA damages and repair deficiencies have been linked to neurodegenerative pathologies as well, such as Alzheimer's, Huntington's and Parkinson's diseases [96]. In addition, three rare degenerative syndromes, which are directly related to the impairment in the NER pathway, act on helix-distorting lesions. This includes xeroderma pigmentosum (XP), an inherited condition of inefficiency of the XP complementing group A-G proteins, which is associated with impaired global-genome NER functionality. Affected individuals are over-sensitive to UV radiation and are likely to develop skin cancer. The disease is sometimes associated with accelerated neurodegeneration when the transcription-coupled NER is also defective. Cockayne's (CS) syndrome is due to defects on CSA and CSB encoding genes. The disease is initiated by an impaired TC-NER activity, causing premature aging, neurodegeneration, cessation of growth and drastic reduction of life span. Trichothiodystrophy is another genetic disorder caused by the alteration of the XPB/XPD helicase activity of NER. Specific symptoms of this disease can vary from hair modifications to neurological deficiencies and skeletal abnormalities [97,98].

The modulation of the activity of DNA repair pathways in the context of cancer treatment has been proposed in two different ways. The first approach involves small molecules that can target the normal repair activity in cancer cells in combination with alkylating agents or other therapies to enhance the DNA damaging effect and the treatment outcomes [99,100]. Another promising approach targets defective repair pathways in cancer cells to induce selective cell death. The latter strategy has been termed as the lethal approach in which DNA repair inhibitors can be used as monotherapy [99].

DNA repair pathways rely on the nucleotide/side pools for the essential DNA synthesis and substitutions' building blocks. Maintaining the correct level of nucleotides in this pool is important for the full activity of the repair mechanisms [101]. In particular, low concentrations of the available nucleotides/sides correlate with deficiencies in the repair mechanism [102]. An interesting difference between tumor and healthy cells is the resistance of the former to oxidative damages occurring to the nucleotide pool [103]. Two recent studies performed by Gad et al. [104] and Huber et al. [105] showed that the 7,8-dihydro-8-oxoguanine triphosphatase (MTH1) protein is essential for the survival of cancer cells by clearing the pool of 8-Oxo-2'-deoxyguanosine-5`-Triphosphate (8-oxoGTP) mutated bases, avoiding their incorporation in the DNA that can lead to cell death. Its activity is not essential for normal cells, probably because of the strict redox balance occurring in the latter compared with the impairment state observed in cancer cells. This impairment state can lead to an accumulation of damaged nucleotides. Worth to mention is the cytotoxic effect of MTH1 inhibitors in cancer cells, which was not altered even by over-expressing many BER proteins such as oxoguanine glycosylase (OGG1) and mutY homolog (MUTYH). MTH1 constitutes therefore an attractive pharmacological target for selective cancer treatment, with the possibility to overcome the BER activity in tumor cells.

Computational methods have been employed for several decades as powerful, inexpensive and fast tools to help investigate the biological world. The recent advances made in the field of high performance computing and the development of multi-scale modeling approaches have in fact given rise to a new era for the use of computers to model biological structures and processes [106,107]. Today it is not only possible but routinely done to investigate, for example, the nanoscale nature of protein-protein interactions, to model chemical reactions and to discover novel drugs using completely *in-silico* strategies. DNA repair mechanisms represent an example of multi-scale biological processes, and are thus perfect candidates for computational investigations. In this review, we focus on the computational studies produced in the last ten years to investigate the NER pathway, which is highly involved in producing cancer treatment resistance. We discuss below studies performed at different time- and space-scales, from the simulation of enzymatic chemical reactions to the coarse-grained prediction of protein-DNA complexes. Special attention is paid to atomic simulations of key proteins and rational drug design efforts of DNA repair inhibitors. An introductory section presents NER mechanism. It is our hope that this work will encourage the research community to find novel solutions for cancer therapy and to develop new computational methodologies for advanced DNA repair studies.

### **3.3 DNA Repair Pathways and Resistance to Cancer Treatment**

DNA damage occurs frequently in each cell of the human body. Some of the endogenous sources of damage are nitrogen and oxygen reactive species, hydrolytic reactions or occasional errors occurring during the replication. DNA damage can be also induced by many external damaging agents such as tobacco products, ionizing radiations and reactive chemicals [96]. The introduction of a lesion within the DNA chain can lead to errors in DNA replication and ultimately to mutations causing cancer or other diseases [98]. In order to repair these damages

before the replication, the cells have developed a complex network of multiprotein pathways to efficiently recognize and fix a wide range of lesions [90]. The deficiency of these mechanisms is associated with a hypersensitivity to damaging agents and can lead to severe pathological conditions; specific examples are the already discussed CS and the XP syndromes, two rare genetic diseases associated to the impaired functionality of the NER pathway (see above).

Although these pathways play an important role in conserving the genome integrity, they can also represent a barrier to cancer therapy. Most of the current cancer treatments are indeed designed to damage the DNA of the tumor cells, damages that these pathways can repair may largely eliminate the benefit of the cure [91]. Therefore, the inhibition of these mechanisms represents a powerful strategy to sensitize the cells to therapies such as chemotherapy or ionizing radiation [99]. Moreover, a novel role of DNA repair inhibition as cancer monotherapy that targets the defective repair mechanisms in cancer cells is attracting considerable interest (synthetic lethality approach) [108]. Cancer is often caused by mutations occurring because of a defective repair mechanism, due to alteration of expression for key proteins or their modifications. The main idea behind the synthetic lethal approach is to destroy the residual repair action of tumor cells, causing them to not be able to fix natural occurring damages and therefore to die [109–111].

Among DNA repair pathways, the BER and the NER mechanisms are two of the most involved in generating cancer treatment resistance. The lesions generated by common chemotherapies such as alkylating agents or radiotherapy are indeed natural substrates for these two pathways, and the modulation of their activities in cancer cells represents a promising strategy to enhance the effect of the treatment [99].

### 3.3.1 Nucleotide Excision Repair

NER pathway is a multi-protein mechanism that acts on bulky lesions. These lesions can distort the helical structure of the DNA and removing them is essential for a successful DNA replication. Similar to BER, depending on the localization of the damage, two different sub-NER pathways can be adopted, namely the transcription-coupled NER (TC-NER) and the global-genome NER (GG-NER). The TC-NER pathway acts on DNA lesions present on the transcribed strands of the genome and is associated with the DNA transcription process [112], while the GG-NER detects damages present on the whole genome [113].

The recognition step is the main difference between the two sub-NER mechanisms. In TC-NER, the blockage of an RNA polymerase II (RNA pol II) at the damaged site is the triggering signal for the pathway. The stalled polymerase engages the CS group A and B (CSA and CSB) proteins. In GG-NER, the XPC-UV excision repair protein RAD23 homolog B (XPC-RAD23B) complex recognizes the DNA lesion instead [114]. Other than the initial recognition of the damage, the two sub-pathways involve the same successive steps. Firstly, both pathways recruit the transcription factor II Human (TFIIH) [115]. TFIIH shows a helicase activity to both DNA directions as a result of involving the XPB (3'-5') and XPD (5'-3') proteins. In this way, the TFIIH complex removes the damaged zone and creates a bubble around the lesion [116,117]. The opening of the DNA duplex facilitates the recruitment of XPA and replication protein A (RPA) [118,119]. XPA then recruits the XPF-DNA excision repair protein (XPF-ERCC1) complex [120], which shows an endonuclease activity by cutting the 5' terminal of the lesion [121,122], while another endonuclease, XPG, is recruited by the TFIIH complex to incise the 3' terminal [123,124]. Finally, the gap-filling is performed by a DNA polymerase ( $\delta$ ,  $\epsilon$  or  $\kappa$ ) in

collaboration with PCNA and RFC, and finally sealed by the XRCC1-LIG3 complex or LIG1 [90]. The steps of NER are shown in Figure 3.1, adapted from the KEGG database [125,126].

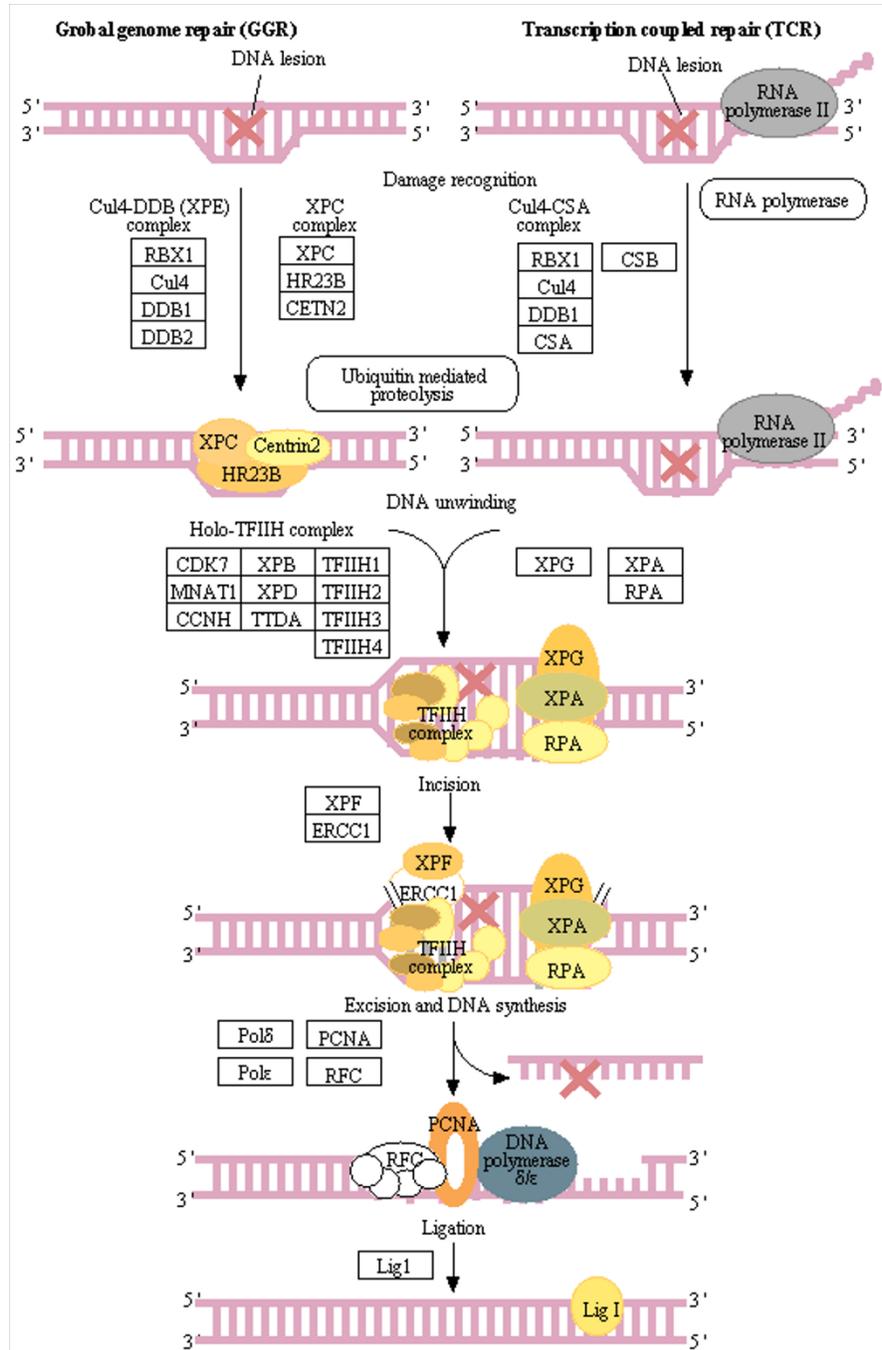


Figure 3.1. Steps of human global-genome and transcription-coupled NER pathways. The recognition step is different for the two pathways. In transcription-coupled NER, a stalled

polymerase at the damaged site is the trigger for the repair activity. The polymerase recruits then the CSA and CSB proteins. In global-genome NER the XPC-RAD23B complex recognizes instead the lesion and starts the repair action. The subsequent steps are common to both the pathways. See text for more details. Adapted from KEGG database (<http://www.genome.jp/kegg/>) [125,126].

NER is responsible for the removal of DNA damages caused by bifunctional alkylators such as cisplatin [99,127]. Several independent studies have provided evidence that the over-expression of NER proteins is correlated with cisplatin resistance in cancer, whereas low-expression causes hypersensitivity [100,128,129]. On the other hand, small molecules able to modulate the activity of NER proteins have shown a sensitizing effect of chemotherapy on cancer cells [80,84,130,131]. The incision of the lesion is the limiting step of the pathway. For this reason, key proteins to target for disrupting NER activity are RPA, XPA, ERCC1, XPF and XPG [132].

### **3.4 Modeling NER Repair Mechanisms at the Atomistic Level**

NER's key proteins have been investigated with the help of computational tools. RPA is a hetero-trimeric protein composed by a 70 kDa (616 amino acids), 32 kDa (270 amino acids) and 14 kDa (121 amino acids) kDa subunits comprising eight domains (RPA70N, RPA70A, RPA70B, RPA70C, RPA32N, RPA32C, RPA32D and RPA14) [133]. Nuss et al. employed MD simulations of the 70 kDa subunit to highlight the flexibility of this structure, especially around the residue K551. In addition, a series of protein-protein docking simulations were employed to propose models for the trimeric form and to identify buried and exposed residues. The best model was structurally close to the crystallographic one [134]. Other studies focused on the conformational changes occurring in the same subunit after single-stranded DNA (ssDNA) binding. MD simulations and X-ray/NMR observation were used in combination to study

possible interactions between ssDNA and the RPA70N domain. The findings elucidated no direct interactions as the zone remain structurally independent after the binding, opposite to the observed coordination of RPA70A/B, supporting its role as a recruiter. The important function of interdomain linkers was also investigated in latter studies [135,136]. BH3 interacting-domain death agonist (BID) protein is one of the RPA70N interacting partners, for which the function is thought to be a recruiter and stabilizer for other proteins such as ataxia telangiectasia and Rad3 related (ATR), and ATR-interacting protein (ATRIP). A model for this interaction was obtained through docking simulations by Liu et al. [143] A series of salt bridges between acidic residues of BID and basic ones from RPA70N were detected and proved the electrostatic complementarity of the two protein surfaces [137]. The binding of BID with ssDNA is mediated by the RPA70A/B structure. Carra et al. [138] employed MD simulations, PCA and MM-PB(GB)SA methods to investigate the nature of this interaction as well as the effect of single and double mutations. In their first study, they were able to reproduce the experimental binding affinities using the MM-PB(GB)SA method over separated MD trajectories for the DNA structure and RPA70A/B complex, elucidating the necessity to include suitable geometries in the calculations. For example, the consideration of a S-S bond between C200 and C289 (overstretched in the X-ray structure of the free protein) was found to yield a better agreement with experimental data. This work provided a solid benchmark for the computational study of this protein. In their second study, the authors investigated the effect of different mutations in RPA70A/B on ssDNA binding using again a computational approach. Here they used the MM/PBSA method over MD trajectories of different mutants-DNA complexes to estimate the energetic differences with the native structure. Again their findings were in good agreement with the experimental data and provided structural insights of this interaction [139]. In Brosey et al.'s

work [133] MD simulations of various RPA and RPA-ssDNA complexes were run in implicit solvent. The DNA binding domain of RPA that was modeled in this study contains residues from RPA70A, B, C and RPA32D. These simulations, combined with X-ray and neutron scattering experiments, provided a novel model where the free form of RPA is flexible and the domain motions are uncorrelated, becoming more compact after ssDNA binding. This proposed mechanism differed from the previously proposed ones also for the absence of an intermediate state between before and after the binding occurs. The study provided a new framework for understanding the action of RPA in incorporating the DNA into the processes. Pavani et al. employed homology modeling, MD simulations and sequence alignment to identify RPA70 from *Leishmania amazonensis* (LaRPA1) as structurally different from other eukaryotic RPA proteins. Their findings support a role of LaRPA1 as telomere end-binding protein [140]. Compared with double-stranded DNA (dsDNA), the prediction of protein-ssDNA complexes is usually difficult due to the high flexibility of the single stranded chain and the heterogeneity of the interaction interface. In a recent work of Mishra and Levy [141], a coarse-grained model was developed and tested to overcome this limit and predict the correct protein-ssDNA complex. The molecular interactions dominating the model potential were the electrostatic interactions established between basic residues of the protein and the DNA phosphates, and the hydrophobic stacking between aromatic residues and DNA bases. The ssDNA was placed at different starting positions and the complex formations were simulated. The final conformations were then compared with the crystallographic structures. The method successfully reproduced six crystallographic complexes, including RPA and ssDNA, opening a new pathway to model these protein-DNA interactions for which an experimental structure is not available. Noticeably, it allows also to explore different binding modes depending on the salt concentration and to observe the binding

process of ssDNA to the protein, as shown in Figure 3.2. Besides NER, RPA is involved in multiple other cellular processes, such as HR and checkpoint processes. During the intra-S checkpoint the RPA32C domain is interacting with the TIMELESS-interacting protein (TIPIN). A homology model of TIPIN was built by Ali et al. [142] and then docked to the RPA domain. Their findings elucidated a binding mechanism occurring at a common interface but different to other proteins binding at the same site, such as XPA and UNG2, where in addition to the helix interaction an interaction with an N-terminal loop is also established. The results of this work supported a role of RPA32C as an exchange point for many proteins involved in different activities within the cell.

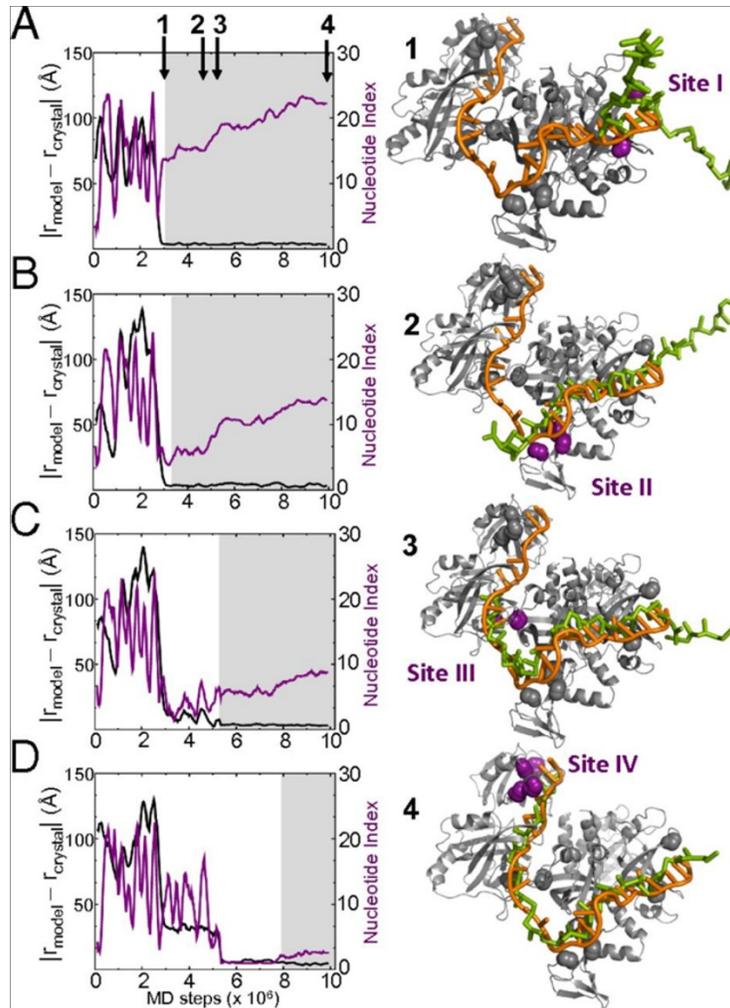


Figure 3.2. Simulation of binding mechanism between ssDNA and RPA. The CG model for RPA-ssDNA interaction not only permits to reproduce the crystallographic complex, but also to simulate the binding process, where DNA gradually interacts with four sites through a sliding process. The approaching of the DNA with each binding site is represented on the left, where the purple line indicates the interacting nucleotide index and the black one the difference between the site-nucleotide minimal distance in the model and the crystal structure. A-D refers to binding site I-IV. The four snapshots occurred at the simulation times indicated by the respective arrow. Adapted with permission from [141].

XPA is a 31 kDa, 273 amino acid long protein involved in damage recognition during NER. The C-terminus binds with TFIIH, whereas the N-terminus interacts with RPA32 and ERCC1. The

central domain is dedicated to recognize the damaged DNA through a zinc finger motif as well as to interact with RPA70 [143]. Its interaction with ATR after UV irradiation is required for a good repair efficiency. ATR phosphorylates XPA at S196, present in a helix-turn-helix (HtH) motif within the DNA binding domain. Molecular modeling and electrostatic calculations allowed to observe the change of charge distribution after this happens. Accumulation of negative charges was detected on the computational model on the HtH motif after the phosphorylation [144]. The interaction between XPA and ERCC1 is essential for the NER activity. Barakat et al. employed MD simulations of an XPA binding peptide-ERCC1 complex and free energy decomposition (MM/PBSA) to study the XPA residue contributions to the binding with ERCC1. F75 was stacked against R110 during the simulation and was found to be the highest XPA contributor to the binding free energy [145]. Fadda's works carefully investigated the nature of the interaction between the XPA binding motif and ERCC1, as well as the recognition process. In her first paper [146], MD simulations of the XPA's ERCC1-binding peptide (native and two mutant forms) free or in complex with ERCC1 were performed. The investigated mutants were F75A and D70A. The 500 ns-long simulations of the wild form were performed for two different protonation states of ERCC1's H149 (N $\delta$ 1 and N $\epsilon$ 2). N $\epsilon$ 2 was found to be the optimal state, in contrast with the NMR structure. Direct residue interactions were identified for the wild type as H149-D70, Y145-D70, R156-G73, Q107-G73, S142-G74 and N110-F75, where the first residue belongs to ERCC1 and the second to XPA. Changes introduced by the mutated peptides were studied, as well as the XPA free form and the effect of the protonation state of H149 of ERCC1 in complex with these forms was assessed. In her second paper on the topic [147], MD simulations on a  $\mu$ s time scale suggested a recognition model of XPA by ERCC1. The motif that was thought to be responsible for the recognition was

a hairpin prominently represented during the simulations (36% of the time over 10  $\mu$ s) and similar to the one of the bound form of XPA. The superposition of the representative structures with the experimental one is reported in Figure 3.3. The same motif was less prominent during the simulation of mutant forms, explaining the decreasing of binding affinity observed by experiments and providing a strong theoretical base for the molecular recognition process of XPA by ERCC1. Another work [148] adopted a combination of bioinformatics and molecular modeling tools to study the effect of mutations of the XPA gene. Firstly, the authors employed the sorting intolerant from tolerant (SIFT) bioinformatics tool to predict deleterious XPA mutations, identifying the C108F as the most deleterious one. The authors then ran MD simulations of the native and mutant form of XPA. Root mean square deviation and root mean square fluctuation (RMSD and RMSF, respectively) comparisons demonstrated a strong structural change in the mutant form, validating the in mutation as deleterious for XPA functions.



Figure 3.3. Superposition of predicted XPA binding motif and NMR structure of bound XPA peptide. The representative structures of the most populated clusters of the XPA peptide, obtained over 10  $\mu$ s seconds of MD simulations of the XPA free form, are represented in cyan and superposed to the NMR bound structure of XPA peptide (red). ERCC1 central domain is represented in grey. The similarities between the hairpin motif and the experimental bound one provided evidences of this structure being responsible for XPA recognition by ERCC1 in Fadda's work. Adapted with permission from [147].

The endonuclease cutting 5' the lesion is a heterodimer formed by XPF and ERCC1, where ERCC1 is the non-catalytic part of it. It associates with XPF through interactions between the (HhH)<sub>2</sub> motifs present in the C-terminals. The protein is composed by 297 amino acids (32 kDa) with two noticeable domains, the C-terminal domain including the (HhH)<sub>2</sub> motif and a central domain. The XPF-ERCC1 complex binds two ssDNA arms with the two HhH<sub>2</sub> domains of XPF and ERCC1 and the 5'-ssDNA arm with the central domain of ERCC1. The central domain interacts also with XPA [121]. XPF contains the catalytic domain of the enzyme. The protein includes 916 amino acids (104 kDa) divided into three domains, the C-terminal domain involved

in the ERCC1 interaction through the (HhH)<sub>2</sub> motif, the active nuclease domain and a helicase-like domain [149]. In an early work [150] the authors used the Eph receptor sterile  $\alpha$  motif (SAM) domain and the UvrC C-terminal as templates for the homology modeling of ERCC1 and XPF binding motifs. Then docking simulations predicted the structure of the heterodimer, subsequently found to be in agreement with another homology model built using the archaeal XPF homodimer as template. The already cited Barakat et al.'s paper [145] provided insights into the ERCC1 residues involved in XPA binding as well (Y145 was shown to be the highest contributor in terms of the binding free energy) and a detailed description of the central domain of ERCC1. The same authors produced a similar work [84] which studied the ERCC1-XPF interaction using MD simulations of the XPF-ERCC1 C-terminals and MM/PBSA calculations. F293 was identified as the most contributing ERCC1 residue towards the binding energy; on the other hand F894 was the highest contributor for XPF. This work investigated in detail the interface between the two proteins aiming to identify druggable sites in order to target it pharmacologically. Similarly to what they have done for XPA, Priya Doss and Nagasundaram's study [151] explored the conformational changes introduced by predicted deleterious mutations in the ERCC1 gene. Three ERCC1 polymorphisms (V116M, R156Q and A199T) were docked to XPF and simulated using MD, in addition to the native form. Solvent Accessible Surface Area (SASA), RMSD/F and hydrogen bond analysis elucidated the effect of these mutations on the protein-protein interaction.

### **3.5 Computer-Aided Drug Design for NER Repair Modulators**

The involvement of NER in repairing DNA damage caused by anticancer drugs prompted the development of inhibitors for this pathway. There are several recent works that employed *in-silico* screening of databases to search for novel drugs that can target essential protein-

protein/DNA interactions or the activity of NER enzymes. As described above, RPA is involved in multiple processes, including ssDNA binding and protein-protein interactions for DNA damage signaling and repair. The binding site and mode of action of compound 505, an inhibitor of the ssDNA binding to RPA, was predicted using docking simulations [152]. The authors of the study targeted three possible binding sites within the RPA70 DNA-binding domain (DBD), identifying the most probable one (DBD-B). A validation with an inactive compound was also performed to confirm their findings. In another study, the binding to RPA70 N-terminal by NSC-15520 (an inhibitor of functionality of RPA70N discovered using *in-vitro* screening) was demonstrated to be competitive with Rad9, essential for DNA damage signaling, through molecular docking simulations [153]. The same authors used MD simulations [154] to generate an ensemble of configurations of the RPA DNA-binding zone interacting with the trans-activation domain of p53, the DBD-F domain. The work produced a novel representation for the RPA70N DNA-binding domain that was demonstrated to be more accurate than the crystallographic data by using docking simulations of the p53 domain and fumaropimaric acid (an inhibitor for the DBD-F domain) to the MD-generated structures.

XPA is a central protein in NER and therefore represents an attractive target for CADD studies. The interaction with ERCC1, RPA and DNA are the key activities influencing the function of this protein. Neher et al. [155] targeted the DNA-binding zone of XPA (loop 2 of the C-terminal) with VS for 200,000 compounds, which were docked to the binding domain. A two-phase scoring protocol followed by clustering and visual inspection was employed, resulting in a hundred of compounds for *in-vitro* testing. One compound shows  $\mu\text{M}$   $\text{IC}_{50}$  in inhibiting XPA binding with ssDNA and dsDNA (both intact and cisplatin-damaged).

The XPF-ERCC1 heterodimer is another key protein for NER, as it is an endonuclease cleaving the 5' extremity of the lesion. For this reason, the disruption of the dimer or the inhibition of the catalytic activity may result in the complete shutting down of NER activities. Barakat et al. focused on the XPA-ERCC1 interaction, through which the endonuclease is recruited to the damaged zone recognized by XPA. In their first study [145] a library composed of about 3,500 compounds was virtually screened against the ERCC1's XPA binding domain. The NCI Diversity Set (NCIDS) [68] and the DrugBank databases were merged to obtain the initial library. A set of representative structures for the free form of ERCC1 was extracted from exhaustive MD simulations and used for flexible docking. The top ranked hits were then used to build a pharmacophore model for XPA-ERCC1 inhibitors. In the second work produced by the authors [80], the NMR structures of ERCC1, XPA-bound form, were used for flexible docking. In addition, MD refinement and MM/PBSA rescoring were used to reduce the number of false positives from the *in-silico* screening. The Chimiotèque Nationale (CN) database (~50,000 compounds) was screened in this work. One compound showed encouraging activity in sensitizing cancer cells to UV irradiation and synergy with cisplatin treatment. The same authors focused also on the inhibition of XPF-ERCC1 protein-protein interaction to disrupt the active form of the endonuclease [84]. After MD simulations and analysis of putative binding pockets, they decided to target one site within the C-terminal domain of XPF. Again a RCS approach was employed to flexibly dock and score the NCIDS and DrugBank repositories to the targeted binding pocket. Compound labelled as NSC-130813 showed synergy with DNA damaging agents in cancer cells, resulting in a sensitizing effect to UV irradiation and ability to disrupt the XPF-ERCC1 interaction. The orientation of this molecule within the XPF binding site and its chemical structure are reported in Figure 3.4. Finally, a recently published study has employed

VS in addition to traditional *in-vitro* screening to identify novel selective inhibitors for this endonuclease. McNeil et al. [85] employed a multistep docking protocol to screen a library composed of several commercial databases (~5,000,000 structures in total, including different conformers). A combination of structure-based and ligand-based VS techniques and the application of a consensus scoring scheme were employed to obtain a set of potential inhibitors. The authors targeted both the XPF active site (obtained through homology modeling) and the XPF-ERCC1 complex on three ERCC1 interacting sites on XPF. Regarding the catalytic activity, two compounds obtained from VS showed a moderate inhibitory effect in the endonuclease assay. On the other hand, one compound showed activity in a NER assay as XPF-ERCC1 inhibitor and enhanced cancer cell sensitivity to cisplatin. The reader is referred to

Table 3.1 for the details about these *in-silico* discovered NER inhibitors.

Table 3.1. *In-silico* discovered NER protein inhibitors. The details about the experimentally measured activities are reported as well. If not indicated, IC<sub>50</sub> values refer to the best obtained in the related experiment. See text and referred papers for more details.

<b>Compound(s)</b>	<b>Target</b>	<b>Database</b>	<b>Activity</b>	<b>IC<sub>50</sub> (μM)</b>	<b>Reference</b>
X80	XPA	ChemDiv	DNA binding <sup>1</sup>	28	[155]
AB-00026258	ERCC1 <sup>2</sup>	CN	Cytotoxicity in cancer cells	4.80	[80]
NSC-130813	XPF <sup>3</sup>	NCIDS	Cytotoxicity in cancer cells	0.79	[84]
E-X PPI2	XPF <sup>3</sup>	Specs	NER in cancer cells	20	[85]
E-X AS1	XPF	Enanime and	Endonuclease	18.30-	[85]

E-X AS2		Asinex		27.60	
---------	--	--------	--	-------	--

<sup>1</sup>Refers to dsDNA containing cisplatin damage.

<sup>2</sup>XPA-ERCC1 interaction.

<sup>3</sup>XPF-ERCC1 interaction.

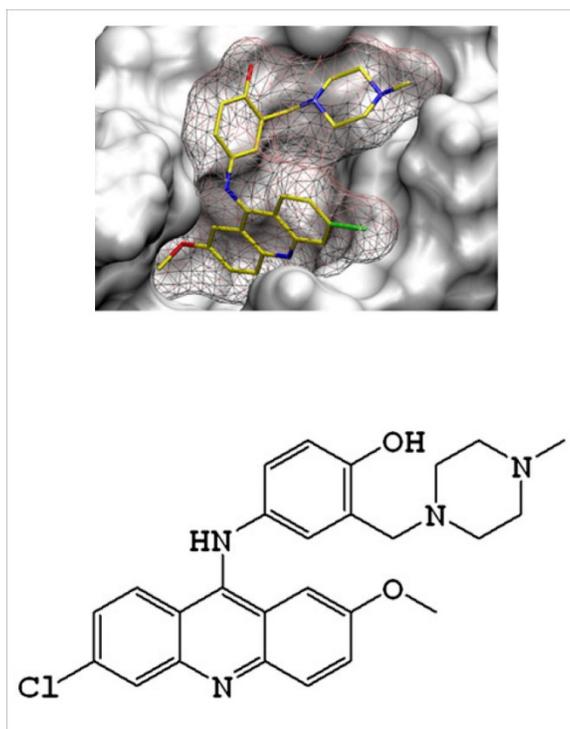


Figure 3.4. Binding mode and chemical structure of NSC-130813. This small molecule binds to XPF, inhibiting the XPF-ERCC1 protein-protein interaction. This compound has been discovered with a structure-based VS protocol targeting the XPF C-terminal by Jordheim et al. Adapted with permission from [84].

### 3.6 Conclusions

The DNA repair mechanisms constitute an interesting, still largely under-explored field of study in cell biology and molecular medicine. We are far from completely understanding the complex network of protein interactions and signaling involved in these intricate processes. However, several efforts have been recently made to investigate the key steps of the repair pathways. The study of these mechanisms is in fact important to understand how the human body is protected

against DNA damaging agents and the consequences occurring when this protection is deficient. Inhibition of DNA repair in cancer cells can be used as adjuvant therapy for treating cancer, with the goal to enhance the effect of standard chemotherapy and radiation therapy. On the other hand, targeting repair defects in tumor cells may be a selective synthetic lethal approach to be adopted in cancer therapy, as indicated in recent advances in research on this topic. In this context, BER and NER are two interesting pathways to target for cancer therapy.

Computational approaches, ranging from micro-scale (QM) to meso-scale (CG), have been largely employed to investigate major aspects of DNA repair pathways. We have reviewed studies that modelled the biochemical reactions of the most important endonucleases, as well as simulations of biological complexes elucidating the role of structural effects introduced by mutations. We have also critically reviewed a large number of reported studies where novel modulators of NER pathway have been discovered using *in-silico* protocols. The insights provided by these efforts constitute a framework for the future introduction of efficacious combination therapies for cancer treatment. The outstanding results of these studies are supportive of a continuously increasing role played by computational modeling in the investigations of different aspects of DNA repair pathways (and biological processes in general) at different scales. Arguably, once these methods reach a satisfactory level of accuracy and reproducibility, biomedical research will benefit from an enormous saving of time and economic resources in an effort to obtain specific and selective results in the search for optimized therapeutic agents.

In spite of these encouraging consideration, there is still a lot of room for improvement in the development of computational techniques. In the specific study of NER and other DNA repair pathways, many key points have not been sufficiently and accurately investigated yet. It also

appears that a real therapeutic solution is still far away from clinical applications. For example, the mechanisms of DNA cleavage have been just partially investigated, which is also true for the key protein-protein/DNA interactions dominating the repair machinery. Moreover, although for some targets (PARP1 for example) a lot of inhibitors have been revealed, many repair proteins are difficult to selectively target due to the structural similarities with others involved in different processes. Future studies focusing on these aspects may result in more effective pharmacological solutions, with reduced side effects of these novel therapies.

More efforts are required to extend the spatial-temporal scales of simulations or, alternatively, to develop more efficacious sampling techniques. Also, there is a need for more accurate and selective structure/ligand-based drug design techniques. Serious limitations involve the lack of computing power. With the continuous development of innovative computational hardware and software, it is likely that simulations that today are prohibitive on a large scale in terms of computational cost will be totally accessible in the next few years, starting a new era of biological system modeling.

## Chapter 4 - New Design of NER Inhibitors for Combination Cancer Therapy<sup>3</sup>

### 4.1 Abstract

Many cancer chemotherapy agents act by targeting the DNA of cancer cells, causing substantial damage within their genome and causing them to undergo apoptosis. An effective DNA repair pathway in cancer cells can act in a reverse way by removing these drug-induced DNA lesions, allowing cancer cells to survive, grow and proliferate. In this context, DNA repair inhibitors opened a new avenue in cancer treatment, by blocking the DNA repair mechanisms from removing the chemotherapy-mediated DNA damage. In particular, the nucleotide excision repair (NER) involves more than thirty protein-protein interactions and removes DNA adducts caused by platinum-based chemotherapy. The excision repair cross-complementation group 1 (ERCC1)-xeroderma pigmentosum, complementation group A (XPA) protein (XPA-ERCC1) complex seems to be one of the most promising targets in this pathway. ERCC1 is over expressed in cancer cells and the only known cellular function so far for XPA is to recruit ERCC1 to the damaged point. Here, we build upon our recent advances in identifying inhibitors for this interaction and continue our efforts to rationally design more effective and potent regulators for the NER pathway. We employed *in-silico* drug design techniques to: 1) identify compounds similar to the recently discovered inhibitors, but more effective at inhibiting the XPA-ERCC1 interactions, and 2) identify different scaffolds to develop novel lead compounds. Two known inhibitor structures have been used as starting points for two ligand/structure-hybrid virtual screening approaches. The findings described here form a milestone in discovering novel

---

<sup>3</sup> This chapter has been published as research article as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy. *J. Mol. Graph. Model.* 2016, *1*, 71–82.

inhibitors for the NER pathway aiming at improving the efficacy of current platinum-based therapy, by modulating the XPA-ERCC1 interaction.

## **4.2 Introduction**

Cancer is a set of diseases whose main cellular result is the transformation of normal cells into anomalous cells, which grow, proliferate, undergo uncontrolled division, and often invade the organism by metastases. The disease is affecting millions of patients worldwide, raising immense economical and social burdens on the patients and their families. DNA damage is among the major factors that contribute to cancer initiation and progression [156]. Several exogenous and endogenous agents can interact with the DNA, leading to critical mutations in the human genome. This includes reactive oxygen and nitrogen species [157], UV radiations [158] and polycyclic aromatic hydrocarbons [159]. To protect the genome from such frequently occurring lesions, the cells have developed an intricate enzymatic network of DNA repair mechanisms to detect, isolate and repair the damaged part of the DNA chain, allowing for proper genome replication and correct cell division and differentiation [90].

The nucleotide excision repair (NER) pathway is responsible for removing bulky DNA adducts induced by UV radiation, external agents, lipid peroxidation or reactive oxygen species [90]. These adducts distort the helix structure, halting the replication cycle and inducing apoptosis [160]. While these adducts could cause severe damage to normal cells, inducing them is, in fact, a desirable and promising way to destroy cancer cells. In this context, platinum-based therapy (e.g. cisplatin) has been developed to deliberately crosslink the DNA molecule, causing bulky lesions in cancer cells resulting in apoptosis. However, in this case, an over-activated NER pathway can act in a reverse way to this therapy, reducing its efficacy and benefit [91]. A

relatively new promising approach is to design selective NER inhibitors to potentiate the efficacy of platinum-based therapy in a synergistic fashion.

NER involves at least thirty proteins that interact with each other at different stages of the pathway. Among these proteins is the excision repair cross-complementation group 1 (ERCC1) protein. ERCC1 is over-expressed in many cancer cells and its high expression is also correlated with resistance to cisplatin treatment [129,161,162]. Targeting this protein has been demonstrated to enhance the sensitivity of cancer cells to cisplatin [100,163]. ERCC1 forms a heterodimer endonuclease complex with the excision repair cross-complementation group 4 (XPF) protein. The ERCC1-XPF endonuclease cleaves the damaged DNA strand at the 5' position [121,122]. The endonuclease complex is recruited to the damage through a specific interaction between ERCC1 and another member of the NER proteins, namely xeroderma pigmentosum, complementation group A (XPA) protein [164]. Besides ERCC1, also the expression level of XPA has been shown to directly influence the response to cisplatin treatment [128,165,166]. Interestingly, the XPA-ERCC1 interaction is specific, essential to NER and no other functions have been known for XPA apart from the recruitment of the ERCC1-XPF complex [120]. The schematic representation of NER is reported in Figure 3.1 (adopted from the KEGG database [125,126]).

Two different strategies to regulate the NER pathway through the XPA-ERCC1 interaction have been proposed so far. The first employs a short peptide that mimics the XPA binding domain to ERCC1, which has the ability to compete with the full-length XPA protein for binding to ERCC1 [164]. The second approach, which we follow here, is to use small molecules binding to ERCC1, preventing its interaction with XPA. The first proof of concept for the latter strategy was provided through the cell cycle checkpoint abrogators, such as UCN-01, which was shown

to inhibit the XPA-ERCC1 interaction [167]. Following that, our group rationally identified a set of novel XPA-ERCC1 inhibitors by modelling the interaction and employing a sophisticated relaxed complex scheme (RCS) docking approach. These efforts led to the discovery of NERI01, a selective inhibitor of the XPA-ERCC1 interaction. This inhibitor was successfully validated experimentally as a regulator of the NER pathway in cancer cell lines. Another similar compound (compound 10) showed a less potent but still interesting modulator activity [80].

Here, we build upon these discoveries aiming at identifying chemically similar but biologically more effective molecules as well as new scaffolds for the XPA-ERCC1 inhibition. In this work, we employed both ligand-based and structure-based virtual screening techniques to identify novel potential inhibitors for the XPA-ERCC1 interaction. On one hand, we concentrated our efforts on searching a large library of compounds similar to the two previously identified lead structures, in order to find more potent structures. On the other hand, we developed a hybrid structure/ligand-based pharmacophore model to screen a library containing diverse structures to identify novel scaffolds that can be further optimized to obtain additional structures inhibiting the interaction.

## **4.3 Results and Discussion**

### *4.3.1 Similarity Search-Based Virtual Screening*

The structures of the two lead compounds (hereafter referred as NERI01 and compound 10) are depicted in Figure 4.1. Both the compounds were able to sensitize cancer cells (especially colon cancer) to UV radiation and NERI01, in particular, was demonstrated to act in synergy with cisplatin. Moreover, the direct binding of both structures to ERCC1 was also validated [80]. The traditional similarity search has been performed using each structure as input; for NERI01, there

were 8,848 compounds from the PubChem database (containing ~68,370,000 compounds) showing a Tanimoto score equal or higher than 0.80, whereas 1,401 have been obtained for compound 10. The turbo similarity search method applied to the top 100 neighbors of the lead structures resulted in 5,937 compounds for NERI01 and 5,798 for compound 10. In total we extracted 21,984 unique structures from PubChem, of which 10,250 came from the traditional similarity search and 11,734 from the turbo approach. The number of unique small molecules has been reduced to 19,828 due to the filtering step in ADMET Predictor. Finally, around 28,600 structures have been considered for protein docking, including various conformations as well as the ionizations states calculated with LigPrep.

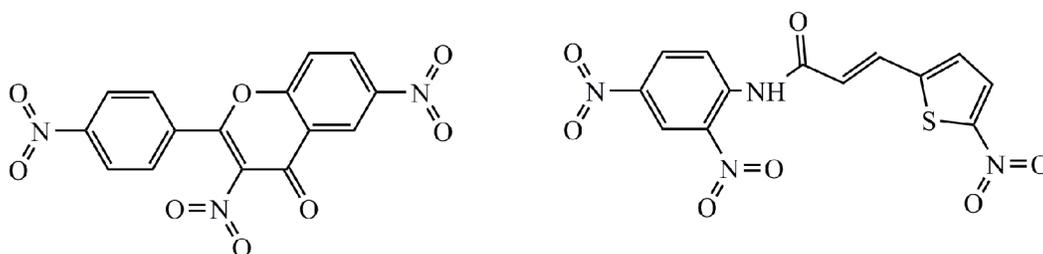


Figure 4.1. Chemical structures of NERI01 (left) and compound 10 (right). The two compounds share similar chemical features, in particular a hydrophobic backbone with two nitro groups at the extremities.

#### 4.3.2 Pharmacophore-Based Virtual Screening

The docked poses of NERI01 and compound 10 on ERCC1 structure #7 (see below) have been used to generate a pharmacophore model. Out of the twenty Consensus Query suggested common features, five of them have been selected for the pharmacophore model, including three hydrogen bond acceptors and two hydrophobic features (Figure 4.2). The choice of the features as well as the size have been manually adjusted based on the acceptor and hydrophobic features

maps calculated as iso-surfaces at -2 kcal/mol and the excluded volume calculated on the binding poses.

The four libraries from the NCI repository contained 2,642 unique molecules. Tautomer and conformer generation resulted in 92,340 entries that have been successively screened using the pre-generated pharmacophore query. The rank of top hits from the pharmacophore search was obtained by scoring their conformations based on the root-mean-squared deviation (RMSD) of their annotation points using the pharmacophore features' model as a reference. A total of 126 unique hits have been identified at the end of the screening, where just the conformation with the lowest RMSD value was considered for each ligand. At this stage, we decided to proceed with a refinement of the results using structure-based techniques for two reasons. First, although pharmacophore generation involved the target structures, the ligand conformer search step, performed in MOE, used a random search technique without considering the target binding site. Therefore, it was possible that some of the predicted configurations would not be physiologically relevant. Second, it was not possible to obtain an energetic score to rank the binding affinities of potential hits using only the pharmacophore search protocol adopted in MOE.

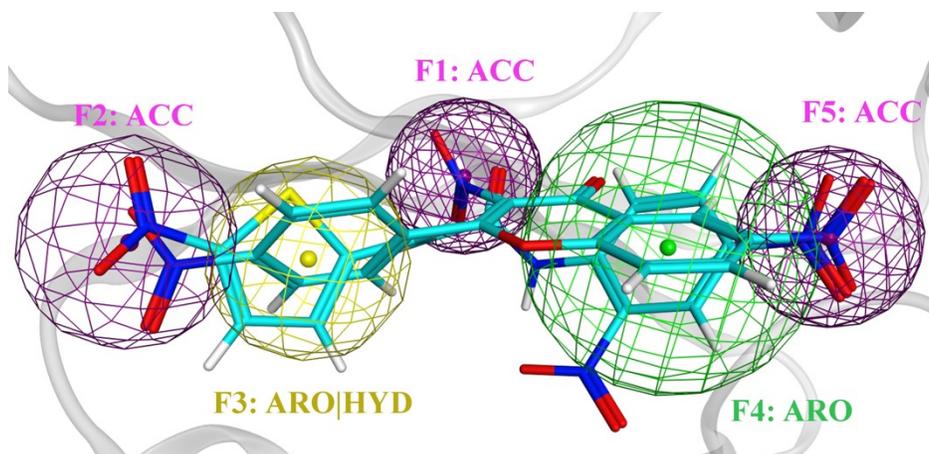


Figure 4.2. The ligand/structure-based pharmacophore model built in MOE. ACC means acceptor feature, HYD means hydrophobic feature and ARO means aromatic feature. The radius and the tolerance of the features have been adjusted in order to satisfy the receptor-based electrostatic maps, calculated in MOE (see text for more details).

#### 4.3.3 Molecular Docking

Ten XPA-ERCC1 structures have been deposited in the Protein Data Bank. Each model includes the central domain of ERCC1 (residues 99-214) interacting with the small peptide from the XPA binding domain (residues 67-77). The ERCC1 residues that energetically contribute the most to the interaction with XPA have been previously identified as Arg106, Gln107, Gly109, Asn110, Pro111, Phe140, Leu141, Ser142, Tyr145 and Tyr152 [145]. From Figure 4.3 it is possible to observe the flexibility of these key residues among the different NMR structures even after the relaxation of the side chains in explicit water.

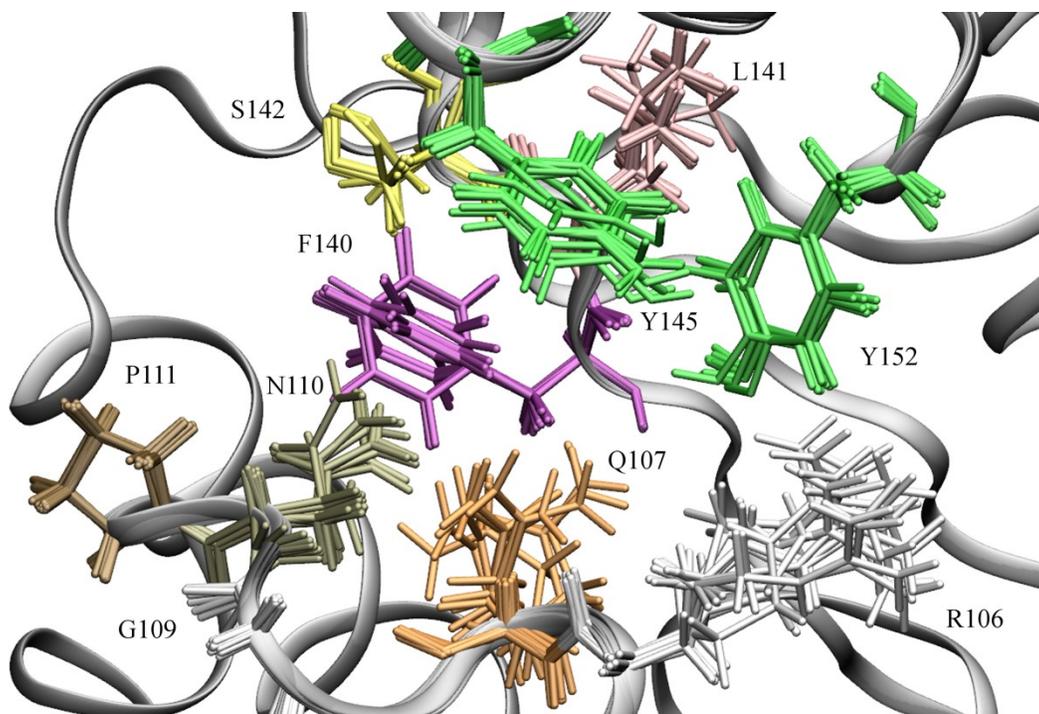


Figure 4.3. Superposition of the NMR configurations of ERCC1 key residues. This set of ten residues is mainly responsible of the interaction with the XPA peptide. The residues are colored according to their names.

Because the number of structures obtained at the end of the similarity search was quite high (more than 28,000), we decided to perform the docking studies on this library in two steps in order to employ the RCS, that is performing rigid docking simulations on multiple relevant conformations of the same target [51,145,168,169] and using a smaller set of molecules. The first step was therefore to dock the whole library of compounds using one target structure and select the top hits for the flexible receptor docking protocol. From the ensemble of available NMR conformations (PDB entry 2JNW, containing ten models), we selected the seventh conformation. We initially docked the two lead structures to all the ten conformations and evaluated the twenty resulting poses. The seventh structure showed the best binding affinities among the ensemble with both the two known inhibitors, as well as complementary orientations

of the poses with respect to the pocket features. The selection of the snapshot showing the lowest binding energy with an active compound as a single starting structure for high-throughput docking has been successfully tested previously [170]. The binding energy predicted by the Autodock scoring function of NERI01 was -8.07 kcal/mol while for compound 10 it was -8.58 kcal/mol. The two nitro groups at the two extremities of the structures allow to establish polar interactions with the Arginine residues of the binding pocket (Arg106, Arg144 and Arg156). Another noteworthy feature is the stacking of the aromatic rings of the two compounds between the hydrophobic bottom of the binding pocket and the aromatic rings of Tyr145 and Tyr152. Finally, the polar zones of the binding pocket are in proximity of the central polar groups of the ligands (Figure 4.4). Therefore, we performed the docking of all the compounds using structure #7 as the target. The range of the obtained binding energies was between -10.27 kcal/mol and -4.09 kcal/mol. We then selected the top 25% unique hits (4,590 structures including different conformations and tautomers) to be docked using the Relaxed Scheme Complex protocol. The binding energy results in this case ranged between -10.54 kcal/mol and -4.32 kcal/mol. After a refining step of visual analysis we selected only 200 compounds for further, more detailed analysis.

62% and 29% of these 200 hits were respectively retained from the traditional and the turbo similarity searches using NERI01 as input. The same indexes were 1% and 8% when compound 10 has been used; in this last case the turbo approach contributed more than the traditional one to the final set.

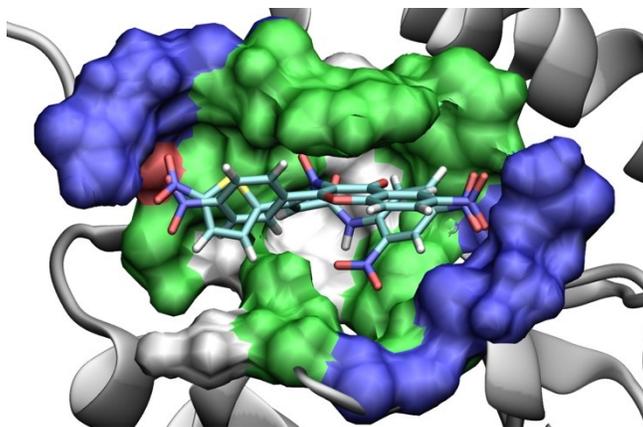


Figure 4.4. Binding poses of NERI01 and compound 10 in the binding site of structure #7. The polar extremities of the two compounds are interacting with the Arg residues delimiting the binding site (Arg106 and Arg144). The hydrophobic backbone is stacked between the aromatic layers constituted by the hydrophobic residues of the binding pocket and Tyr145, 152. The surface of the binding pocket is colored depending on the type of residues, in white where hydrophobic, in green where polar, in blue where basic and in red where acidic.

For the hits that resulted from the pharmacophore-based virtual screening, we performed docking simulations using only the ERCC1 structure #7 as a target, as we have designed and refined the pharmacophore model using this structure (see above). The predicted binding affinities following this approach ranged from -8.79 kcal/mol to -4.80 kcal/mol. After the visual inspection to identify the best solution for each ligand, we selected only thirteen structures for further analysis.

#### *4.3.4 Molecular Dynamics Simulations and MM/PBSA Rescoring*

Although docking methods have been shown to well predict the binding pose of a ligand, their scoring functions usually lack in accuracy in favor of speed and reduced computational cost. Moreover, just the final docked pose is considered for the energy calculation, without accounting for the natural dynamicity of the binding. For these reasons, we performed 2ns-long molecular

dynamics simulations of the complexes, applying the MM/PBSA method to the resulting trajectories for the estimation of the binding energies.

The entropic contributions calculated using normal mode analysis over a diverse sample of the identified hits were in the same order of magnitude. Therefore, entropy was neglected for the rest of the compounds (Table A.1 in Appendix A) and, consequently, we relied mainly on the relative binding energies of the compounds. This is consistent with the fact that all the identified hits came originally from a similarity search approach (i.e. either using Tanimoto or pharmacophore-based methods), consequently they all share similar structural features and the entropy can be neglected for the final ranking purposes as it will be cancelled out when calculating the relative binding energies [171]. The computed binding energy for NERI01 was -23.24 kcal/mol, whereas for compound 10 it was -24.86 kcal/mol. The binding pose of NERI01 is shown in Figure 4.5A. The two nitro groups established hydrogen bonds with Arg144 and Arg106, where the negatively charged oxygens interacted with the positive charges of the Arg. The central polar groups formed other hydrogen bonds with Ser142 and Gln107. The visualization of the trajectory revealed possible interactions also with Arg156, which was however involved in a solid salt bridge with Asp129 and in a hydrogen bond network with the surrounding residues. It is worth to note that these hydrogen bonds, especially the ones involving the Arg residues were disappearing and recreating continuously during the trajectory, often substituted with other hydrogen bonds or mediated by water molecules. The hydrophobic rings were stably stacked within the hydrophobic cleft created by the Tyr rings (Tyr145 and Tyr152), the backbones of Arg106 and Gln107, the interaction between the side chains of Asn110 and Phe140, and Pro111. Compound 10 showed a similar binding mechanism, dominated by hydrophobic stacking rather than really stable electrostatic interactions.

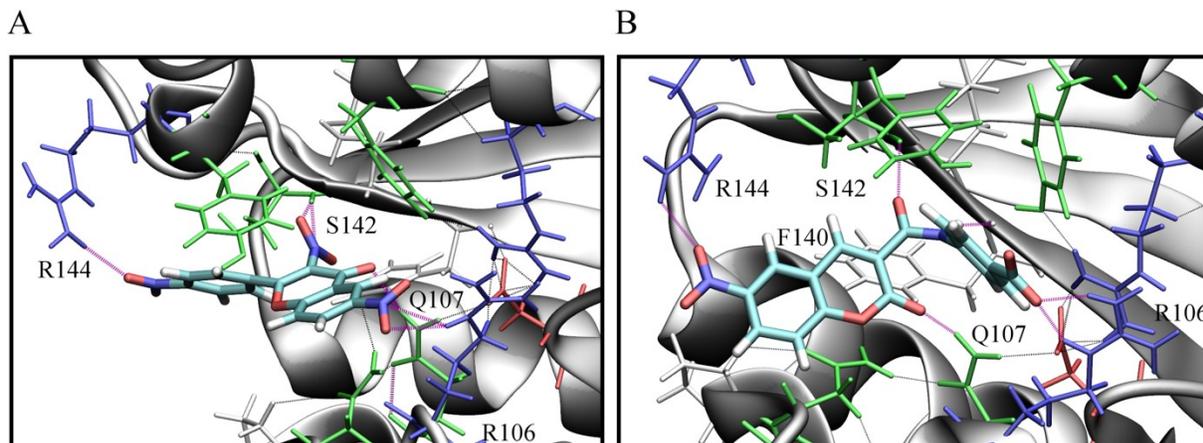


Figure 4.5. Predicted binding poses of A) NERI01 and B) 6971912. Hydrogen bond networks are also represented. The nitro/positive groups can interact with the positive charged parts of the Arg residues. The hydrogen bond acceptors of the ligands interact with Ser142 or Gln107.

Sixty-nine compounds from the similarity set showed an MM/PBSA score more favorable than the original lead compounds. The top hits shared a common hydrophobic backbone composed of aromatic rings which included at least one polar atom. Additionally, one or two polar terminals are usually present in most of the identified structures and are often represented by negatively charged groups. In particular, at least one nitro group was always conserved in the highly ranked hits, highlighting the importance of this moiety in this class of inhibitors. By visually analyzing the top hit complexes, we identified a number of common binding features, as expected. Noticeably, the hydrophobic rings constituting the backbone of the compounds were stably stacked within the hydrophobic cleft of the binding pocket, constituted by the rings of Tyr145 and Tyr152, the interactions between Asn110 and Phe140, the backbones of Arg106 and Gln107. We also observed conserved interactions between polar/charged groups of the ligands and the basic zones of the binding site, especially with Arg106. The polar atoms of the backbone were able to interact with the central polar zone of the binding pocket, constituted by Ser142 and Gln107. During the simulations, we often observed the cyclic disruption and reconstruction of

the hydrogen bonds. The top fifteen structures are reported in Figure 4.6 and the energetic scores are reported in Table 4.1. The contributions to the binding energies of the residues in the binding pocket are reported in Figure 4.7A. A common interaction pattern can be identified among the hits and the lead compounds. For example, Tyr145 was reported as the highest contributor to the binding with the XPA peptide by Barakat et al. [145]. The residue was the highest contributor for the binding of all the ligands as well. Arg106 and Gln107 constituted another group of conserved interacting residues, as well as Phe140, Leu141, Ser142 and Tyr152. Interestingly the interactions with Asp129 were often unfavorable, because the presence of the negative charge on the residue and the negative charged atoms of the small molecules (e.g. the oxygens of the nitro group). The ligand binding was dominated by the van der Waals interactions due to the hydrophobic stacking of the ligand backbone's rings, as observed also for the lead compounds. Arg106, Gln107, Asn110, Phe140 and Tyr145 were in particular involved in hydrophobic interactions (Figure 4.7B). The pattern of these interactions reflected the one of the total energy, and unfavorable contributors were not observed. A more fragmented pattern was observed for the electrostatic contributions (Figure 4.7C), consistently with the instabilities observed during the simulations. Arg106 was in general the most important contributor for these terms. Due to its positive charge, the residue was indeed involved in hydrogen bonds with partial negative charges of the hits in many cases. As expected and already discussed, the Asp129 electrostatic contribution was unfavorable in many cases. Noticeably, most of the newly discovered hits showed an attenuated unfavorable electrostatic interactions established with Arg108, Asn110 and Phe140, when compared to the lead compounds.

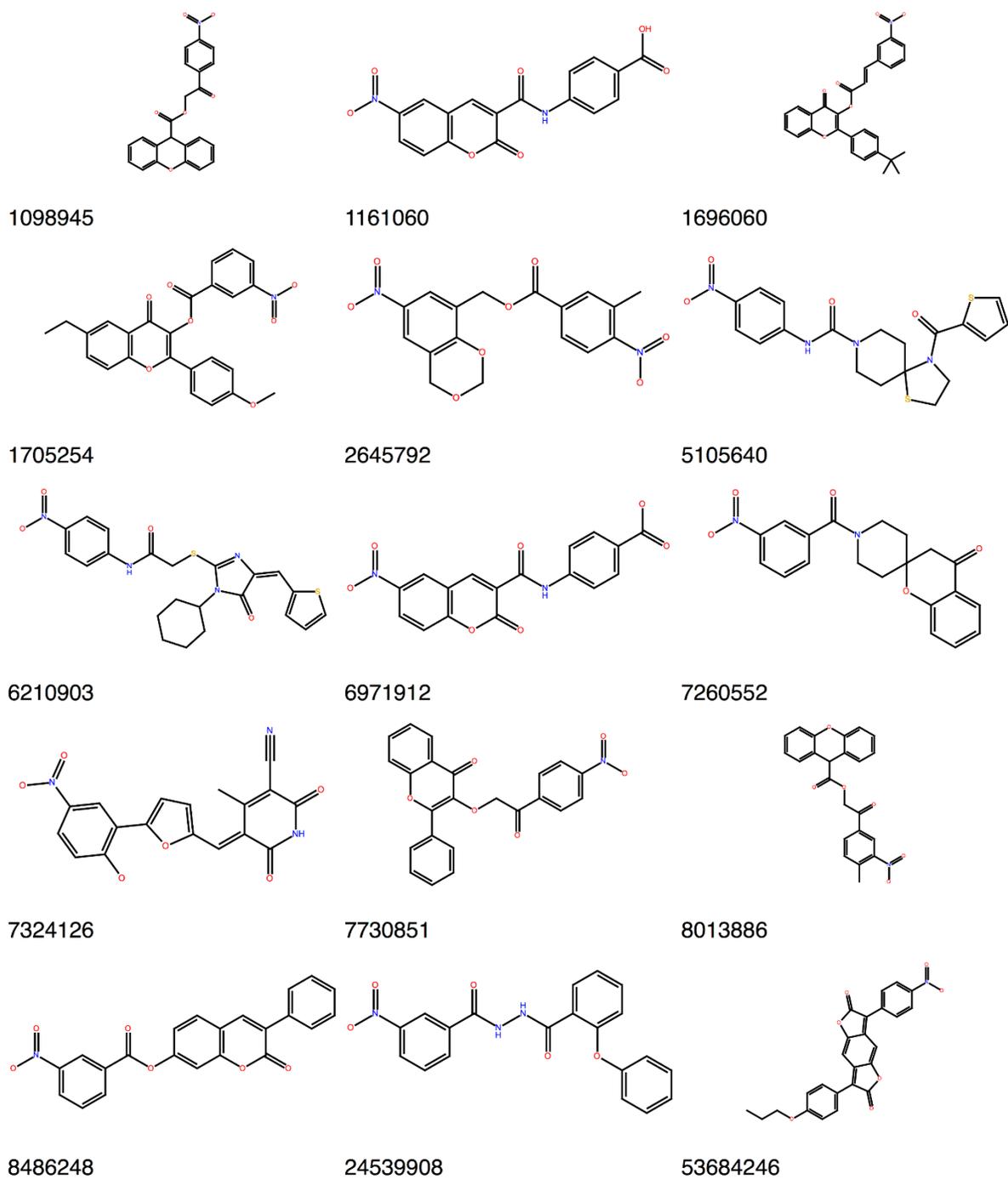


Figure 4.6. Chemical structures of the top fifteen hits derived from the turbo similarity search. The compounds have been ranked using the MM-PBSA binding energies. The PubChem Compound ID is reported below each structure.

Table 4.1 MM/PBSA binding energies of the top fifteen hits.

<b>Rank #</b>	<b>PubChem ID</b>	<b>MM/PBSA binding energy (kcal/mol)</b>
1	6210903	-33.84
2	7324126	-33.41
3	8486248	-31.49
4	7730851	-31.25
5	1696060	-30.29
6	1161060	-30.11
7	1098945	-30.11
8	53684246	-29.75
9	7260552	-29.48
10	6971912	-29.47
11	5105640	-29.33
12	2645792	-28.56
13	8013886	-28.23
14	1705254	-28.16
15	24539908	-28.06

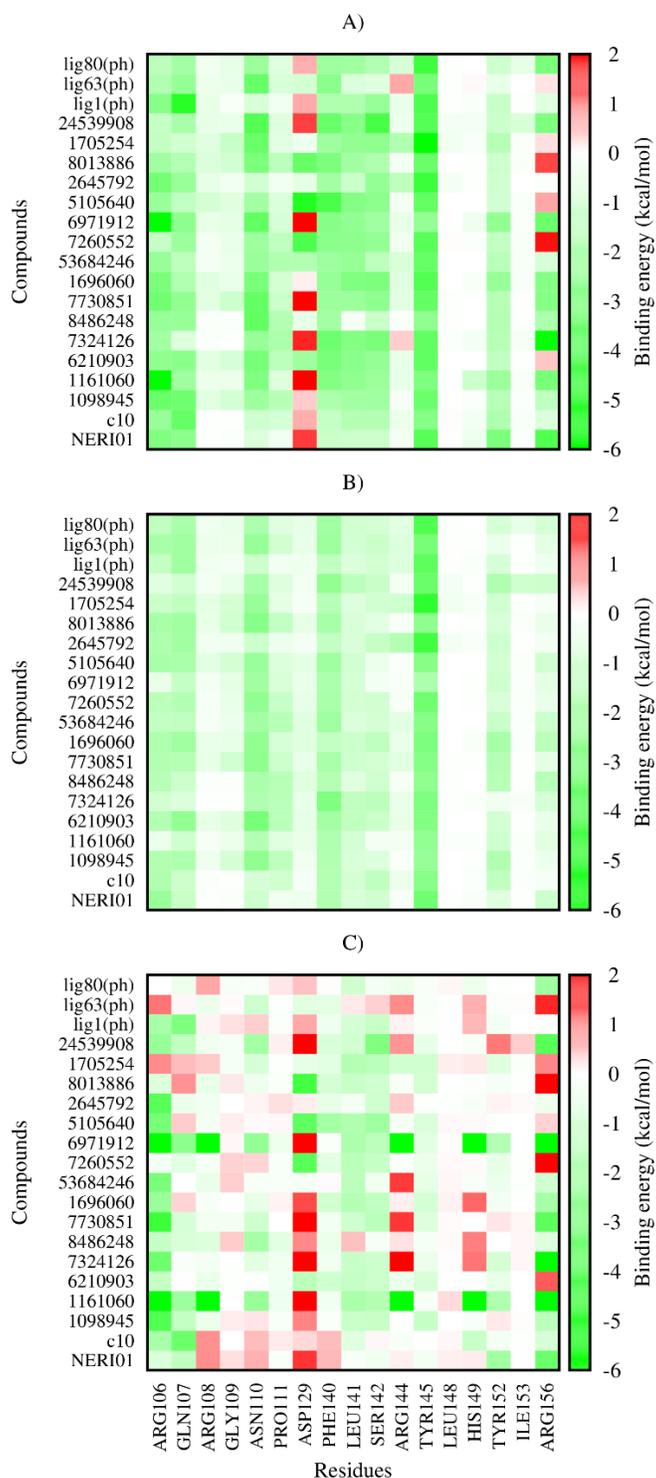


Figure 4.7. Heat maps of the A) total B) van der Waals and C) electrostatic contributions of the ERCC1 residues to the binding energies. NERI01 and c10 are the two lead structures. Compound names followed by (ph) are derived from the pharmacophore-based virtual screening. The other

ligands are indicated with the PubChem Compound ID. A clear binding pattern can be observed for the total energy, with Tyr145, Arg106, Gln107 and Phe140 being the highest and most conserved contributors. The pattern is very similar for the van der Waals contributions, indicating the dominating role of hydrophobic interactions due to the stacking of ligand rings. Consistently with the instabilities observed during the molecular dynamics simulations, it is not possible to observe a well-defined pattern for the electrostatic interactions, although Arg106 is the main partner for the major part of the compounds. Asp129 contribution is often unfavorable, due to the electrostatic interference with the negatively charged atoms of the ligands.

For example, the hit coded as 6971912 in PubChem scored -29.47 kcal/mol, and also established hydrogen bonds with Arg106, Arg144, Ser142, Phe140, Gln107 and water molecules. Noticeably, the net negative charge on the carboxylic group strongly interacted with the positive charge of Arg106, making the charge-assisted interaction completely stable during the simulation. The hydrophobic features of the binding are similar to the one present in the two lead compounds (Figure 4.5B). The RMSD of the heavy atoms of the compound during the production simulation shows fluctuations around an average value of 0.5 Å relative to the first snapshot after the release of restraints (Figure 4.8). The docked pose of the compound seems to be more stable than NERI01 (about 1 Å of displacement from the beginning of the simulation), a fact that can be explained by the larger number of stable electrostatic interactions present in the former (evident also from the decomposed electrostatic contributions in Figure 4.7C).

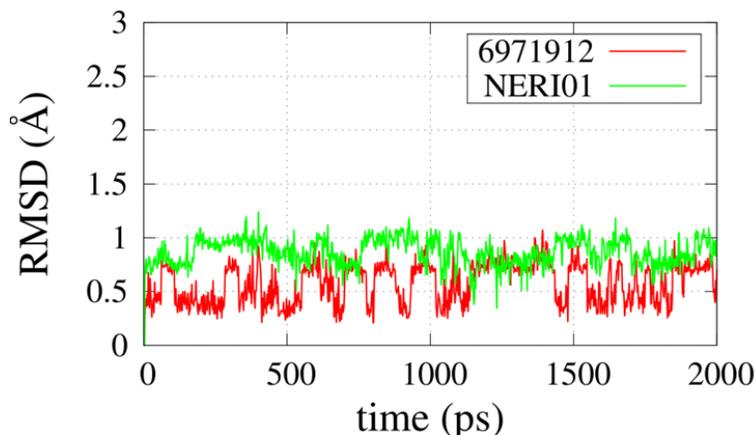


Figure 4.8. RMSD trend of NERI01 and 6971912 during the 2 ns of production simulation. Only the heavy atoms of the compounds have been considered. 6971912 shows a smaller displacement from the original docking pose during the simulation, but also higher, still stable fluctuations.

Regarding the thirteen hits obtained from the pharmacophore-based virtual screening followed by docking simulations, three structures showed a MM/PBSA score better than that of the two lead compounds (hereinafter denoted as ligand 1, ligand 63 and ligand 80) and can therefore be considered potential candidates for new scaffolds. The binding poses are reported in Figure 4.9. As expected, these compounds establish strong interactions with the key residues that have been reported to interact with the lead compounds. These three structures possess bulky, hydrophobic cores with a number of polar branches that allowed them to establish hydrogen bonds with the residues of the binding pocket. The first compound (ligand 1, 106408 from the Mechanistic set) showed a binding energy of -25.41 kcal/mol, the second one (ligand 63, 5113 from the Natural Products set) was the best ranked with -28.24 kcal/mol whereas the last (ligand 80, 107582 from the Diversity set) scored -25.25 kcal/mol. Noticeably, ligand 63 was already the best ranked hit among the pharmacophore driven hits in the docking simulations. The residue contributions to the binding energies for these molecules are also reported in Figure 4.7, showing consistency with the patterns discussed before.

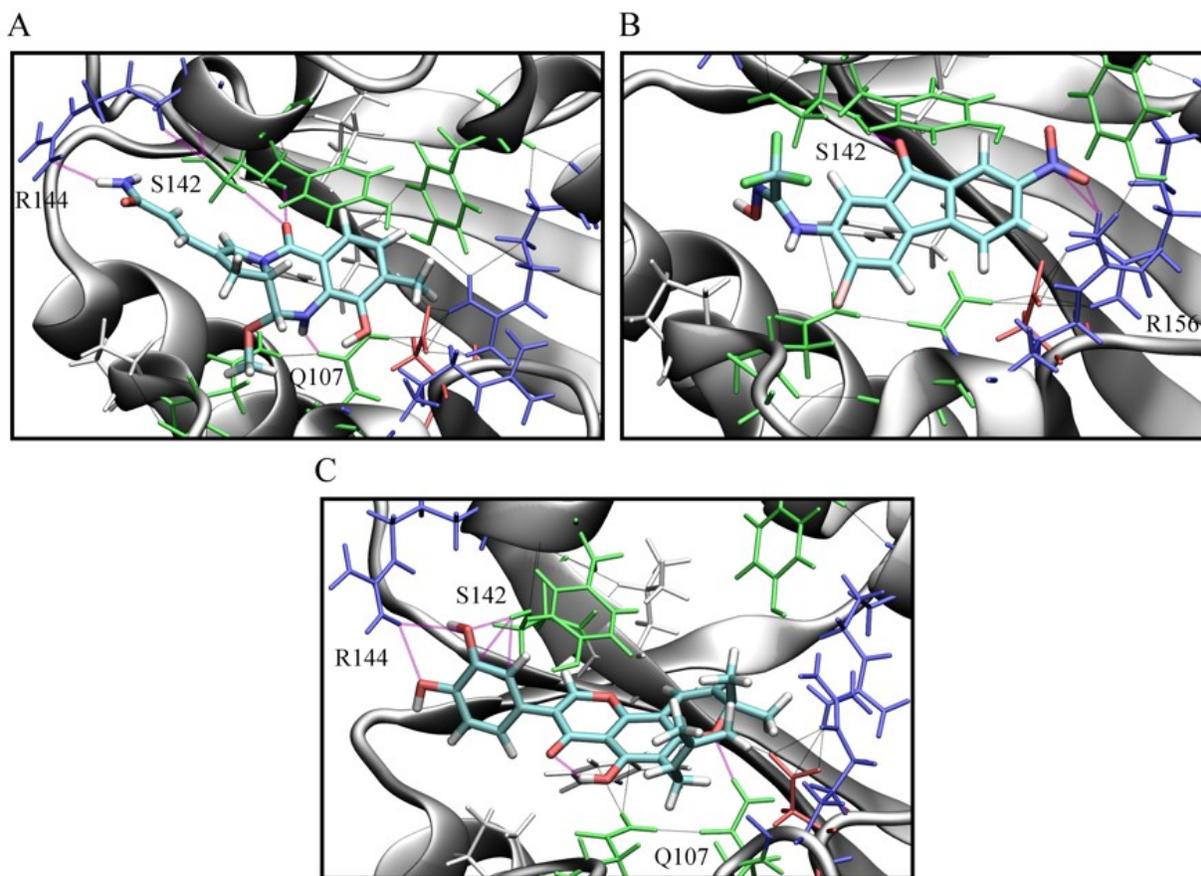


Figure 4.9. Binding poses and hydrogen bond networks of the new scaffolds from the NCI sets, A) ligand 1, B) ligand 80 and C) ligand 63. These three compounds are the top hits identified through the pharmacophore-based screening and represent promising scaffolds to be optimized. The hydrophobic backbones are well-stacked within the hydrophobic cleft of the ERCC1 central domain. The polar extremities can form hydrogen bonds with the Arg residues, although these interactions were generally unstable during the simulations.

#### 4.4 Conclusion

The inhibition of the NER repair pathway has been demonstrated as a promising approach to overcome drug resistance to platinum-based therapies in cancer cells. One of the essential steps of this machinery is the interaction between the XPA and the ERCC1 proteins, which represents an attractive target for pharmacological intervention. This is mainly due to the specificity of this

interaction to the NER pathway and its correlation with resistance to platinum-based treatment. Our group has previously identified two lead structures that target this interaction. Here, we continue our efforts to discover similar but more potent molecules, as well as new potential scaffolds. We employed an extensive and modified similarity search approach to screen the PubChem Compounds database, one of the largest available databases, for novel hits. The outcomes of this ligand-based virtual screening have been refined using structure-based techniques, resulting in a set of promising compounds to be validated experimentally. In addition, we also constructed a pharmacophore model and used this model to screen four NCI databases containing diverse sets of molecules to identify new scaffolds with the same chemical features as the reference lead structures. Again, target-based methods have been used to refine the results, resulting in three scaffolds that can potentially result in novel inhibitors of the XPA-ERCC1 interaction. In addition, we performed the decomposition of the binding energies among the residues constituting the binding pocket. A common pattern, dominated mainly by stable and conserved hydrophobic interactions rather than the electrostatic ones, was observed for the identified hits. The key residues for ligand binding have been therefore defined as Arg106, Gln107, Phe140 and Tyr145. The ligand motifs that seemed to be specific for the binding to the target are a backbone constituted by chemical rings and at least one negatively charged extremity.

Experimental validation of our findings will be the next goal, with two aims: first, to validate novel XPA-ERCC1 inhibitors; second, to demonstrate the effectiveness of our computational workflows in identifying active molecules.

In conclusion, we have employed two hybrid ligand/target-based protocols that can be adopted in virtual screening to identify new active compounds starting from the structures of already known

active ones; these two strategies can be indeed used to screen very large databases of small molecules in a reasonable amount of time. Our future directions will include the optimization and the use of these protocols to discover small molecules modulating other protein-protein interactions involved in DNA repair mechanisms.

## 4.5 Materials and Method

### 4.5.1 Similarity Search

A similarity score is usually calculated between the chemical fingerprints of known active molecules and the compounds in a database. The objective is to identify new active structures [172]. Among the different similarity scores, here we used the Tanimoto index due to its rapidity and reliable performance [71]. The score can be computed between two binary chemical fingerprints of the same length as Equation 2.7. In this study, we selected the 2D PubChem Compounds database [65] as the library from which to identify molecules similar to our previously identified hits. The PubChem database is suitable for virtual screening [66] and was chosen for two main reasons: first, the database contains more than 68 million structures; second, for each structure in the repository a PubChem substructure fingerprint is already available. These fingerprints are composed of 881 binary bits describing the element counting, ring presence, atom pairs and neighbors, and SMART patterns [173]. We employed the turbo similarity search technique in this study [71,174,175], using the following protocol. The *sdf2fps* tool from chemfp1.1 package [72] was used to extract all the fingerprints from the PubChem repository. Following that we used the *simsearch* routine to select the compounds that showed a Tanimoto score equal or higher than 0.80. The top 100 hits for each reference structure have been then used as references for similarity search of the PubChem compounds, using a Tanimoto

threshold of 0.85. The overlap between the four obtained libraries was finally eliminated, resulting in a non-redundant collection of molecular structures.

#### 4.5.2 Ligand Preparation

In order to focus our study on drug-like molecules, the ligand structures obtained from the turbo similarity search step have been filtered using an upper limit of 6.5 for the ADMET Risk score (combination of physicochemical, metabolic and toxicity predictions) computed with ADMET Predictor 7.1, an artificial intelligence software able to predict the ADMET properties (Adsorption, Distribution, Metabolism, Elimination and Toxicity) of small molecules based on their chemical structures [176]. The compounds that successfully passed the filtering step have been imported into LigPrep where relevant 3D conformations, tautomers and ionization states at pH 7 have been obtained [177]. The AutodockTools *prepare\_ligand4.py* [56] script was finally used to prepare the output structures for the docking simulations in Autodock.

#### 4.5.3 Pharmacophore-Based Virtual Screening

The concept of pharmacophore was first introduced by Ehrlich in 1909 as a molecular framework describing the essential physicochemical features for a molecule to successfully interact with a specific biological target [178]. The design of a 3D pharmacophore is mainly used to identify new scaffolds (scaffold hopping) from pre-computed, multi-conformational compound libraries [179]. In this technique novel inhibitors/activators can be discovered by changing the central core of the active compounds while still conserving their essential binding features [180]. To identify new inhibitory scaffolds for the XPA-ERCC1 interaction, we developed a pharmacophore model using MOE software [181]. The predicted binding poses of the two known active compounds obtained from the previous studies have been used to define

the model using the Pharmacophore Consensus Query tool [182], with 2 Å of spatial tolerance for the superposition of features of the two molecules to be considered as pharmacophore annotations. The model was then manually refined by computing the electrostatic features maps in MOE within 4.5 Å of the docked ligands for the binding site of ERCC1. These maps are indicative of favourable zones for acceptor, donor and hydrophobic atoms as they are iso-surfaces of the potential  $qu + v$ , where  $q$  is the partial charge of the probe atom depending on the type of map,  $v$  is the van der Waals potential of the same atom and  $u$  is the electrostatic potential calculated by solving numerically the Poisson-Boltzmann equation [183]. Four libraries from the National Cancer Institute (NCI) Compound Sets repository [68], namely the Approved Oncology Drugs Set VI, the Diversity Set V, the Mechanistic Set III and the Natural Product Set III, have been selected as targets for the pharmacophore-based virtual screening, as they are generally designed to guarantee structural diversity of the compounds and their proper pharmacological features. A multi-conformational database has been created using Conformation Import by generating at maximum 10 tautomers and 500 conformations for each structure; two conformations were considered different if the RMSD value was equal or higher than 0.6 Å. The Oprea's filter has been applied to the results to remove non-lead-like compounds [184]. Finally, the Pharmacophore Search routine was run using the designed pharmacophore as the query and the multi-conformational filtered database as the target. The AutodockTools *prepare\_ligand4.py* script was used to prepare the output structures for the docking simulations.

#### 4.5.4 Target Preparation

NMR experiments can accurately determine the 3D structures of sufficiently small proteins and also extensively sample their conformational space [185,186]. Structures obtained from an NMR ensemble are therefore reflecting the dynamical nature of the protein structure and are suitable

for docking experiments based on a RCS approach. The goal is to introduce the full flexibility of both the backbone and the side chains of the target protein during docking [51]. In this context, here, we used the available NMR structure of the ERCC1 central domain in complex with the XPA binding peptide (PDB entry: 2JNW) [164]. Prior to docking, we removed the XPA peptide from each NMR snapshot and the protonation states for the ERCC1 structures have been assigned using the H++ online server [187], using pH 7, a salinity of 0.15 M, an internal dielectric constant of 10 (for protein) and an external dielectric constant of 80 (for water) as input parameters. Each structure has been then solvated with an octahedral box of TIP3P explicit water molecules (buffer of 12 Å). The systems have been neutralized at a physiological concentration of 0.15M of explicit Na - Cl ions using the AmberTools13 tleap program [188] with the parameters from the ff14SB force field [189]. The systems have been minimized in Amber pmemd.cuda [190,191] using the following protocol. Initially the positions of water and ions have been relaxed with 1,000 steps of the conjugate gradients method, keeping restrained the whole protein; then a strong harmonic restraint was applied just to the protein backbone atoms and other 1,000 steps were run to relax the side chain conformations. The pdbqt files for the structures have been prepared for Autodock simulations by computing Gasteiger charges, Autodock atom types and by merging nonpolar hydrogen charges on the respective carbons using the AutodockTools *prepare\_receptor4.py* script [56].

#### *4.5.5 Molecular Docking Simulations*

Docking simulations have been prepared using AutodockTools and run on Autodock 4.2. For each target, a 70 X 70 X 70 points grid has been centered on the XPA binding site within ERCC1 to include all key residues mediating this interaction, using 0.375 Å as the spacing between the grid points. Affinity, electrostatic and desolvation maps have been calculated using Autogrid. The

Lamarckian Genetic Algorithm [55] has been used for the docking protocol. We used the docking parameters as described in our earlier studies [45,56]. This included 10,000,000 maximal energetic evaluations, 27,000 maximal generations, 300 individuals and one survival for each generation with the rates of mutation and crossover set to 0.02 and 0.8. Each docking simulation employed 50 trials and the solutions have been clustered based on the RMSD metric using 2 Å as tolerance. Only the simulations showing at least a cluster with 12 poses (25%) or more have been retained for further analysis; Autodock scoring function has been used for the estimation of the binding affinities [47].

#### *4.5.6 Molecular Dynamics Simulations*

The force field parameters for selected ligands have been obtained using the GAFF force field [192] and the Antechamber utility [193]. The AM1-BCC charges were assigned for the ligands [194] and the ff14SB parameters were assigned for the protein. The complexes have been neutralized, solvated in a Na-Cl concentration of 0.15 M with tleap using the same process as described above. The simulations have been performed in pmemd.cuda: an initial minimization step was performed in order to relax the water and ionic positions. The whole system was then minimized and heated gradually up to 300 K in 100 ps using Langevin dynamics. During the heating process we restrained the backbone of the protein and the heavy atoms of the ligand; a time step of 0.5 fs and constant volume conditions have been employed during this phase. The time step has been set to 2 fs, constant pressure conditions (1 atm) have been imposed and the restraints have been gradually released in four phases of 50 ps each; finally 2 ns of production simulation have been performed; the coordinates of the system were recorded every 2 ps during this last phase. The choice of the length for production simulations was a critical point, as we needed to balance the time required and the accuracy of the calculations. In the past, we have

employed the 2 ns scale for the refinement of docking-derived poses with excellent results [80,168]. A recent study evaluated the impact of different times of simulations to the MM/PBSA performance. A deterioration of the correlation coefficients between experimental and computational data was observed among the six investigated systems for longer simulations (3-5 ns) [171]. A similar conclusion was reached by Xu et al., where MM/PBSA showed consistent ranking performances among different lengths of simulation (1-4 ns) [195]. Really recently, Su et al.'s work aimed to testing the effect of different radii sets on different lengths of simulation to MM/PBSA results: just 0.25-2 ns trajectories led to satisfactory determination coefficients ( $> 0.7$ ), whereas trajectories longer than 2 ns seemed to cause a decrease in accuracy. By successively investigating this behavior, the authors observed a rapid convergence of the binding energy values, concluding that the predictive power was likely to decrease due to the amplification of force field errors [196]. The 2 ns length of the simulations was therefore chosen based on our previous experience and the literature discussed above.

#### 4.5.7 MM/PBSA

The enthalpic contribution to the change in free energy due to the binding of the ligands has been estimated using the MM/PBSA method using the script MMPBSA.py [197]; the change of the energy of a complex configuration is calculated as

$$\Delta G_{bind,solv} = \Delta G_{MM,vac} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,protein}) - T\Delta S \quad (4.1)$$

where  $\Delta G_{MM,vac}$  is calculated as the sum of electrostatic and van der Waals interactions occurring between the protein and the ligand.  $T\Delta S$  is the term modeling the change in conformational entropy due to the binding. The solvation terms are modeled as

$$\Delta G_{solv} = \Delta G_{solv,polar} + \Delta G_{solv,npolar} \quad (4.2)$$

where the polar contribution of the solvent is calculated using the numerical solution of the Poisson-Boltzmann equation and the Amber *pbsa* solver with a grid spacing of 0.5 Å; the optimized radii from Luo and Tan were used [198] because this set of parameters performs better in ranking small molecules when compared with other PB and GB methods [171,195]. The ionic concentration was set to 0.15M. The hydrophobic contribution to the solvation free energy has been calculated as

$$\Delta G_{solv,npolar} = \gamma \cdot SASA \quad (4.3)$$

where  $\gamma$  is the surface tension equal to 0.0072 kcal/mol/Å<sup>2</sup> and *SASA* is the solvent accessible surface area. The change in free energy has been calculated for 200 snapshots (extracted every 10 ps) of each production simulation and the final value has been obtained as the average. A pairwise decomposition of the binding energy was performed for seventeen residues located in the binding zone, in order to quantify the role of the key interactions with the ligands. In particular, van der Waals and electrostatic solute-solute interactions have been separately investigated to determine the predominant nature of the binding. In addition, the changes of conformational entropy ( $T\Delta S$  term) were calculated for a diverse set of the top hits using normal mode analysis. This term is critical to calculate, highly fluctuating and computationally demanding. In fact, focused studies observed no improving or even worsening of MM/PBSA results when the entropies were added [171,199,200]. Although it is necessary for estimating the absolute affinities, it can be neglected for ranking purposes if the ligands are sufficiently similar.

#### *4.5.8 Visual Analysis*

The visual inspections of the ERCC1 structures, docking poses and molecular dynamics conformations and trajectories have been performed using Visual Molecular Dynamics (VMD) [201] and MOE software. The RMSD values have been computed using AmberTools ptraj, considering the first snapshot after the release of the harmonic restraints as the reference structure.

## **Chapter 5 - Computational Characterization of Small Molecules Binding to the Human XPF Active Site and Virtual Screening to Identify Potential New DNA Repair Inhibitors Targeting the ERCC1-XPF Endonuclease<sup>4</sup>**

### **5.1 Abstract**

The DNA excision repair protein ERCC-1-DNA repair endonuclease XPF (ERCC1-XPF) is a heterodimeric endonuclease essential for the nucleotide excision repair (NER) DNA repair pathway. Although its activity is required to maintain genome integrity in healthy cells, ERCC1-XPF can counteract the effect of DNA-damaging therapies such as platinum-based chemotherapy in cancer cells. A promising approach to enhance the effect of these therapies is, therefore, to combine their use with small molecules able to inhibit the repair mechanisms in cancer cells. Currently, there are no structures available for the catalytic site of the human ERCC1-XPF which performs the metal-mediated cleavage of a DNA damaged strand at 5'. We adopted a homology modeling strategy to build a structural model of the human XPF nuclease domain, containing the active site, and to extract dominant conformations of the domain using molecular dynamics simulations followed by clustering of the trajectory. We investigated the binding modes of known small molecule inhibitors targeting the active site to build a pharmacophore model, and we then performed virtual screening of the ZINC Is Not Commercial 15 (ZINC15) database to identify new ERCC1-XPF endonuclease inhibitors. Our work provides structural insights regarding the binding mode of small molecules targeting the ERCC1-XPF

---

<sup>4</sup> This chapter has been published as research article as Gentile, F.; Barakat, K.; Tuszynski, J. Computational Characterization of Small Molecules Binding to the Human XPF Active Site and Virtual Screening to Identify Potential New DNA Repair Inhibitors Targeting the ERCC1-XPF Endonuclease. *Int. J. Mol. Sci.* 2018, *19*, 1328.

active site that can be used to rationally optimize such compounds. We also propose a set of new potential DNA repair inhibitors to be considered for combination cancer therapy strategies.

## **5.2 Introduction**

The human genome is continuously exposed to damage caused by endogenous and exogenous agents. The effects resulting from these lesions range from interfering with cellular processes to inducing mutations which can lead to several pathological conditions. In order to maintain genome integrity, cells have developed a series of DNA repair pathways, which are able to recognize and repair specific DNA damages through the action of dedicated proteins [90]. In the context of cancer, DNA repair pathways can be seen as both friends and enemies at the same time. Indeed, although the obvious role of these pathways is to maintain genome stability and remove mutation-causing damages, they can interfere with cancer therapies, which aim to damage the cancer cell genome and hence induce apoptosis. Examples of such therapies are platinum-based chemotherapy and ionizing radiation therapy. Unsurprisingly, the success of these strategies highly depends on the DNA repair capability of the targeted cell population [128,129,202]. Hence, a relatively new direction to improve the efficacy of these treatments is to use them in combination with drugs able to inhibit the DNA repair mechanisms [99,100].

Among the five main repair pathways found in humans, the nucleotide excision repair (NER) pathway is dedicated to the repair of bulky DNA lesions, which distort the helix structure, thus interfering with the replication cycle. Such lesions can be caused by ultraviolet light radiation (UV), environmental chemical agents or reactive oxygen species [90,160]. NER is also responsible for the removal of DNA damages caused by platinum-based chemotherapy drugs such as cisplatin [99]. Over-expression of NER proteins results in cisplatin resistance in cancer

cells, whereas cell populations with low-expression of NER proteins are hypersensitive to DNA damaging agents [128,129]. In addition, modulation of NER results in sensitizing cancer cells to DNA-damaging chemotherapy [83].

The NER pathway involves about thirty proteins whose role is to recognize, remove and replace a damaged DNA strand. One of the essential players of NER action is the DNA excision repair protein ERCC1-DNA repair endonuclease XPF (ERCC1-XPF) complex, a 5'-3' structure-specific endonuclease, which cuts the strand at 5' of the damaged zone. ERCC1-XPF is also involved in inter-strand crosslink (ICL) and double-strand break (DSB) repair pathways. ERCC1-XPF is an heterodimer composed by two proteins: ERCC1 that contains 297 residues divided in a central domain and a double helix-hairpin-helix (HhH2) domain; and XPF, which is constituted by 916 residues divided in an helicase-like domain, which lacks the helicase activity, a nuclease domain, which contains the catalytic site, and finally a HhH2 domain as well. Dimerization occurs mainly through the two HhH2 domains. An excellent review article about the structure and function of ERCC1-XPF can be found in McNeil *et al.* [203].

ERCC1-XPF is an attractive target for designing small molecule inhibitors of DNA repair. In order to inhibit the activity of the ERCC1-XPF endonuclease, three major ways have been explored in the last few years. The first approach is to target the interaction between the central domain of ERCC1 and the DNA repair protein complementing XP-A cells (XPA), through which the endonuclease is recruited to the damage site in NER [80,82]. This approach would, however, be effective just in inhibiting NER, while the ICL and DSB repair activity would be preserved as XPA is not involved in these pathways [203]. A second approach is to target the ERCC1-XPF protein-protein interaction. Our group and others identified and targeted binding pockets at the interface of the dimerized HhH2 domains to inhibit the dimerization of the ERCC1

and XPF, essential to build a functional endonuclease [84,85]. Although this approach would result in stopping any activity of ERCC1-XPF, it presents some difficulties due to the high-affinity, hydrophobic nature of the involved protein-protein interaction [203]. Finally, the third approach is to target the XPF active site. Recently, McNeil *et al.* [85], Chapman *et al.* [204,205] and Arora *et al.* [206] discovered several small molecule inhibitors targeting the catalytic site of XPF with promising biological activities. The lack of an experimentally determined crystal structure for the XPF nuclease domain, as well as the similarity of the active site with related nucleases are the two main drawbacks of this latter approach. Nevertheless, targeting the XPF active site is a promising strategy to inhibit the endonuclease activity due to the presence of metal ions in the catalytic site (ideal for metal chelators), the weak contacts established by the domain and the DNA, and the number of successful drug discovery programs targeting DNA repair-related similar enzymes [203].

In this work, we employed computational methods to: 1) build a structural model of the human XPF nuclease domain, which can be used in structure-based drug design and virtual screening (VS); 2) investigate the binding modes of known XPF active site inhibitors, identifying key residues involved in small molecule binding; 3) perform a pharmacophore and structure-based VS campaign against the ZINC Is Not Commercial 15 (ZINC15) compound database [67] in order to propose potential novel inhibitors binding to the XPF catalytic site. The reported results provide for the first time a detailed investigation of the interactions between the XPF active site and small molecules binding to it. Our findings should be of considerable interest to rationally modify these molecules in order to improve the binding affinities as well as their specificity to the target. Additionally, we provided a set of commercially-available compounds which can potentially bind to the XPF catalytic site and inhibit the endonuclease activity of ERCC1-XPF,

and that can be therefore considered in combination with DNA-damaging cancer therapies to amplify their effects.

## 5.3 Results and Discussion

### 5.3.1 Homology Modeling and Molecular Dynamics Simulation

The results obtained from the Molecular Operating Environment (MOE) *MOE-Search PDB* protocol are reported in Table 5.1. We identified top templates based on an expectation value (E-value) equal or lower than  $1e-12$ . Other potential templates with E-values between the accepted value and the cutoff were retained only if the associated Z-scores were at least 6.

Table 5.1. Results from different substitution matrices available in MOE 2013 for the detection of homologs of the human XPF nuclease domain. Proteins are reported with their PDB ID. Top templates were identified as sequences with a E-value lower than the E-value acceptance ( $1e-12$ ). Others templates showed a E-value between E-Value acceptance and E-value cutoff (10), and a Z-score of at least 6.

Substitution matrix	Top templates	Others
BLOSUM50	2BGW, 2BHN	1J22, 2ZIU, 4P0P
BLOSUM62	2BGW, 2BHN, 1J22	2ZIU, 2ZIX, 4P0P
Gonnet	2BGW, 2BHN	1J22
PAM250	2BGW, 2BHN	1J22

All the four tested matrixes identified XPF-related proteins from the *Aeropyrum pernix* (PDB ID 2BGW, 2BHN) [149] and *Pyrococcus furiosus* (PDB ID 1J22) [207] archaea. In addition to the hits identified using the Gonnet and Point Accepted Mutation 250 (PAM250) methods, the BLOcks SUBstitution Matrix (BLOSUM) matrices led to the identification of the Mus81 protein

(PDB ID 2ZIU (human/*Dario rerio*), 2ZIX (human) and 4P0P (human)) as template as well, known to be related to XPF [208]. In contrast to the other three matrices, BLOSUM62 included the Hef protein (1J22) from *Pyrococcus furiosus* within the top templates. Hence, we selected the BLOSUM62 results for the successive steps, this matrix having showed the best performances in detecting biological relationships even for distantly-related proteins [209–211]. The nuclease motif is conserved among XPF family, putative RNA helicases (SF2) and Mus81 family and it is represented in human XPF by residues D687, E690, D715, E725, R726, K727 and D731 [212]. In addition to this motif, we observed seven other conserved residues from the multiple sequence alignment, corresponding to V686, L711, G714, S733, G739, Q744 and E760 in the human XPF sequence. The sequence alignments of the XPF nuclease domain and the six templates are reported in Figure A.1 in Appendix A.

The top templates identified by MOE were 2BGW, 2BHN and 1J22. The metal-binding site of the XPF is likely to employ a two-metal-ion catalysis process to cleave the DNA [213]. However, the available structures contain from zero to one metal ion. The absence of a second ion may be due to the requirement of a catalytic complex for its stable binding, as in the case of the related Mus81-Eme complex [214]. Also, the vast majority of known XPF active site inhibitors contain at least one metal-binding motif. For these reasons, we included also the HeF protein from *Pyrococcus furiosus* associated with the PDB ID 1J25, which is the same structure as 1J22 but containing one coordinated metal ion. The four nuclease domains (from 2BGW, 2BHN, 1J22 and 1J25) shared a very similar and superimposable structure (Figure 5.1). Finally, we selected the nuclease domain of the 1J25 structure as template to build the homology model of the human XPF nuclease domain, based on the highest sequence identity (35.2%) and similarity (60.7%) scores observed among the four sequences and the presence of one metal ion.

The alignment of the sequences of the human XPF nuclease domain, 2BGW and 1J25 is reported in Figure A.2 in Appendix A.

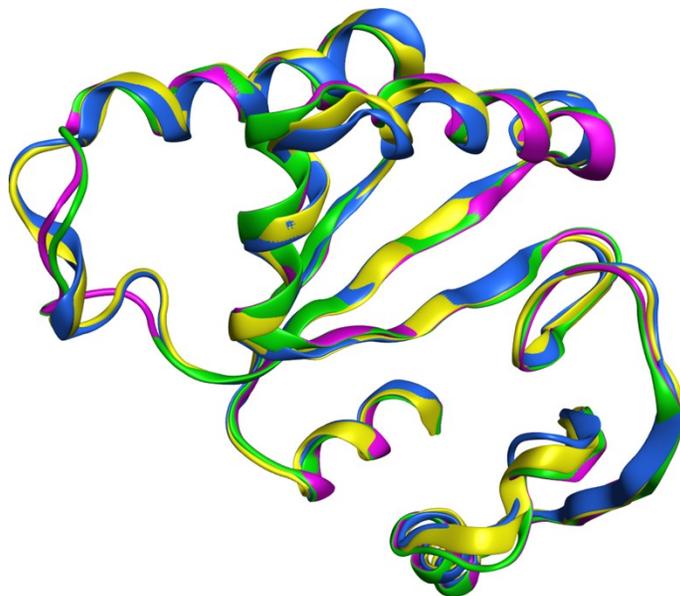


Figure 5.1. Structural superposition of the four top templates for the human XPF nuclease, 2BGW (blue, XPF from *Aeropyrum pernix*), 2BHN (yellow, XPF from *Aeropyrum pernix*), 1J22 (purple, HeF from *Pyrococcus furiosus*) and 1J25 (green, HeF from *Pyrococcus furiosus*).

Once the homology model of the human XPF nuclease domain was obtained, we manually modified the  $Mn^{2+}$  ion deriving from the HeF structure to a  $Mg^{2+}$  ion, which is the biologically-relevant cofactor for the ERCC1-XPF endonuclease [203]. The metal ion was stably coordinated by the negatively-charged side chains of residues D715 and E725 and the backbone oxygen of R726.

The best predicted structure derived from the 1J25 template was simulated with molecular dynamics (MD) for 170 ns. The root-mean-square deviation (RMSD) trend of the backbone atoms of the modeled domain reached a plateau after about 60 ns, with stable fluctuations around 3 Å for the remaining time of the simulation. The backbone atoms of the active site residues

fluctuated steadily around 1 Å for all the simulated time, after the initial equilibration phase following the restrain release. The RMSD plots are reported in Figure 5.2. After visually inspecting the zone surrounding the metal ion during the simulation, and considering the previously sequence alignments, we defined the active site as the residues D687, R689, E690, D715, E725, R726, K727 and D731. During the simulation, three stable water molecules completed the coordination of the  $Mg^{2+}$  ion (coordination number of 6).

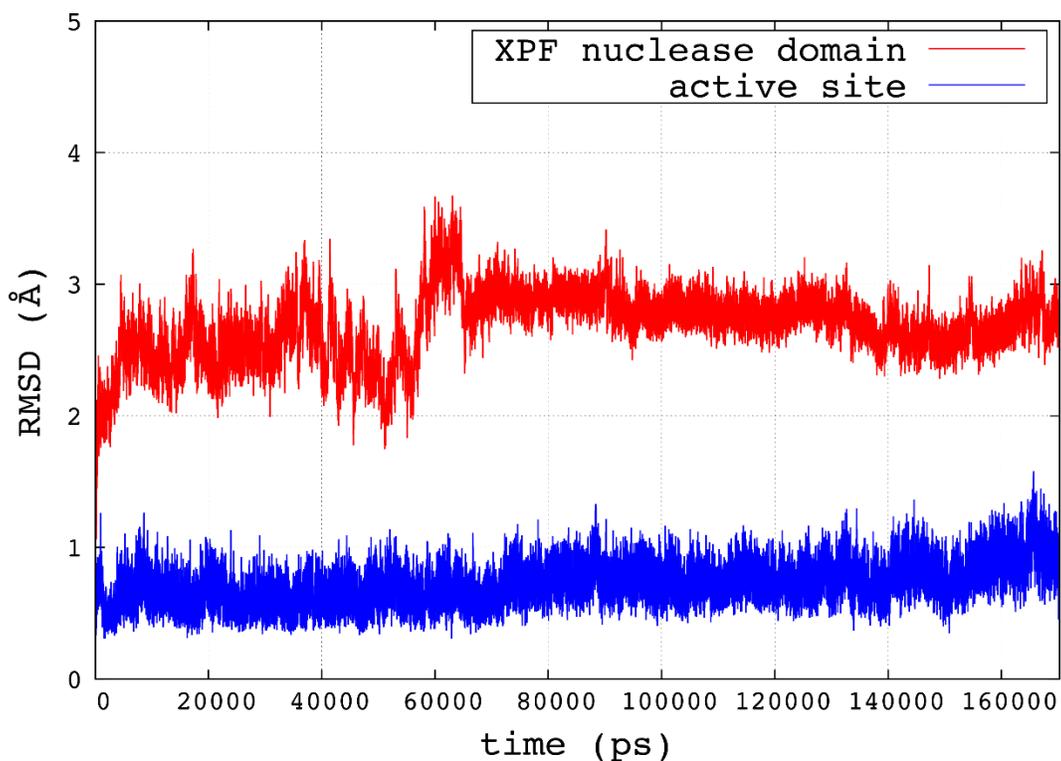


Figure 5.2. RMSD trends for the backbone atoms of the human XPF nuclease domain (red) and the active site (blue), defined as residues D687, R689, E690, D715, E725, R726, K727 and D731.

Clustering of active site conformations was performed over the last 106 ns of the MD simulation. At cluster counting equal to 10, we observed the highest peak of the pseudo-F statistic (pSF) value, a kink in the curve of the ratio between the sum of square regression and the number of

total squares (SSR/SST) and a local minimum for the David-Bouldin (DBI) index, indicating optimal cluster counting (see Figure A.3 in Appendix A). Cluster compositions are reported in Table A.2 in Appendix A. In order to exclude rarely occurring active site conformations from the molecular docking simulations, we selected the representative structures from the top six most populated clusters including 99% of the total conformations, to be used as targets. We also included the lowest potential energy structure ( $\sim$ -98.831 kcal/mol) of the XPF domain extracted by the equilibrated part of the MD simulation.

### 5.3.2 Modeling of Small Molecules Binding to the Human XPF Active Site

In order to account for the flexibility of both the side chains and the backbone of the active site, we considered the seven XPF structures described previously as single targets for our docking protocol. A detailed view of these conformations is reported in Figure 5.3.

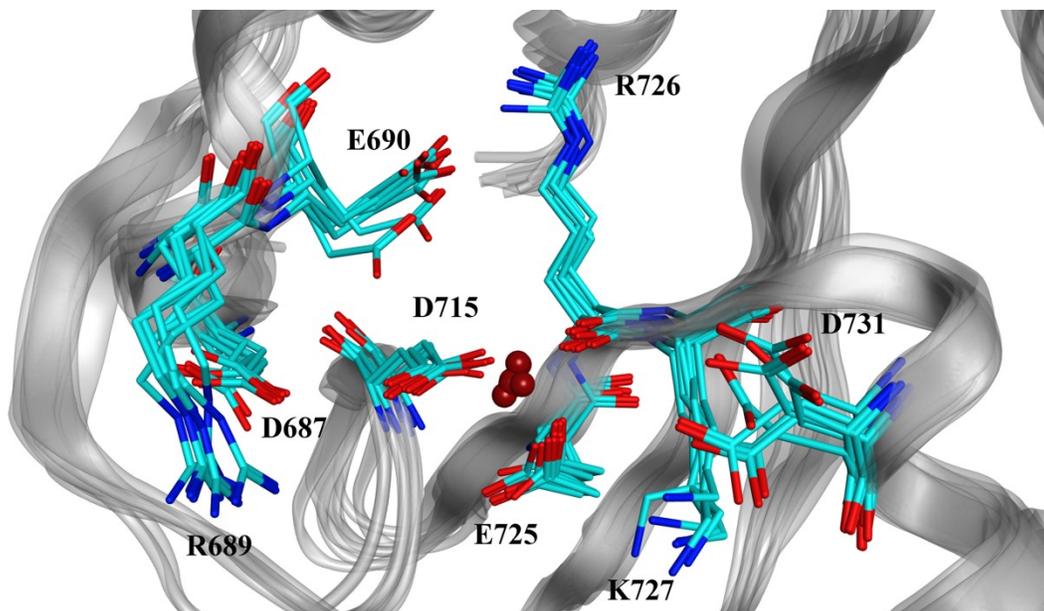


Figure 5.3. Superposition of the seven conformations of the human XPF active site obtained by clustering the MD trajectory and including the lowest potential energy conformation as well. The

set of conformations was then used as target for the molecular docking simulations. Red spheres indicate the positions of the  $Mg^{2+}$  ion present in the active site and coordinated by D715, E725 and R726.

From the resulting binding poses, we identified a pattern of conserved interactions between the small molecules and specific parts of the XPF active site. As expected, the metal-binding motifs present in the ligand structures carried a negative charge and were in close proximity of the  $Mg^{2+}$  ion. Also, we observed two hydrogen bonds being consistently established between ligand hydrogen bond donor and acceptor groups and E712 and K727, respectively. Therefore, the resultant three-point pharmacophore model included three features, namely one anionic (Ani) with radius of 2 Å, one donor projection (Don2) with radius of 3.2 Å and one acceptor projection (Acc2) with radius of 2.7 Å, as represented in Figure 5.4. It is noteworthy that this pharmacophore model accounted for multiple conformations of the active site, as we designed it considering ligands bound to different XPF structures.

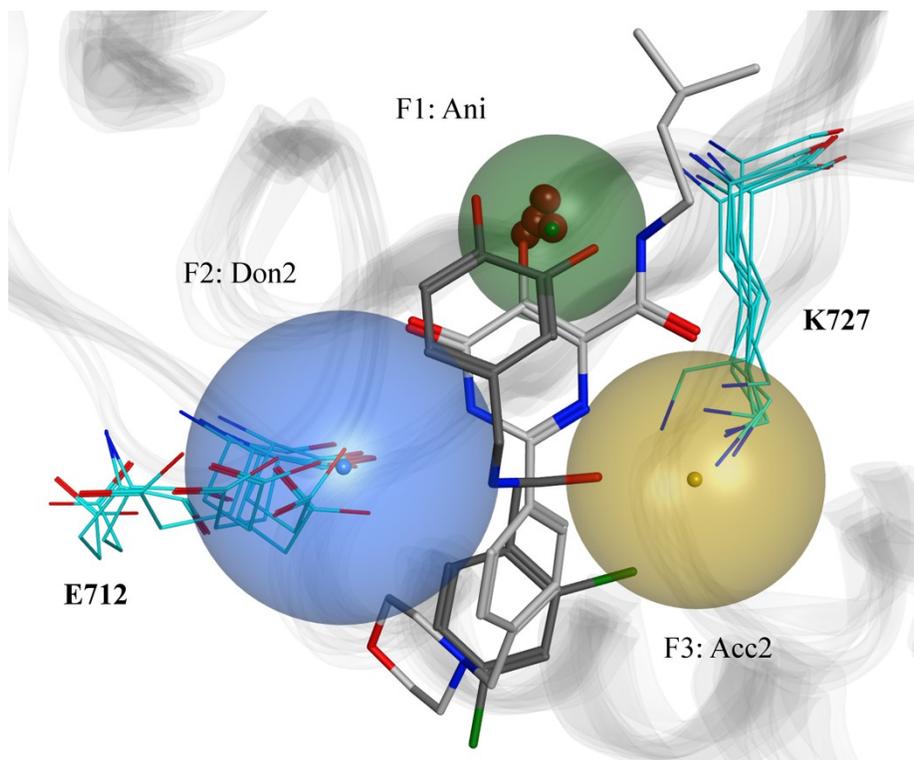


Figure 5.4. Pharmacophore model designed by considering the predicted binding poses of the known XPF inhibitors. Three features were present for all the ligand-protein complexes, namely the anionic one (Ani, in green) in proximity of the  $Mg^{2+}$  ion, the donor projection (Don2, in blue) close to E712, and the acceptor projection (Acc2, in gold) close to K727. Two docked ligands satisfying the pharmacophore model are also reported (13 [205] in dark grey, 33 [204] in light grey). All the conformations of the active site extracted from clustering the MD trajectory are represented.

### 5.3.3 Virtual Screening

Around 80,600,000 structures from ZINC15 were downloaded. After the filtering step and the pharmacophore-based screening, we reduced the number of compounds to undergo the structure-based VS step to 2,013,120. We then performed VS of the compounds against the set of structures of the XPF nuclease domain and retained just the resulting binding modes satisfying the pharmacophore features leading to only 104,714 unique compounds kept for consideration. The highest-ranked XPF inhibitor was compound 15, with a London dG score of -29.543 kcal/mol. 285 hits from the ZINC15 resultant set showed a better score than compound 15. Visual analysis of the binding modes was then performed, in order to further narrow down the hit set. Additional details about the resulting top fifty hits, including chemical structures, ZINC IDs and London dG binding energies, are reported in Figure A.4 in Appendix A.

Among the resulting binding modes of the top hits, we observed charge-assisted hydrogen bonds between charged groups of the compounds and the charged residues of the XPF active site as dominant interactions. The predicted binding modes of two VS-derived hits are reported in Figure 5.5, as are examples of such non-bonded interactions. Hit #5 (Figure 5.5A) showed hydrogen bonds between the guanidine group and the side chains of E712 and D715, in addition to a hydrogen bond between the same group and the backbone oxygen of L711. The  $Mg^{2+}$  ion

interacted with one of the oxygens of the compound, while the other one was involved in a hydrogen bond with K727. Hit #13 (Figure 5.5B), a smaller and less flexible compound, interacted with the  $Mg^{2+}$  ion and K727 via one of its carboxyl groups, and with E712 through a hydrogen bond with the imidazole ring.

Based on the binding poses of our predicted hits, we were able to identify in detail the electrostatic features of the active site, which are important for ligand binding. The active site of the human XPF is divided between two zones with different electrostatic properties: a negatively-charged part constituted by acidic residues such as E712, D715 and E725 (in red in Figure 5.5), favorable to establish interactions with electropositive moieties of the ligands; a positively-charged part constituted by the metal ion and K727, favorable for interactions with electronegative moieties of the ligands.

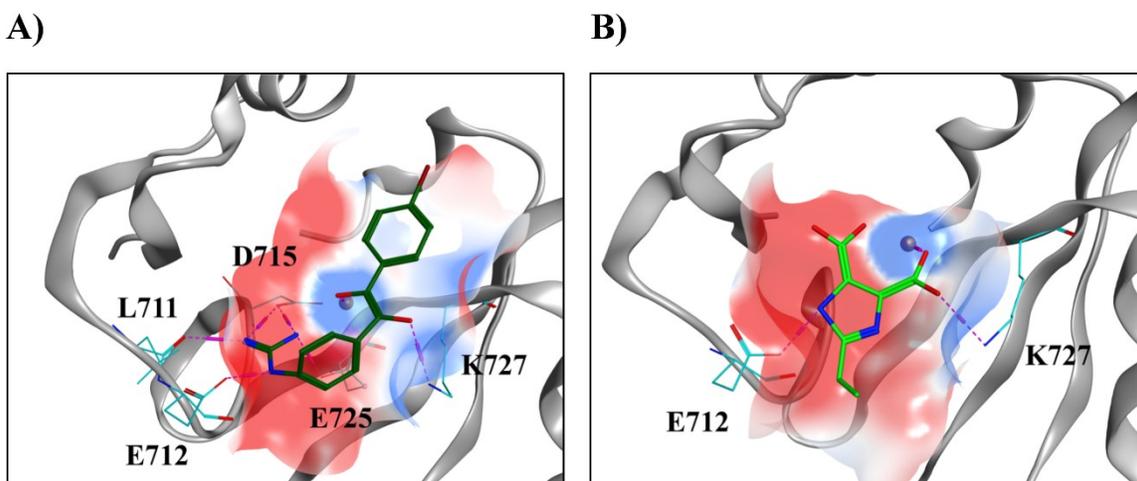


Figure 5.5. Binding modes of two promising hits derived from the VS. A) ZINC000049131978, hit #5 B) ZINC000038550857, hit #13. Hydrogen bonds are represented in purple dotted lines. The surface of the active site is colored depending on the electrostatic potential, ranging from -40 (intense red) to +40 (intense blue). See text for a detailed description of the electrostatic interactions.

## 5.4 Materials and Methods

### 5.4.1 Homology Modeling of the Human XPF Nuclease Domain

The amino acid sequence of the human XPF nuclease domain was defined as residue 658 to 813 from the entry Q92889 in the UniProt database [215]. For the entire homology modeling process, we used the MOE2013 package [181]. Initially, we used the *MOE-SearchPDB* module [216] to align the target sequence with a database of pre-clustered families of proteins [217] for which experimental structures are available in the Protein Data Bank (PDB) [32], identifying in this way potential template structures to be used for homology modeling. The parameters for the homology search have been chosen as follows: a gap start penalty of -12, a gap extend penalty of -2, an E-value cutoff of 10, an E-value acceptance of  $1e-12$ , 100 Z-iterations and a Z-score cutoff of 6. As a substitution matrix, we tested the BLOSUM62, BLOSUM50 [218], Gonnet [219] and PAM250 [220], all available in MOE 2013. *MOE-Align* [216] using sequence and structural alignment was used for the multiple alignment in the following way: first, the whole XPF sequence was aligned to the identified templates. Second, the XPF nuclease sequence was aligned to the first multiple alignment, in order to obtain a better alignment of the nuclease domains of the templates. Just the nuclease domain sequences were used for the successive steps, by trimming the sequences of the templates to the residues aligned within residues 658 and 813 of the human XPF nuclease domain. The best template obtained from this step was then used for the homology model building. The parameters were set as 10 intermediate models, one side chain model for each intermediate at 300 K, Medium refinement for intermediates, the Generalized Born/volume integral (GB/VI) [221] scoring for the selection of the final model; the final refinement was set to “Fine” with a root-mean square (RMS) gradient of 0.1 kcal/mol and the protonation states of the final model were assigned using *Protonate3D* [222]. Amber ff12SB

force field [223] was selected for the entire process. Coordinated metal ions present in the template were included in the process as environment for the induced fit.

#### 5.4.2 Molecular Dynamics Simulation and Clustering of the Trajectory

Amber ff14SB force field parameters were assigned to the protein [224], whereas the Li, Song and Merz's 12-6-4 parameters for mono and divalent ions in TIP3P water were assigned to the ions [225,226]. The protein was solvated with an octahedral box of TIP3P explicit water molecules with 15 Å of minimal distance between the protein atoms and the box edges. Na<sup>+</sup> and Cl<sup>-</sup> ions were added to neutralize the system and simulate a physiological ionic concentration of 0.15 M. The system was simulated in Amber *pmemd.cuda* [190,191] using the following protocol: relaxation of the NaCl ions and water molecules using 1,000 steps of steepest descent and 1,000 steps of conjugate gradients minimization, keeping the whole protein and the metal co-factor harmonically restrained (force constant of 500 kcal/mol/Å<sup>2</sup>). 2,000 steps of steepest descent followed by 3,000 of conjugate gradients method for the whole system. We then performed gradual heating of the system from 0 to 300 K in 100 ps using the Langevin thermostat, keeping the backbone atoms and the co-factor restrained (force constant of 2 kcal/mol/Å<sup>2</sup>) and using an integration time step of 0.5 fs and constant volume conditions. Gradual release of the restraints followed from 2 to 0 kcal/mol/Å<sup>2</sup> in four phases of 50 ps each at constant pressure (1 atm), using an integration time step of 2 fs. We then ran 170 ns of NPT production simulation, recording the atomic coordinates every 2 ps. The SHAKE algorithm was used to keep the bonds involving hydrogens frozen [227]. The cutoff for long-range interactions was set to 9 Å. In order to assess the equilibration of the system we evaluated the time evolution of the mass-weighted RMSD calculated over the backbone atoms of the protein and the active site, using *cpptraj* from AmberTools12 [223]. In addition, the trajectory was visually analyzed

using Visual Molecular Dynamics (VMD) program [201]. In order to extract a set of representative and diverse conformations of the active site to be used a relaxed complex scheme (RCS) docking protocol [51], we performed RMSD-based clustering of the last 106 ns of the simulation, using the conformations extracted every 10 ps. Firstly, all the translation and rotation motions were eliminated by RMS-fitting the backbone atoms' positions of the trajectory to the first frame. The average-linkage clustering algorithm as implemented in *cpptraj* was then used to divide the frames in clusters and extract the centroid or representative conformation of each one, based on the positions of all the atoms of the active site. In general, the optimal number of clusters is not known a priori. In order to identify the optimal number of clusters in which the trajectory should be divided, we varied it from 1 to 200 and evaluated three metrics each time, namely the DBI index, pSF and the SSR/SST. A local minimum of the DBI, a maximum of the pSF and a kink in the SSR/SST plot are expected at the optimal cluster counting [80,169,228].

#### 5.4.3 Molecular Docking of Known Inhibitors and Pharmacophore Modeling

Molecular docking simulations were run for ERCC1-XPF endonuclease small molecule inhibitors, which are likely to bind to the XPF active site: compounds E-X AS7 from [85], 3, 14, 15, 21, 27, 33 and 34 from [204], 4, 13, 25, 29, 36, 37 from [205], and NSC16168 and NSC143099 from [206]. The selection criteria we used to select these compounds were the high potencies as ERCC1-XPF activity inhibitors and the specificities to the target. Indeed, we aimed to identify the intramolecular interactions which were essential for a strong and specific binding to the XPF active site. Different accessible protonation states and tautomeric forms of the compounds were obtained using the *MOE Database Wash* tool. The chemical structures of the small molecules are reported in

Figure 5.6. We extracted the centroid conformations of the top six clusters found by clustering the MD trajectory, in order to use them as target structures for the docking. In this set, we included also the lowest potential energy protein conformation found in the equilibrated fraction of the MD trajectory. Because all the selected active site inhibitors contained a metal-binding motif in their structure, we used MOE Site Finder to identify a potential binding zone in each structure, by selecting each time the highest ranked site close to the metal ion. For the docking

simulations, we used the Triangle Matcher placement algorithm [229] returning thirty poses, and the Rigid Receptor refinement method returning one final pose, as implemented in *MOE Dock*. The London dG method [221,230] was used to score the poses in both the steps. The binding energy of a ligand-receptor complex was calculated with the London dG method as

$$\Delta G_{LdG} = c + E_{flex} + \sum_{hbonds} c_{hb} f_{hb} + \sum_{metal-lig} c_m f_m + \sum_i \Delta D_i \quad (5.1)$$

where  $c$  is an empirically-derived term modeling the change in rotational/translational entropy upon binding,  $c_{hb}$  and  $c_m$  are the energies of ideal hydrogen bonds and metal ligations, respectively,  $f_{hb}$  and  $f_m$  range between 0 and 1 and measure the geometric imperfections of hydrogen bonds and metal ligations, respectively, and  $\Delta D_i$  is the desolvation contribution modeled by using a volume integral London dispersion [221]. Pharmacophore features common to all the docked compounds were automatically generated using the Consensus method in the Pharmacophore Editor in MOE and the Unified pharmacophore scheme [182].

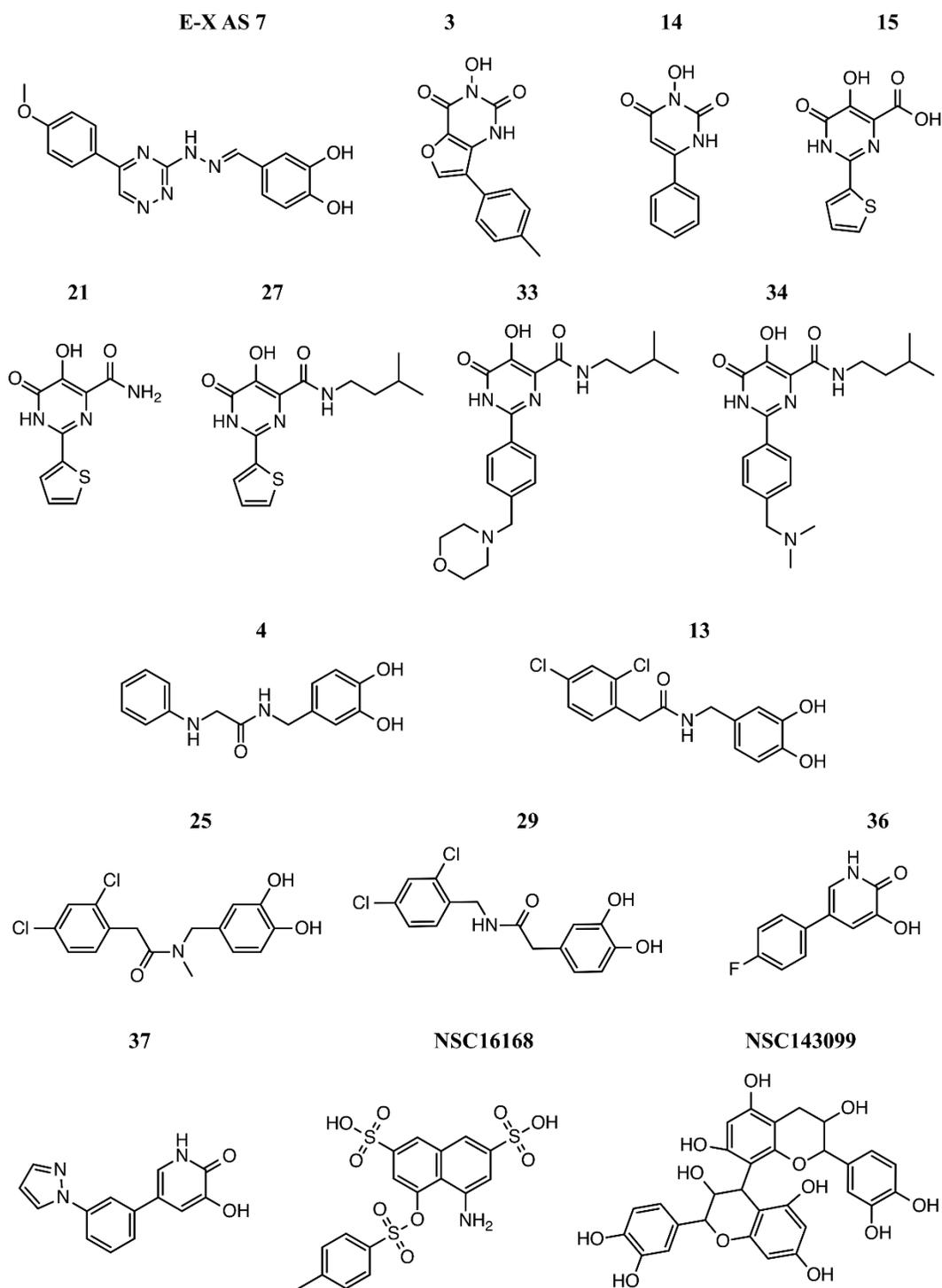


Figure 5.6. Chemical structures of the ERCC1-XPF inhibitors used to build the pharmacophore model after molecular docking simulations targeting the human XPF nuclease domain were run. All the small molecules investigated in this step included at least one metal-binding motif.

#### 5.4.4 Virtual Screening of the ZINC15 Database

To identify potential novel ERCC1-XPF inhibitors acting by binding to the XPF active site, we performed a VS campaign against the ZINC15 database, containing ~120 millions of compounds overall. Prior to performing the VS, the database needed to undergo different filtering and preparation steps, in order to reduce the enormous number of compounds and at the same time to consider different states of the retained small molecules. Hence, we downloaded all the compounds which satisfied the following ZINC15 filters: availability of 3D conformers, standard reactivity, commercial purchasability as wait-ok, standard protonation state at pH of 7.4, and charges ranging from -2 to +2. We then used again the *MOE Database Wash* tool to calculate other accessible protonation states and tautomers, and we filtered the resulting database in order to retain just compounds satisfying the Oprea's lead-like filter [184]. The Oprea's rules are a) the number of N or O atoms that are hydrogen bond donors must be maximum 5; b) the number of N and O atoms must be maximum 8; c) the molecular weight must be maximum 450; d) the logP must be between -3.5 and 4.5; e) the number of rings of size three through eight must be maximum 4; and, f) the number of rotatable bonds must be maximum 10. A compound is considered a lead-like one if its structure violates at most one of Oprea's rules. Our goal was to screen just lead-like compounds which can be optimized in drug-like compounds once the activities have been assessed. As the last preparation step, we generated up to five 3D conformations for each compound in MOE, imposing a strain limit of 4 kcal/mol. The pharmacophore obtained previously was then used to screen the resulting multi-conformational database, retaining just compounds with at least one conformation satisfying the pharmacophore. We then performed structure-based VS of the resulting ZINC15 subset of small molecules, using the same docking parameters described previously. The resulting top-scored poses were filtered

again using the same pharmacophore model, in order to retain just the binding modes satisfying the pharmacophore features. Duplicates of the same compound scoring worse than the top pose were removed.

## 5.5 Conclusions

The ERCC1-XPF endonuclease plays a primary role in several DNA repair pathways, including NER, ICL and DSB. Since in cancer cells these pathways are involved in the repair of damages caused by DNA-damaging cancer therapies, blocking its activity is expected to result in enhancement of the effect of such therapies. Inhibiting the endonuclease activity by using small molecules binding to the catalytic site of XPF is a relatively new strategy, which has not been fully explored yet. Indeed, despite the recent discovery of several XPF active site inhibitors, there is a lack of structural information about the mode of binding of these compounds, due mainly to the unavailability of experimental structures of the human XPF nuclease domain. Here, we generated a homology model for such a domain, based on templates that were carefully selected between all the structures of protein domains related to the human XPF one. Consequently, we used MD simulations and iterative clustering of the MD trajectory to identify dominant conformations of the active site and used the resulting set of structures as targets in molecular docking simulations of the most potent and selective known XPF inhibitors. As a result, we built a pharmacophore model elucidating the key interactions required for an effective ligand binding to the site, involving E712 and K727 and the coordinated Mg<sup>2+</sup> ion. A multi-step VS campaign was then performed to identify potential novel XPF inhibitors by sequentially filtering the ZINC15 database.

This work provides a detailed picture of the binding modes of small molecules to the human XPF active site. The results presented here can be effectively used in the rational design of XPF inhibitors, which are potent and specific to the target. In addition, we identified a set of commercially available chemical compounds, which can potentially show improved binding compared to the set of known inhibitors. In order to validate our model of ligand binding, mutation studies regarding the residues identified as essential for binding would be ideal. Based on previous studies, residue E712 would be the best candidate for a mutation study, as it is not conserved among other XPF-related proteins and its mutation does not affect the endonuclease activity. K727, on the other hand, would stop the activity when mutated [212]. Also, the top hits we reported in this study represent a good starting point to rationally develop optimized analogues, after their experimental validation as XPF binders and DNA repair inhibitors.

## Chapter 6 – Novel ERCC1-XPF Inhibitors for Resistant Colorectal Cancer Treatment Based on Targeting DNA Repair<sup>5</sup>

### 6.1 Abstract

The ERCC1-XPF heterodimer is a 5'-3' structure-specific endonuclease, which plays an essential role in several DNA repair pathways in mammalian cells. ERCC1-XPF is primarily involved in the repair of chemically-induced helix-distorting and bulky DNA lesions, such as cyclobutane pyrimidine dimers (CPDs), and DNA inter-strand crosslinks. Inhibition of ERCC1-XPF has been shown to potentiate cytotoxicity of platinum-based drugs and cyclophosphamide in cancer cells. In this study, the previously described ERCC1-XPF inhibitor 4-((6-chloro-2-methoxyacridin-9-yl)amino)-2-((4-methylpiperazin-1-yl)methyl)phenol (F06) was used as a reference compound. Following the outcome of docking-based virtual screening (VS), we synthesized seven novel derivatives of F06 that were identified *in silico* as being likely to have high binding affinity for the ERCC1-XPF heterodimerization interface by interacting with the XPF double helix-hairpin-helix (HhH2) domain. Two of the new compounds, 4-((6-chloro-2-methoxyacridin-9-yl)amino)-2-((4-cyclohexylpiperazin-1-yl)methyl)phenol (compound 3) and 4-((6-chloro-2-methoxyacridin-9-yl)amino)-2-((4-(2-(dimethylamino)ethyl) piperazin-1-yl)methyl) phenol (compound 4), were shown to be potent inhibitors of ERCC1-XPF activity *in vitro*. Compound 4 showed a significant inhibition of the removal of CPDs in UV-irradiated cells and the capacity to sensitize cells to UV radiation and cyclophosphamide.

---

<sup>5</sup> This chapter is under consideration for publication as Elmenoufy A.H.; Gentile F.; Jay D.; Karimi-Busheri F.; Yang X.; Soueidan O.M.; Weilbeer C.; Mani R.S.; Barakat K.H.; Tuszynski J.A.; Weinfeld M.; West F.G. Targeting DNA Repair in Tumor Cells via Inhibition of ERCC1-XPF in *ACS Cent. Sci.*

## 6.2 Introduction

The ERCC1-XPF heterodimer is a highly-conserved DNA repair endonuclease. It plays a pivotal role in nucleotide excision repair (NER) of bulky adducts and helix-distorting DNA lesions such as UV-induced pyrimidine-(6,4)-pyrimidone photoproducts (6-4PPs) and CPDs [85,203,231,232]. ERCC1-XPF is also involved in DNA inter-strand crosslink (ICL) repair [233] in cells treated with platinum-based and other chemotherapeutic agents such as cyclophosphamide and mitomycin C (MMC) [234]. There is also evidence that ERCC1-XPF participates in DNA double-strand break (DSB) repair [235,236]. It thus contributes significantly to the response of cancer cells to a range of DNA-damaging chemotherapeutic agents and radiotherapy. In the ERCC1–XPF complex structure, ERCC1 is considered to be catalytically inactive but rather regulates DNA–protein and protein–protein interactions, whereas the endonuclease activity is provided by XPF, which also contains an inactive helicase-like motif that is likely to be involved in protein–protein interactions and DNA binding [237,238].

Dimerization and localization of ERCC1 and XPF is essential for the enzyme's stability and endonuclease activity [239]. The dimerization of C-terminal regions of ERCC1 and XPF is the key interaction to form a stable heterodimer. C-terminal regions dimerize through the interaction of their HhH2 motifs [240]. It is thought that XPF acts as a scaffold for ERCC1 during protein folding, and ERCC1 may exhibit improper folding *in vitro* in the absence of XPF [241]. It was demonstrated that without dimerization the activity of ERCC1-XPF was abolished because neither protein was stable, and therefore they were rapidly degraded [121,150]. As a result, development of small molecule inhibitors that can disrupt the HhH2 domain interactions between ERCC1 and XPF would be expected to sensitize cancer cells to chemotherapy whose DNA-damaging effects are repaired by ERCC1–XPF-dependent pathways [203]. Moreover, rational

drug design methodology can be employed due to the availability of multiple experimental structures of the dimerized HhH2 domains (for example, PDB code 2A1J and 1Z00) [121].

Recently, McNeil *et al* employed an *in-silico* screening approach targeting three sites on the XPF HhH2 domain to identify possible inhibitors for the dimer. They discovered a small molecule able to inhibit the NER activity in melanoma cells and slightly sensitize them to cisplatin treatment. However, the reported  $K_d$  and  $IC_{50}$  values for this compound were suboptimal, in the medium-high  $\mu$ M range, and the cisplatin  $IC_{50}$  reduction was only 1.3-fold [85]. In addition, Chapman *et al.* [204,205], and Arora *et al.* [206] identified and optimized different small molecules targeting the active site on the XPF nuclease domain. These efforts resulted in several endonuclease inhibitors with  $IC_{50}$  in the nanomolar range, able to diminish NER activity and enhance the cytotoxicity of platinum-based drugs in cancer cells. Although the specificity of these inhibitors to the ERCC1-XPF endonuclease was assessed in some cases, the similarity between endonuclease active sites could still lead to off-target interactions of these compounds; as well as no structural insights of the ligand-protein complexes were provided. Finally, Yang *et al.* proposed the cellular delivery of therapeutic peptides mimicking the ERCC1 HhH2 domain (residues 220-297) as a promising alternative strategy to inhibit NER activity and sensitize cancer cells to DNA-damaging agents [242].

Development of small molecule inhibitors of the HhH2 domain interaction in our study would be expected to sensitize cells to radiotherapy whose DNA-damaging effects are repaired by ERCC1-XPF-dependent pathways. Jordheim *et al.* focused on developing small molecule inhibitors of the NER pathway acting through the inhibition of ERCC1-XPF heterodimerization and reported that F06 (also called NSC-130813 or NERI02) interacted with XPF, repressed the interaction between ERCC1 and XPF *in vitro* and sensitized cancer cells to MMC and cisplatin

[84]. This affirms that targeting this protein-protein interaction can enhance the cytotoxic activity of crosslinking agents such as cisplatin. The initial hit, F06, arose from a VS of a large compound library, and the study also provided a characterization of the XPF binding pocket and the binding mode of the compound to it [84]. F06 was predicted to interfere with the heterodimerization of ERCC1 and XPF, a necessary step to attain DNA repair activity.

Preliminary *in-vitro* assays confirmed that F06 shows promising inhibitory activity, and acts synergistically with cisplatin. However, the activity of F06 is suboptimal in terms of clinical properties including potency and safety, and a derivatization strategy, suggested by Jordheim *et al.* [84], was adapted to optimize the action of the compound. Using F06 as a reference hit, we employed computer-aided drug design techniques such as electrostatic mapping of the F06 binding pocket, molecular docking, pharmacophore modeling and molecular dynamics (MD)-based rescoring to rank F06 derivatives based on their predicted binding affinities to the XPF domain. The reference compound (compound 1) and top hits (compounds 2-8) were synthesized and tested for their ability to inhibit the *in-vitro* endonuclease activity of ERCC1-XPF, and the most active compound was further assessed as an inhibitor of the repair of UV-induced thymidine dimers in cells, as well as a sensitizing agent to UV radiation and cyclophosphamide.

## **6.3 Materials and Methods**

### *6.3.1 Molecular Docking of F06*

The first step of our VS study was to investigate the binding mode of the lead compound F06 [84] to the pocket on the XPF C-terminus. We used twenty XPF structures as single targets for a relaxed complex scheme (RCS) docking protocol, in order to accurately account for protein flexibility [185,186]. These structures were extracted from the Protein Data Bank (PDB) [32]

entry 1Z00 [241], reporting the NMR ensemble of the dimerization complex between the ERCC1 and XPF HhH2 domains. The XPF structures were optimized with a minimization process, details of which are reported in section B.1.1 in Appendix B. The binding site of each target was defined as the geometric center of the residues Y833, N834, P837, Q838, M856, K860, N861 and I862 on the XPF HhH2 domain, as reported in our previous study [84], where the residues are numbered according to the 1Z00 PDB file. Molecular docking simulations were performed using the Lamarckian Genetic Algorithm (LGA) [55] and the built-in scoring function [47] of Autodock4 [56]. Technical details regarding the docking simulations are reported in section B.1.1 in Appendix B.

### *6.3.2 Characterization of the Binding Pocket and Pharmacophore Modeling*

To characterize the binding site and the binding mode of F06, electrostatic maps were generated in the Molecular Operating Environment 2015 software package (MOE2015) [181] based on the docked pose of F06, using -2 kcal/mol for the potential iso-surfaces of hydrophobic, acceptor and donor probe atoms, and the Poisson-Boltzmann equation to compute the potentials. The maps were built within 4.5 Å of the ligand pose obtained from docking. Using the information from the docking and electrostatic mapping, a pharmacophore model for F06 was built with the MOE Pharmacophore Query Editor and the EHT scheme [182].

### *6.3.3 Docking-Based Virtual Screening of F06 Analogues*

57 analogues of the F06 structure were designed by replacement of the piperazine N-methyl group with other moieties expected to capture additional binding interactions of the piperazine ring. The structures were prepared using Schrödinger LigPrep [177] to account for multiple tautomers, protonation states and low-energy ring conformations, using the same approach as for

F06 (section B.1.1 of Appendix B). The molecular docking simulations were performed with MOE Dock. Only the XPF structure from the best F06 complex was used as a target for the VS study. Different conformations of the analogues were obtained using Conformation Import. The previously built pharmacophore model was used for the placement step, in which 30 poses are returned according to the London dG scoring method [221]. To account for the local arrangement of the pocket residues upon ligand binding, the Induced Fit method was selected for the refinement step, where the side chains of the binding pocket were left to move freely. At the end, one pose scored with the Generalized Born Volume Integral/Weighted Surface Area (GBVI/WSA) function was returned [221]. Water/octanol partition coefficients (logP) of the molecules were calculated in MOE using the SlogP function [243], which takes into account the given protonation state of the molecule under examination. Ligand efficiencies were calculated as the ratio between the computed binding energies and the number of heavy atoms of each analog.

#### *6.3.4 Molecular Dynamics Simulations and MM/GBSA Rescoring of the Analogues*

To calculate an average binding energy for the hits, we performed 2 ns of MD simulations for the top ranked compound-XPF complexes. Details of the preparation method and the parameters used in these simulations are reported in section B.1.2 of Appendix B. Free energy calculations were performed over the trajectories with the Molecular Mechanics-Generalized Born Surface Area (MM/GBSA) method, using the MMPBSA.py script [197]. For technical details regarding the calculations refer to section B.1.2 of Appendix B. The calculations were performed on snapshots extracted every 10 ps from the MD trajectories, and per-residue decompositions of the binding energies were also performed for the residues within 10 Å of any analog atom at the beginning of the simulations. We also calculated the entropic contribution of ligand binding

using the normal mode analysis (NMA) method [244]. The final binding energies used to rank the compounds were calculated combining the MM/GBSA and entropy contributions with the Equation B.1 in Appendix B.

### *6.3.5 General Experimental Procedures for Top F06-Based Analogues Synthesis*

Synthesis and characterization of compounds 1-8 were reported in Appendix B, section B.2.2. Reactions were carried out in flame-dried glassware under a positive argon atmosphere unless otherwise stated. Transfer of anhydrous solvents and reagents was accomplished with oven-dried syringes or cannulae. Solvents were distilled before use. Chemicals were purchased from Sigma Aldrich Inc., and were used without further purification. Thin layer chromatography was performed on glass plates pre-coated with 0.25 mm silica gel. Flash chromatography columns were packed with 230-400 mesh silica gel. Proton nuclear magnetic resonance spectra ( $^1\text{H}$  NMR) were recorded at 500 MHz, and coupling constants (J) are reported in hertz (Hz). Standard notation was used to describe the multiplicity of signals observed in  $^1\text{H}$  NMR spectra: broad (br), multiplet (m), singlet (s), doublet (d), triplet (t), etc. Carbon nuclear magnetic resonance spectra ( $^{13}\text{C}$  NMR) were recorded at 125 MHz and are reported (ppm) relative to the center line of the triplet from chloroform-d (77.0 ppm) or the center line of the heptuplet from methanol-d<sub>4</sub> (49.0 ppm). Infrared (IR) spectra were measured with a FT-IR 3000 spectrophotometer. Mass spectra were determined on a high-resolution electrospray positive ion mode spectrometer.

### *6.3.6 ERCC1-XPF Protein Preparation*

Human ERCC1-XPF wild-type protein was obtained as previously described [245]. Basically, the recombinant protein was expressed from a bicistronic plasmid (kindly provided by Dr. Richard Wood, University of Texas MD Anderson Cancer Center, Smithville, TX) in the *E. coli*

BL21 (DE3) strain. Since both XPF and ERCC1 contained a polyhistidine (His-6) tag, the proteins extracted from *E. coli* were incubated with a ProBond Nickel-Chelating Resin (Thermo Fisher Scientific). Protein eluted from the Ni affinity column was subsequently loaded into a Hi-trap heparin column (GE Healthcare). Fractions recovered from the heparin column that contained ERCC1-XPF were dialyzed, concentrated and stored at -80°C in 10 mM HEPES pH 7.4, 2.5 mM  $\beta$ -mercaptoethanol, 0.01% CHAPS, 0.25 mM EDTA, 50% glycerol and 25 mM NaCl.

### *6.3.7 Microplate Fluorescence Incision Assay*

We followed a previously described protocol [85,245]. Briefly, reactions were carried out in 384-well black, flat-bottomed microtiter plates (OptiPlate – 384 F, Perkin Elmer) in a total volume of 20  $\mu$ l containing the indicated concentrations of inhibitor compounds, 100 nM stem-loop substrate [6-FAM-5'-CAGCGCTCGG(20T)CCGAGCGCTG-3'-dabcyl], 25 ng ERCC1-XPF, 50 mM Tris-Cl pH 8, 20 mM NaCl, 0.5 mM DDT and 0.75 mM MnCl<sub>2</sub> at 25°C. Fluorescent readings were obtained using a FLUOstar Optima fluorimeter (BMG Labtech) with Optima software at an excitation of and emission wavelengths of 485 and 520 nm respectively, for 12 minutes.

### *6.3.8 Steady-State Fluorescence Assays*

Steady-state fluorescence spectra were measured at room temperature on a Perkin-Elmer LS-55 spectrofluorometer (Freemont, CA) with 5 nm spectral resolution for excitation and emission using 30-80 nM solution of purified recombinant ERCC1-XPF protein complex. Protein fluorescence was excited at 295 nm, and fluorescence emission spectra were recorded in the 300-400 nm range: changes in fluorescence intensity was monitored at the emission maximum (330

nm). In studying the effects of inhibitors on protein fluorescence intensities, additions to protein samples were made from inhibitors stock solutions, keeping the protein dilution below 3%.

### 6.3.9 Cell Culture

The human colorectal cancer HCT-116 cell line was obtained from the American Type Culture Collection (ATCC). The cell population was expanded immediately after arrival, aliquoted and stored frozen in liquid nitrogen. Freshly thawed cells were used for each experiment. The cells were cultured in a 1:1 DMEM/F12 media supplemented with 10% FBS, 50 units/mL penicillin, 50 µg/mL streptomycin, mM l-glutamine, 0.1 mM nonessential amino acids, 1 mM sodium pyruvate and maintained under 5% CO<sub>2</sub> in a humidifier incubator at 37°C. All the supplies for cell culture were obtained from Gibco/BRL.

### 6.3.10 Cellular Repair of Cyclobutane Pyrimidine Dimers

We followed the protocol of Mirzayans *et al.* with minor modifications [246]. Approximately 1 x 10<sup>5</sup> HCT-116 cells were seeded on each coverslip and allowed to attach overnight. Medium was then removed and cells were treated for one hour with the desired compound. The medium was removed, and cells were exposed to 8 J/m<sup>2</sup> UV-C radiation (G15T8 254 nm lamp, Ushio America Inc, Cypress, CA) followed by adding fresh medium containing the compound. Plates were incubated at 37°C for different periods of time up to 24 hours and fixed in 50:50 methanol/1X phosphate-buffered saline (PBS) solution, followed by replacing the methanol/PBS solution with 100% methanol and incubation in -20°C. After 20 minutes, methanol was removed, and cells were treated with PBS for 5 minutes in room temperature. After fixing, cells were permeabilized in 0.5% Triton/PBS and washed with PBS, denatured in 2 N HCl, and neutralized by twice washing with 0.1 M borate buffer pH 8.5. Cells on cover slips were washed once with

PBS followed by blocking with 5% non-fat dry milk/PBS for 30 minutes. Mouse anti-thymine dimer monoclonal antibody (Cat. No. MC-062, Kamiya Biomedical Company, Seattle, WA) was applied to the cover slips for one hour in the dark and at room temperature. Cells were then washed with PBS containing 0.1% Tween-20 and incubated with rabbit anti-mouse IgG-Alexafluor antibody (Invitrogen, Carlsbad, CA) followed by two washes with PBS/(0.1%)Tween-20. Cells on coverslips were then washed with PBS, rinsed with water, and mounted on slides using DAPI glycerol mounting solution. Slides were kept at 4°C before fluorescent microscopic evaluation and measurement of fluorescence intensity with MetaXpress Version: 6.2.1.704 software (Molecular Devices, Sunnyvale, California).

#### *6.3.11 Clonogenic Survival Assay*

UV treatment: HCT-116 cells (100-800 cells depending on the UV dose) were plated in triplicate in 60-mm petri dishes. The cells were incubated overnight at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> to allow for cell attachment. Medium was then removed, and cells were treated with 1 or 2 µM compound 4 (TFA salt) for one hour. Medium was removed again, and cells were exposed to increasing doses (0-10 J/m<sup>2</sup>) of UV-C radiation and then incubated for a further 10 days in the presence of inhibitor at 37°C in a 5% CO<sub>2</sub> atmosphere to allow for colony formation. After this period plates were stained with crystal violet and colonies were counted using a Colcount instrument (Oxford Optronix, Abingdon UK). Finally, plating efficiency and surviving fraction were calculated.

Cyclophosphamide treatment: A similar protocol was followed as described for the UV treatment, except that cells were treated with 1 or 2 µM compound 4 (TFA salt) for 4 hours followed by addition of increasing doses of cyclophosphamide (0-300 µM). After 24 hours

medium was replaced with fresh medium containing compound 4 alone. Plates were incubated for another 8 days at 37°C in a 5% CO<sub>2</sub> atmosphere for colony formation. After this period plates were stained with crystal violet, colonies were counted, and plating efficiency and surviving fraction were calculated.

## 6.4 Results and Discussion

### 6.4.1 Molecular Docking of F06

The first step was to investigate the binding mode of the lead compound, F06, to the pocket on the XPF C-terminus, followed by functionalization and extension of the piperazine ring in the F06 compound to: 1) provide potential key interactions via hydrogen bond formation or hydrophobic interactions, and 2) optimize the physicochemical properties and binding affinity for better potency and reduced toxicity of the F06-based compounds.

The binding energy for the best docked conformation of the F06 molecule to the XPF structures was -10.23 kcal/mol, as calculated by the Autodock scoring function. The compound showed a high shape complementarity with the XPF pocket, which interacts with the F293 residue of ERCC1 in the dimerized complex (Figure B.2A in Appendix B). Three ligand-receptor hydrogen bonds were observed in the docked pose, namely with the side chain of E829, the backbone of N834 and the backbone of K860. In addition, the hydrophobic core of F06, constituted by the three aromatic rings, was positioned within a hydrophobic zone constituted by XPF residues Y833, Q838, M856 and H857 (Figure B.2B in Appendix B). Also, the binding mode of the ligand was in accordance with the spatial distribution of the affinity maps calculated in MOE (Figure B.2C in Appendix B).

#### 6.4.2 Virtual Screening of F06 Analogues

To correctly place the scaffold common to all the analogues within the binding site, a pharmacophore model based on the binding pose of F06 was built using MOE. This model included six features, namely four aromatic moieties, one hydrogen bond donor, and one hydrogen bond acceptor features (Table B.1 in Appendix B). Upon generation of such features using MOE, we manually adjusted their radii consistently with the spatial distribution of the affinity maps. The total number of molecular structures that were docked was 57, accounting for all the different states generated during the preparation step. After the docking was performed, just the top scored molecular structure for each analog was retained. The results were ranked according to their Generalized Born Volume Integral/Weighted Surface Area (GBVI/WSA) score, as calculated by MOE. Then, we performed the MD simulations of the complexes, and rescored the analogues using the Molecular Mechanics-Generalized Born Surface Area (MM/GBSA) method. A subset of the top ranked hits was selected for chemical synthesis.

The results of the docking-based VS for this subset are reported in Table B.2. The first two hits (compounds 3 and 4) were particularly interesting because of their binding energies in the range of the lead compound, reasonable logP values and conserved ligand efficiencies with respect to the hit structure. The decomposition of the binding energies among the residues of the binding pocket revealed a similar pattern of ligand-receptor interactions of the top hits (compound 3 and 4) and the lead compound.

The most important residues for ligand binding of F06 analogues were observed to be E829, N834 and V859 where the interaction energy is dominated by the electrostatic component, and Y833, M856, H857 and K860 where the van der Waals interactions dominated the binding

(Figure 6.1A). A more detailed analysis of the binding mode of one of our best hits, compound 4, is reported in Figure 6.1B. The structure of the analogue common to the lead compound interacts in a similar fashion with the residues constituting the binding site, although the backbone of V859 replaces K860 as hydrogen bond partner during the simulation. In addition, compound 4 interacted with several water molecules surrounding the binding site, during the 2-ns simulation (Figure 6.1C), although the charge-charge interactions with the E829 and E831 residues were maintained in terms of average magnitude, when compared with the lead compound.

#### 6.4.3 Synthesis of F06-based analogues

Synthesis of compounds 1-8 was achieved through a one-pot sequential addition reaction in 3 steps as shown in Figure B.1 in Appendix B. Mannich reaction of p-acetamidophenol with formaldehyde and appropriate secondary amine in 2-propanol was carried out under reflux for 12 h. The solvent and excess of the unreacted formaldehyde from the resulted mixture was removed under vacuum and without isolating the compound, the resulting viscous residue was treated with 6 M HCl to deacetylate the acetamido group to result in primary amine as depicted in Figure B.1 in Appendix B. Afterwards, equimolar of 6,9-dichloroacridine was added in the mixture to result in compounds 1-8 in moderate to good yields after isolation. The reaction is general, facile and reproducible. The intermediates expected to be formed are illustrated in Figure B.1 in Appendix B.

According to the *in-silico* screening, seven compounds were selected to be synthesized among the top hits due to their promising ligand efficiencies and binding energy values in comparison to F06. Derivatization focused primarily on extension of the piperazine ring with different functionalities (i.e alicyclic (compound 3), substituted aromatics with electron donating group

(EDG) or EWG (compounds 2, 5, 6 and 7), and an aliphatic chain installed with HBA atom (compound 4) as shown in Table 6.1. Compound 8 was synthesized as a control to investigate whether the piperazine ring is crucial for the activity of compounds.

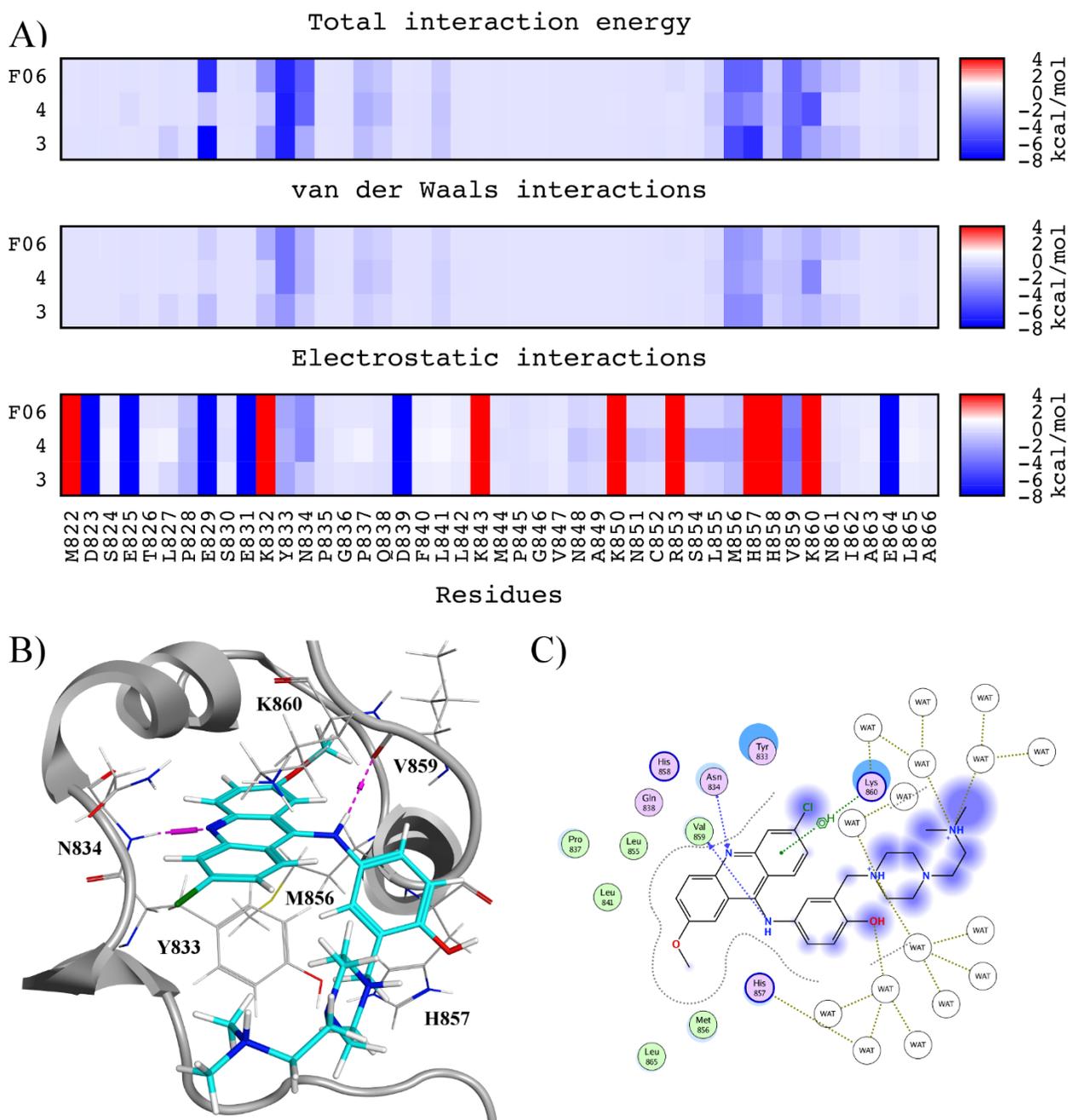
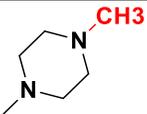
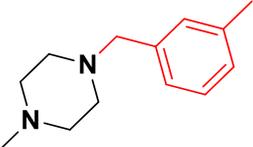
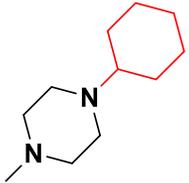
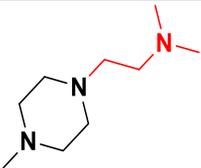
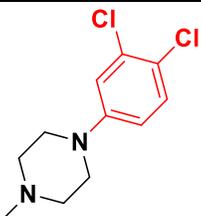
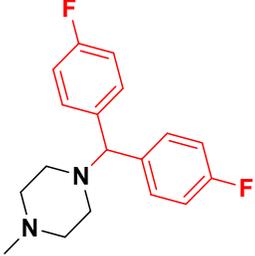
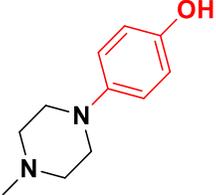


Figure 6.1. Analysis of the binding features of the top hits. A) Per-residue decomposition of ligand-receptor binding energy. Compound 4 showed a slightly different pattern of total

interaction energy, while the lead compound and compound 3 had more similar patterns. Blue cells indicate favorable interactions, while red cells indicate unfavorable interactions. B) Lowest potential energy snapshot extracted from the MD simulation of compound 4 bound to XPF. Hydrogen bonds are colored in purple. The highest pocket contributors in terms of total interaction energy are labelled. C) Interaction diagram for compound 4 and surrounding residues and water molecules. K860 was a strong contributor to the binding energy due to arene-H interaction with one of the aromatic rings of compound 4. The diagram refers to the lowest potential energy snapshot extracted from the MD simulation.

Table 6.1. F06-based analogues functionalized with different R substituents.

Compound	R	Yield
1 (reference)		78 %
2		73 %
3		81 %
4		80 %
5		71 %

6		65 %
7		59 %
8	H	73 %

#### 6.4.4 Inhibition of ERCC1-XPF Endonuclease Activity

We used an *in-vitro* real-time fluorescence-based assay to assess the inhibitory effect of the synthesized compounds (except compound 2 due to its instability over time) on ERCC1-XPF endonuclease activity. This assay has been previously described [245], and utilizes a stem-loop substrate (composed of a 10mer-duplex stem and a 20mer-oligodT single-strand loop), labelled on the 5'-terminus with 6-FAM and 3' with the quencher dabcyI. Upon ERCC1-XPF cleavage in the 10mer-duplex region an 8-base 5' FAM-labelled product is released resulting in a fluorescent signal (Figure 6.2A, control).

The other tracings in Figure 6.2A indicate to what extent the different compounds (10  $\mu$ M each) could inhibit incision of the substrate.

As shown here, three compounds exhibited a marked capacity to inhibit the incision activity of ERCC1-XPF, i.e. compounds 1 (F06), 3 and 4. For these three compounds, different

concentrations of the drugs were plotted against the initial velocity ( $V_o$ , change in relative fluorescence units (RFU)/time) of the enzyme.

Figure 6.2B (inset) shows an example of data obtained with compound 4. Half-maximum inhibitory concentrations ( $IC_{50}$ ) for these compounds were estimated from at least three different experiments for every compound (Table B.3).

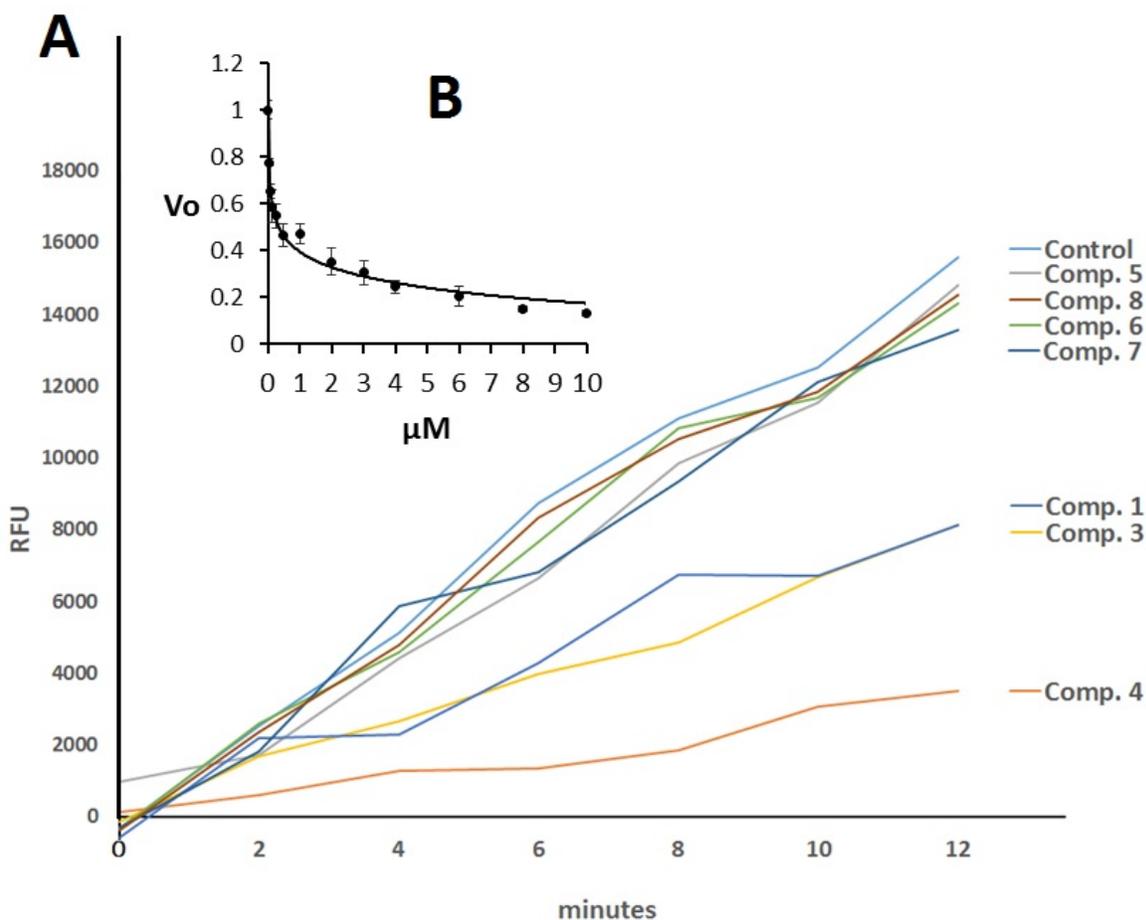


Figure 6.2. A microplate fluorescent assay to measure inhibition of ERCC1-XPF endonuclease activity by the different compounds. Incubation of the stem-loop substrate with ERCC1-XPF resulted in the release of a fluorescent 8-base fragment. A) Shows the increase in fluorescence (RFU) with time. A representative tracing of the effect of the different compounds (8  $\mu M$  each)

on the incision activity is shown. B) (Inset) shows a representative plot of enzyme rate ( $\Delta$ RFU/time) vs compound (4) concentration.

Since acridine compounds are known to bind nucleic acids, the possibility existed that a disruption of the structure of the stem-loop substrate by compound 4 could have been responsible for the inhibition observed in the ERCC1-XPF kinetic assays. To rule out this possibility the experiment shown in Figure B.4 in Appendix B was carried out. Pre-incubation of ERCC1-XPF with compound 4 (3  $\mu$ M) in the reaction medium resulted in a time-dependent loss of enzyme activity. The reactions were started either with addition of the substrate (T0) or with substrate that had been pre-incubated in the reaction medium in the presence of compound 4 (3  $\mu$ M) for 12 minutes (T12). The data indicate that the time-dependent inactivation of ERCC1-XPF by compound 4 was the consequence of the interaction of the drug with the enzyme, and not altered by addition of substrate that had been pre-incubated with the inhibitor. The data in Figure B.4A in Appendix B were obtained with a concentration of compound 4 approximately 9 times higher than the observed  $IC_{50}$  value of the compound and in the presence of 10% DMSO. When the same test was carried out in the presence of 5% DMSO, a similar time-dependent inhibition was observed and the substrate exhibited comparable behavior (Figure B.4B in Appendix B). (The  $IC_{50}$  values reported in the present study were obtained by incubating ERCCI-XPF in the presence of the indicated concentrations of the drug for 30 minutes in the reaction medium containing 5% DMSO.)

#### *6.4.5 Inhibitor Binding to ERCC1-XPF*

Intrinsic fluorescence spectroscopy (of the protein tryptophan residues) was utilized to study the binding affinity of compound 4 (active compound) and compound 5 (non-active compound) to ERCC1-XPF (Figure 6.3). Addition of 2  $\mu$ M of compound 5 had no significant effect on protein

fluorescence and the observed fluorescence quenching at 330 nm was only  $3 \pm 1\%$ , thus providing no evidence of any interaction. In contrast, addition of compound 4 induced nearly  $23 \pm 2\%$  quenching in protein fluorescence at 330 nm, clearly indicating interaction of compound 4 with ERCC1-XPF complex as depicted in Figure 6.3B. Binding affinity (in terms of dissociation constant,  $K_d$ ) of compound 4 for ERCC1-XPF complex was determined by following fluorescence quenching (a measure of ligand binding) as a function of ligand concentration. A representative plot of relative fluorescence intensities versus the concentration of compound 4 is shown in Figure 6.3C (inset). Nonlinear regression analysis (GraphPad Prism Software, San Diego, CA) of the binding data was carried out as described in our earlier paper [247], and revealed a unimodal binding with a  $K_d$  value of  $100 \pm 5$  nM as also shown in Figure 6.3.

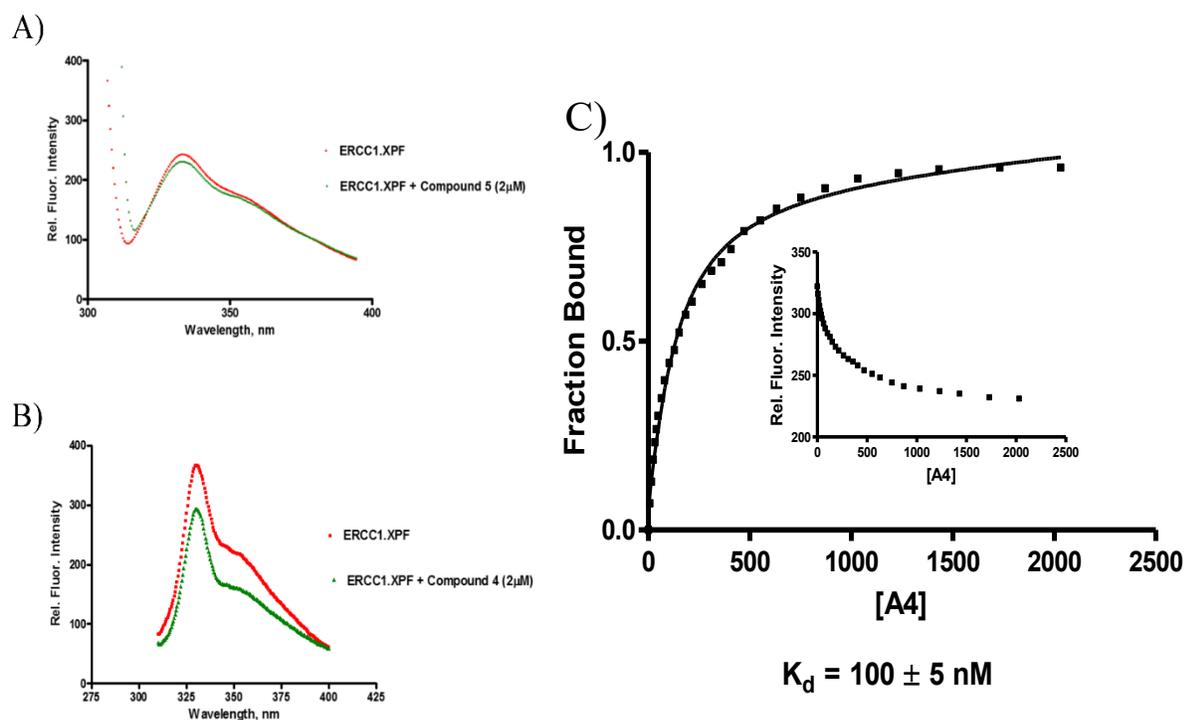


Figure 6.3. Fluorescence spectroscopy. A) Fluorescence of ERCC1-XPF (13 nM and 20 nM) incubated with compound 5 (negative control), and B) compound 4 (active inhibitor)

respectively. C) Unimodal binding pattern and the binding affinity of compound 4 with ERCC1-XPF (70 nM).

#### 6.4.6 Inhibition of Cellular Repair of Cyclobutane Pyrimidine Dimers

Immunofluorescence results of the exposure of HCT-116 colorectal cells to 8 J/m<sup>2</sup> treated with compound 4 showed a significant inhibition of NER determined by the removal of CPDs compared with control cells over 24 hours as depicted in Figure 6.4. Based on our results and others [246], approximately 80% of CPDs are removed over 24 hours after exposure of HCT-116 cells, but this was reduced to approximately 60% in the presence of compound 4. On the other hand, compound 5, which demonstrated very limited inhibition *in vitro*, failed to inhibit the cellular removal of CPDs. Figure 6.4B shows the relative quantification of CPDs after treatment with positive and negative compounds.

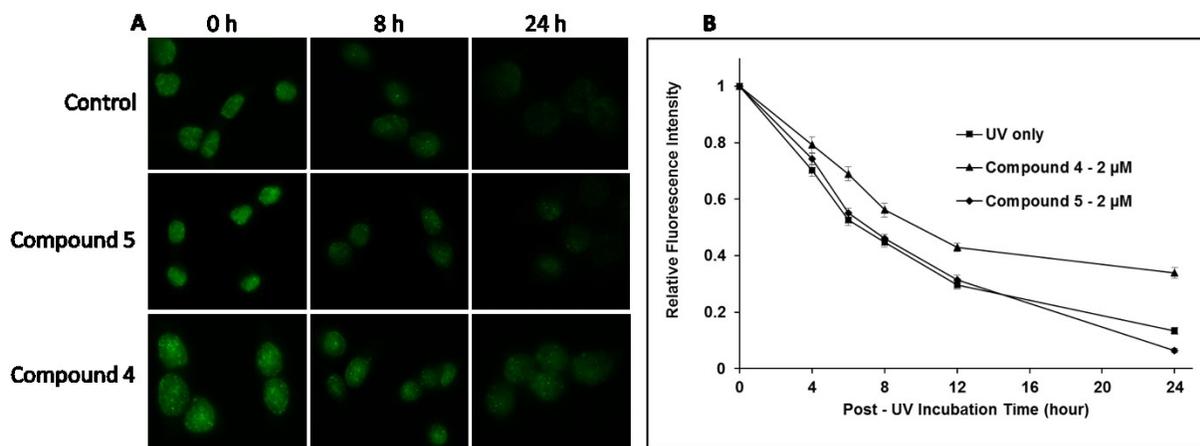


Figure 6.4. A) Immunofluorescence images of the UV-based assay for detecting CPDs in cells treated with compounds 4 and 5. B) Normalized fluorescence intensity of the treated cells.

#### 6.4.7 Increased Sensitization to UV Radiation and Cyclophosphamide

We also evaluated the effectiveness of compound 4 (1 and 2  $\mu\text{M}$ ) to sensitize cells to UV irradiation based on clonogenic survival as shown in Figure 6.5A. In accord with the repair data, 2  $\mu\text{M}$  compound 4, significantly reduced survival of the UV irradiated cells.

We then examined the combined effect of compound 4 (1 and 2  $\mu\text{M}$ ) on cellular survival following exposure to the DNA interstrand crosslinking agent cyclophosphamide. The survival curves shown in Figure 6.5B indicate that, at a concentration of 2  $\mu\text{M}$ , compound 4 significantly sensitized the cells to cyclophosphamide (starting at 50  $\mu\text{M}$ ) and no colonies were detected on the plates with a cyclophosphamide concentration of 300  $\mu\text{M}$ .

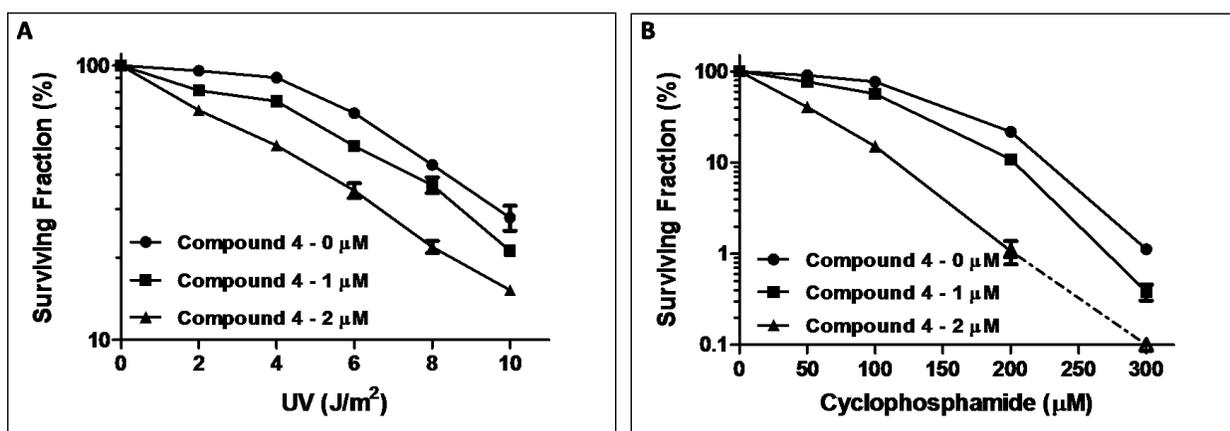


Figure 6.5. A) Survival of HCT-116 cells exposed to increasing doses of 254 nm UV radiation and treated with 1 and 2  $\mu\text{M}$  compound 4 determined by the clonogenic survival assay. B) Survival of HCT-116 cells exposed to increasing doses of cyclophosphamide and treated with 1 and 2  $\mu\text{M}$  compound 4 determined by the clonogenic survival assay.

## 6.5 Conclusion

The ERCC1-XPF heterodimer is a structure-specific endonuclease, which is required especially for NER and ICL DNA repair pathways. Although its action is essential to maintain genome integrity and to protect against damage-induced mutations, as part of the NER and ICL machinery it can counteract the effect of DNA damaging therapies such as platinum-based chemotherapy and radiotherapy. A promising approach to enhance the effect of such therapies is to inhibit the action of DNA repair in cancer cells using small molecules. In this work, we used a computational drug design workflow to provide a rational design for novel analogues of the F06 molecule, a lead inhibitor targeting the dimerization between XPF and ERCC1, which is required for endonuclease activity. We identified seven compounds whose *in-silico* simulations predicted attractive properties, namely binding affinities and ligand efficiency towards an XPF site on the dimerization interface. The synthesis of the computationally designed compounds was successfully carried out using a feasible, robust and reproducible synthetic strategy. Interestingly, compound 4 has good physicochemical properties i.e reasonable logP and ligand efficiency values, and small molecular weight, thus compound 4 is a potential candidate that obeys Lipinski's rules. Following structure-activity relationship studies and *in-vitro* screening, this approach yielded compounds 3 and 4 as potent inhibitors of ERCC1-XPF activity. An *in-vitro* ERCC1-XPF endonuclease assay identified compound 4 as the best ERCC1-XPF inhibitor with an IC<sub>50</sub> value of 0.33  $\mu$ M compared to 1.86  $\mu$ M for F06. In addition, the K<sub>d</sub> value for this compound was experimentally measured as 100 nM. Compound 4 also showed a significant inhibition of the removal of cyclobutane pyrimidine dimers compared with control cells after exposure of HCT-116 cells to UV radiation and sensitized the cells to UV and cyclophosphamide-induced cytotoxicity, indicating inhibition of NER and ICL repair. We

suggested that the *in-vitro* activity of compound 4 is well correlated with our *in-silico* results through enhanced van der Waals binding interactions with V859 and charge-charge interaction with E831 relative to the binding patterns of either F06 or the other non-active compounds. In summary, the use of *in-silico* modelling to design a superior compound to F06 led to compound 4 that can potentially be used in combination with other existing DNA-damaging therapies to amplify their effects by sensitizing cancer cells.

## **Chapter 7 - Conclusions and Future Works**

In this thesis I discussed the application of CADD techniques to identify novel DNA repair NER inhibitors. The target of my research has been the ERCC1-XPF endonuclease, a heterodimeric complex responsible for the excision step of NER and involved in drug resistance to platinum-based chemotherapy. The second chapter of the thesis provided a review of the employed CADD methods. Theoretical details are provided for molecular dynamics simulation techniques, molecular docking, and structure/ligand-based virtual screening (S/LBVS). The third section of the thesis is a systematic review of the state-of-the-art of NER proteins modeling and CADD studies targeting the pathway. The three successive chapters presented the results obtained by targeting with CADD three different sites of the ERCC1-XPF complex, namely the interaction site with XPA, the active site and the dimerization interface. In this conclusive chapter, I summarize the results achieved in each thesis section, and I suggest possible future directions which I hope will be pursued with the final goal to translate ERCC1-XPF inhibitors from the research environment to the clinical one.

### **7.1 Identification of XPA-ERCC1 Inhibitors Using Virtual Screening**

In NER, the recruitment of ERCC1-XPF endonuclease to the damaged DNA zone occurs through an interaction with XPA protein with the ERCC1 central domain. If this does not happen, the NER machinery is blocked. A drug able to inhibit the XPA-ERCC1 interaction can therefore prevent repair of damages induced by DNA damaging cancer therapies, although this would be limited to NER and not to other pathways in which ERCC1-XPF is involved (e.g. ICL repair pathway). The choice of this target is encouraged by the near-to-optimal results of cisplatin in the treatment of testicular cancer, due to low expression of XPA in testicular germ

cells. Early efforts in this direction were made by using VS to identify small molecules able to bind to the central site of ERCC1.

The study presented in Chapter 4 was performed by conducting a VS campaigns to identify potential, more potent compounds similar to the top hits identified in [80] and able to sensitize cancer cells to UV radiation and synergize cisplatin action. The study had two distinct goals: 1) Identify compounds similar to NERI01 and compound 10 (the two original hits) but with improved potency and drug-like properties. 2) Identify compounds with a different scaffold to propose new starting points for drug optimization. In order to accomplish the first objective, a ligand-based VS campaign as pursued using the two original hits as input. Using the Tanimoto score (Equation 2.7) as comparison score and the turbo similarity search approach, it was possible to quickly filter the whole PubChem repository (~68,370,000 compounds) and retain a subset of compounds similar to the two hits. Because of the goal was to identify molecules with drug-like properties, ADMET Predictor was used to furtherly refined the subset. This subset was then subjected to a multi-step structure-based VS protocol to predict the binding poses and rank the compounds by their computed binding affinities. The NMR structure of the ERCC1 central domain bound to XPA was chosen as target for this step. On the other hand, four NCI compound repositories were screened using a pharmacophore model based on the binding pose of the two hits. The use of these diverse databases and the pharmacophore, which identifies compounds sharing similar moieties but not necessarily similar structures, led to a filtered set of new different scaffolds. A structure-based VS protocol similar to before was adopted then to rank the resulting compound based on their binding affinities. In future, the top ranked compounds deriving from both the subsets should be purchased and validated experimentally. Once active compounds have been identified, structures coming from the similarity search will in theory be

already drug-like and not many steps of derivatives would be needed to optimize their properties. On the other hand, compounds resulting from the pharmacophore screening are in fact new scaffolds which can be fully explored through CADD medicinal chemistry techniques to improve binding affinities and drug-like properties.

## **7.2 Exploring the ERCC1-XPF Active Site as Potential Target for Inhibitors**

The endonuclease activity of ERCC1-XPF occurs through the catalytic site present on the XPF nuclease domain, where the damaged DNA strand is cleaved at its 5' extremity. Targeting this site to stop ERCC1-XPF-mediated repair would present many advantages, due to the presence of metal ions in the site (ideal for metal chelators), weak contacts made by the domain containing the active site and the DNA, and previous successful studies targeting similar sites in DNA repair enzymes. In addition, the first series of inhibitors targeting the XPF active site were recently reported. Nevertheless, targeting directly the ERCC1-XPF nuclease activity is not an easy task. First, the 3D structure of the human XPF nuclease domain has not been solved experimentally yet. Second, the active site is highly conserved in other related proteins, and a detailed description of the key binding residues is required to design highly-specific ERCC1-XPF inhibitors, in order to limit off-target interactions. The work reported in Chapter 5 of this thesis employed CADD techniques in order to address both these issues. Additionally, I performed a VS campaign aimed to identify new inhibitors of the ERCC1-XPF nuclease activity.

The first step of the study was to obtain a 3D model of the human XPF nuclease domain. Using bioinformatics techniques, we identified similar domains with a publicly available experimental structure. The nuclease domain was then built using the HeF protein from *Pyrococcus furiosus* as template for the homology modeling. A molecular dynamics simulation was then performed for

the model, followed by RMSD-based clustering of the active site residues to extract dominant conformations. The resulting structures were then used as single targets for molecular docking simulations, where the known XPF active site inhibitors were used as ligands. The docking poses allowed to identify a repeated binding pattern involving residues E712, K727 and the coordinated Mg<sup>2+</sup> ion, as well as to build a dynamic pharmacophore, accounting for multiple receptor conformations. In this way, we described in details the features required for successful ligand binding to the XPF active site. The pharmacophore model was successively employed to filter lead-like compounds from a subset of nearly 80,600,000 molecules from the ZINC15 database, in order to retain only the ones with the appropriate chemical features. Finally, SBVS was used to rank the compounds based on their predicted binding affinity. To summarize, the results of Chapter 5 provide, for the first time, a detailed biophysical picture of small molecule inhibitors binding to the XPF active site, opening a new venue for rational drug design targeting this domain. At the same time, future steps will include the experimental validation of the potential top hits and the design of more potent and drug-like analogues.

### **7.3 Targeting the ERCC1-XPF Heterodimerization Interface with Small Molecules to Disrupt the Functional Complex**

The last strategy to inhibit ERCC1-XPF encompassed in this thesis is the target of the protein-protein interaction. Heterodimerization of ERCC1-XPF is required to generate a functional complex with full endonuclease activity. Successful inhibition of the heterodimerization would be ideal because, differently to targeting the XPA-ERCC1 interaction, will be effective against all the activities of the complex and not only the NER-related one. Also, it will be less difficult to design drugs specific to the ERCC1-XPF interface rather than to the XPF active site, well conserved in similar endonuclease. Targeting the ERCC1-XPF interface, however, is challenging

due to the hydrophobic, strong nature of the interaction between the two proteins. Chapter 6 of this thesis reports the results of a CADD effort for the pre-clinical development of inhibitors of the ERCC1-XPF dimer. Chemical synthesis and testing of the compounds in protein-based assays and cell lines were performed at the University of Alberta by the Alberta DNA Repair Consortium team.

Our lab previously identified a promising hit compound able to disrupt the ERCC1-XPF dimer and stop the repair activity. The compound, named F06, was discovered by carrying out an extensive VS campaign targeting a binding pocket at the XPF of the interface. Starting from this lead structure, we built a library of analogue structures and carried out molecular docking simulations followed by MM-GBSA rescoring to identify a limited set of optimized compounds to synthesize and test. Following our rank of compounds, top molecules were synthesized. Following *in-vitro* assessment, compound 4 showed the most promising potential as drug candidate, with an observed  $IC_{50}$  value of 0.33  $\mu$ M compared to 1.86  $\mu$ M for F06. The  $K_d$  value for compound 4 was measured as 100 nM. The compound also showed a significant inhibition of the removal of UV-induced damages in colorectal cancer cell lines and sensitized colorectal cancer cell lines to cyclophosphamide-induced cytotoxicity, indicating inhibition of NER and ICL repair. Compound 4 is therefore a promising drug candidate for combination cancer therapy. Currently, the compound is under the patenting process, and future steps will include evaluation of the ADMET profile and validation in animal models. At the same time, two new generations of derivatives are being synthesized and tested with the objective to further optimize the profile of compound 4.

Chapter 6 concludes the work of my thesis with the hope that my findings will benefit the future of cancer treatment for patients in Canada and around the world.

## Bibliography

1. Vitalis, A.; Pappu, R. V Methods for Monte Carlo simulations of biomacromolecules. *Annu Rep Comput Chem* **2009**, *5*, 49–76, doi:10.1016/S1574-1400(09)00503-9.
2. Karplus, M. Molecular dynamics simulations of biomolecules. *Acc Chem Res* **2002**, *35*, 321–323.
3. Jorgensen, W. L.; Tirado-Rives, J. Monte Carlo vs molecular dynamics for conformational sampling. *J. Phys. Chem.* **1996**, *100*, 14508–14513.
4. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
5. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854, doi:10.1093/bioinformatics/btt055.
6. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
7. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–802, doi:10.1002/jcc.20289.
8. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197, doi:10.1021/ja00124a002.

9. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–25, doi:10.1002/prot.21123.
10. MacKerell, A. D.; Bashford, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M.; MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616, doi:10.1021/jp973084f.
11. Fujihashi, M.; Ishida, T.; Kuroda, S.; Kotra, L. P.; Pai, E. F.; Miki, K. Substrate Distortion Contributes to the Catalysis of Orotidine 5'-Monophosphate Decarboxylase. *J. Am. Chem. Soc.* **2013**, *135*, 17432–17443, doi:10.1021/ja408197k.
12. Tiwari, G.; Mohanty, D. An In Silico Analysis of the Binding Modes and Binding Affinities of Small Molecule Modulators of PDZ-Peptide Interactions. *PLoS One* **2013**, *8*, doi:ARTN e7134010.1371/journal.pone.0071340.
13. Tuszynski, J. A.; Wenger, C.; Friesen, D. E.; Preto, J. An Overview of Sub-Cellular Mechanisms Involved in the Action of TTFIELDS. *Int. J. Environ. Res. Public Health* **2016**, *13*, 1128, doi:UNSP 112810.3390/ijerph13111128.
14. Zahn, D. Molecular dynamics simulation of ionic conductors: perspectives and limitations. *J. Mol. Model.* **2011**, *17*, 1531–1535, doi:10.1007/s00894-010-0877-3.
15. Wood, N. T.; Fadda, E.; Davis, R.; Grant, O. C.; Martin, J. C.; Woods, R. J.; Travers, S. A. The Influence of N-Linked Glycans on the Molecular Dynamics of the HIV-1 gp120 V3 Loop. *PLoS One* **2013**, *8*, doi:ARTN e8030110.1371/journal.pone.0080301.

16. Chaudhuri, A.; Sarkar, I.; Chakraborty, S. Comparative analysis of binding sites of human meprins with hydroxamic acid derivative by molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1969–1978, doi:10.1080/07391102.2013.848173.
17. Leonis, G.; Steinbrecher, T.; Papadopoulos, M. G. A contribution to the drug resistance mechanism of Darunavir, Amprenavir, Indinavir, and Saquinavir complexes with HIV-1 protease due to flap mutation I50V: A systematic MM–PBSA and thermodynamic integration study. *J Chem Inf Model* **2013**, *53*, 2141–2153.
18. Yoda, T.; Sugita, Y.; Okamoto, Y. Hydrophobic Core Formation and Dehydration in Protein Folding Studied by Generalized-Ensemble Simulations. *Biophys. J.* **2010**, *99*, 1637–1644, doi:10.1016/j.bpj.2010.06.045.
19. Yoda, T.; Sugita, Y.; Okamoto, Y. Salt effects on hydrophobic-core formation in folding of a helical miniprotein studied by molecular dynamics simulations. *Proteins-Structure Funct. Bioinforma.* **2014**, *82*, 933–943, doi:10.1002/prot.24467.
20. Hu, S. Y.; Yu, H. J.; Liu, Y. J.; Xue, T.; Zhang, H. B. Insight into the binding model of new antagonists of kappa receptor using docking and molecular dynamics simulation. *J. Mol. Model.* **2013**, *19*, 3087–3094, doi:10.1007/s00894-013-1839-3.
21. Hu, W.; Deng, S. W.; Huang, J. Y.; Lu, Y. M.; Le, X. Y.; Zheng, W. X. Intercalative interaction of asymmetric copper(II) complex with DNA: Experimental, molecular docking, molecular dynamics and TDDFT studies. *J. Inorg. Biochem.* **2013**, *127*, 90–98, doi:10.1016/j.jinorgbio.2013.07.034.
22. Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* **1994**, *7*, 385–391.
23. Guitierrez-de-Teran, H.; Aqvist, J. Linear Interaction Energy: Method and Applications in Drug Design. *Comput. Drug Discov. Des.* **2012**, *819*, 305–323, doi:10.1007/978-1-61779-465-0\_20.
24. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E.

- Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897, doi:10.1021/ar000033j.
25. Chen, F.; Liu, H.; Sun, H. Y.; Pan, P. C.; Li, Y. Y.; Li, D.; Hou, T. J. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Phys. Chem. Chem. Phys.* **2016**, *18*, 22129–22139, doi:10.1039/c6cp03670h.
  26. Sanders, J. M.; Wampole, M. E.; Thakur, M. L.; Wickstrom, E. Molecular Determinants of Epidermal Growth Factor Binding: A Molecular Dynamics Study. *PLoS One* **2013**, *8*, doi:ARTN e5413610.1371/journal.pone.0054136.
  27. Zhu, T.; Lee, H.; Lei, H.; Jones, C.; Patel, K.; Johnson, M. E.; Hevener, K. E. Fragment-Based Drug Discovery Using a Multidomain, Parallel MD-MM/PBSA Screening Protocol. *J Chem Inf Model* **2013**, *53*, 560–572, doi:10.1021/ci300502h.
  28. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461, doi:10.1517/17460441.2015.1032936.
  29. Sun, H. Y.; Li, Y. Y.; Tian, S.; Xu, L.; Hou, T. J. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16719–16729, doi:10.1039/c4cp01388c.
  30. Chakravarty, S.; Varadarajan, R. Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* **1999**, *7*, 723–32, doi:http://dx.doi.org/10.1016/S0969-2126(99)80097-5.
  31. Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. **2005**, doi:10.1021/JM049131R.
  32. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.

33. Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111–4119, doi:10.1021/jm048957q.
34. Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. Protein clefts in molecular recognition and function. *Protein Sci.* **1996**, *5*, 2438–52, doi:10.1002/pro.5560051206.
35. Zheng, X.; Gan, L.; Wang, E.; Wang, J. Pocket-Based Drug Design: Exploring Pocket Space. *AAPS J.* **2013**, *15*, 228–241, doi:10.1208/s12248-012-9426-6.
36. Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.* **1995**, *13*, 323–330, doi:10.1016/0263-7855(95)00073-9.
37. Glaser, F.; Morris, R. J.; Najmanovich, R. J.; Laskowski, R. A.; Thornton, J. M. A method for localizing ligand binding pockets in protein structures. *Proteins Struct. Funct. Bioinforma.* **2005**, *62*, 479–488, doi:10.1002/prot.20769.
38. Li, B.; Turuvekere, S.; Agrawal, M.; La, D.; Ramani, K.; Kihara, D. Characterization of local geometry of protein surfaces with the visibility criterion. *Proteins Struct. Funct. Bioinforma.* **2008**, *71*, 670–683, doi:10.1002/prot.21732.
39. Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* **1997**, *15*, 359–63, 389, doi:10.1016/S1093-3263(98)00002-3.
40. Huang, B.; Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **2006**, *6*, 19, doi:10.1186/1472-6807-6-19.
41. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–57.
42. Laurie, A. T. R.; Jackson, R. M. Q-SiteFinder: an energy-based method for the prediction

- of protein-ligand binding sites. *Bioinformatics* **2005**, *21*, 1908–16, doi:10.1093/bioinformatics/bti315.
43. Ghersi, D.; Sanchez, R. EasyMIFs and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics* **2009**, *25*, 3185–3186, doi:10.1093/bioinformatics/btp562.
  44. Harris, R.; Olson, A. J.; Goodsell, D. S. Automated prediction of ligand-binding sites in proteins. *Proteins Struct. Funct. Bioinforma.* **2007**, *70*, 1506–1517, doi:10.1002/prot.21645.
  45. Hetényi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002**, *11*, 1729–37, doi:10.1110/ps.0202302.
  46. Hetényi, C.; van der Spoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **2006**, *580*, 1447–1450, doi:10.1016/j.febslet.2006.01.074.
  47. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–52, doi:10.1002/jcc.20634.
  48. Thomas, P. D.; Dill, K. A. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* **1996**, *257*, 457–69, doi:10.1006/jmbi.1996.0175.
  49. Zheng, Z.; Merz, K. M.; Jr. Development of the knowledge-based and empirical combined scoring algorithm (KECSA) to score protein-ligand interactions. *J. Chem. Inf. Model.* **2013**, *53*, 1073–83, doi:10.1021/ci300619x.
  50. Feher, M. Consensus scoring for protein–ligand interactions. *Drug Discov. Today* **2006**, *11*, 421–428, doi:10.1016/j.drudis.2006.03.009.
  51. Amaro, R. E.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided. Mol. Des.* **2008**, *22*, 693–705, doi:10.1007/s10822-007-9159-2.

52. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288, doi:10.1016/0022-2836(82)90153-X.
53. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156, doi:10.1002/jcc.23905.
54. Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986–2000, doi:10.1021/ci1001982.
55. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662, doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
56. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–91, doi:10.1002/jcc.21256.
57. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–61, doi:10.1002/jcc.21334.
58. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–49, doi:10.1021/jm0306430.
59. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609–23, doi:10.1002/prot.10465.
60. Mooij, W. T. M.; Verdonk, M. L. General and targeted statistical potentials for protein-

- ligand interactions. *Proteins Struct. Funct. Bioinforma.* **2005**, *61*, 272–287, doi:10.1002/prot.20588.
61. Nissink, J. W. M.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins Struct. Funct. Bioinforma.* **2002**, *49*, 457–471, doi:10.1002/prot.10232.
  62. Avorn, J. The \$2.6 Billion Pill — Methodologic and Policy Considerations. *N. Engl. J. Med.* **2015**, *372*, 1877–1879, doi:10.1056/NEJMp1500848.
  63. Dickson, M.; Gagnon, J. P. The Cost of New Drug Discovery and Development. *Discov. Med.* **2009**, *4*, 172–179.
  64. Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* **2013**, *20*, 2839–60, doi:10.2174/09298673113209990001.
  65. Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217–241, doi:10.1016/S1574-1400(08)00012-1.
  66. Xie, X.-Q. Exploiting PubChem for Virtual Screening. *Expert Opin. Drug Discov.* **2010**, *5*, 1205–1220, doi:10.1517/17460441.2010.524924.
  67. Sterling, T.; Irwin, J. J. ZINC 15--Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–37, doi:10.1021/acs.jcim.5b00559.
  68. The National Cancer Institute (NCI) NCI Compound Sets Available online: <https://wiki.nci.nih.gov/display/NCIDTPdata/Compound+Sets> (accessed on Jul 20, 2015).
  69. Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discov.* **2010**, *5*, 597–607, doi:10.1517/17460441.2010.484460.
  70. Johnson, M. A.; Maggiora, G. M. *Concepts and applications of molecular similarity*; Wiley, 1990; ISBN 0471621757.

71. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **2006**, *11*, 1046–53, doi:10.1016/j.drudis.2006.10.005.
72. Dalke, A. The FPS fingerprint format and chemfp toolkit. *J. Cheminform.* **2013**, *5*, P36, doi:10.1186/1758-2946-5-S1-P36.
73. Horvath, D. Pharmacophore-based virtual screening. *Methods Mol. Biol.* **2011**, *672*, 261–98, doi:10.1007/978-1-60761-839-3\_11.
74. Melville, J.; Burke, E.; Hirst, J. Machine Learning in Virtual Screening. *Comb. Chem. High Throughput Screen.* **2009**, *12*, 332–343, doi:10.2174/138620709788167980.
75. Harris, C. J.; Hill, R. D.; Sheppard, D. W.; Slater, M. J.; Stouten, P. F. W. The design and application of target-focused compound libraries. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 521–31, doi:10.2174/138620711795767802.
76. Ripphausen, P.; Nisius, B.; Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug Discov. Today* **2011**, *16*, 372–376, doi:10.1016/j.drudis.2011.02.011.
77. Chiang, Y.-K.; Kuo, C.-C.; Wu, Y.-S.; Chen, C.-T.; Coumar, M. S.; Wu, J.-S.; Hsieh, H.-P.; Chang, C.-Y.; Jseeng, H.-Y.; Wu, M.-H.; Leou, J.-S.; Song, J.-S.; Chang, J.-Y.; Lyu, P.-C.; Chao, Y.-S.; Wu, S.-Y. Generation of Ligand-Based Pharmacophore Model and Virtual Screening for Identification of Novel Tubulin Inhibitors with Potent Anticancer Activity. *J. Med. Chem.* **2009**, *52*, 4221–4233, doi:10.1021/jm801649y.
78. Vuorinen, A.; Engeli, R.; Meyer, A.; Bachmann, F.; Griesser, U. J.; Schuster, D.; Odermatt, A. Ligand-Based Pharmacophore Modeling and Virtual Screening for the Discovery of Novel 17 $\beta$ -Hydroxysteroid Dehydrogenase 2 Inhibitors. *J. Med. Chem.* **2014**, *57*, 5995–6007, doi:10.1021/jm5004914.
79. Kurczyk, A.; Warszycki, D.; Musiol, R.; Kafel, R.; Bojarski, A. J.; Polanski, J. Ligand-Based Virtual Screening in a Search for Novel Anti-HIV-1 Chemotypes. *J. Chem. Inf. Model.* **2015**, *55*, 2168–2177, doi:10.1021/acs.jcim.5b00295.
80. Barakat, K. H.; Jordheim, L. P.; Perez-Pineiro, R.; Wishart, D.; Dumontet, C.; Tuszynski,

- J. A. Virtual screening and biological evaluation of inhibitors targeting the XPA-ERCC1 interaction. *PLoS One* **2012**, *7*, e51329, doi:10.1371/journal.pone.0051329.
81. Thompson, D. C.; Humblet, C.; Joseph-McCarthy, D. Investigation of MM-PBSA rescoring of docking poses. *J. Chem. Inf. Model.* **2008**, *48*, 1081–91, doi:10.1021/ci700470c.
  82. Gentile, F.; Tuszynski, J. A.; Barakat, K. H. New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy. *J. Mol. Graph. Model.* **2016**, *65*, 71–82, doi:10.1016/j.jmngm.2016.02.010.
  83. Gentile, F.; A. Tuszynski, J.; H. Barakat, K. Modelling DNA Repair Pathways: Recent Advances and Future Directions. *Curr. Pharm. Des.* **2016**, *22*, 3527–3546, doi:10.2174/1381612822666160420141435.
  84. Jordheim, L. P.; Barakat, K. H.; Heinrich-Balard, L.; Matera, E.-L.; Cros-Perrial, E.; Bouledrak, K.; El Sabeh, R.; Perez-Pineiro, R.; Wishart, D. S.; Cohen, R.; Tuszynski, J.; Dumontet, C. Small molecule inhibitors of ERCC1-XPF protein-protein interaction synergize alkylating agents in cancer cells. *Mol. Pharmacol.* **2013**, *84*, 12–24, doi:10.1124/mol.112.082347.
  85. McNeil, E. M.; Astell, K. R.; Ritchie, A.-M.; Shave, S.; Houston, D. R.; Bakrania, P.; Jones, H. M.; Khurana, P.; Wallace, C.; Chapman, T.; Wear, M. A.; Walkinshaw, M. D.; Saxty, B.; Melton, D. W. Inhibition of the ERCC1-XPF structure-specific endonuclease to overcome cancer chemoresistance. *DNA Repair (Amst).* **2015**, *31*, 19–28, doi:10.1016/j.dnarep.2015.04.002.
  86. Nunes, R. R.; Costa, M. D. S.; Santos, B. D. R.; Fonseca, A. L. da; Ferreira, L. S.; Chagas, R. C. R.; Silva, A. M. da; Varotti, F. de P.; Taranto, A. G. Successful application of virtual screening and molecular dynamics simulations against antimalarial molecular targets. *Mem. Inst. Oswaldo Cruz* **2016**, *111*, 721–730, doi:10.1590/0074-02760160207.
  87. Bajusz, D.; Ferenczy, G.; Keserű, G. Structure-Based Virtual Screening Approaches in Kinase-Directed Drug Discovery. *Curr. Top. Med. Chem.* **2017**, *17*, 2235–2259,

doi:10.2174/1568026617666170224121313.

88. Gu, W.-G.; Zhang, X.; Yuan, J.-F. Anti-HIV drug development through computational methods. *AAPS J.* **2014**, *16*, 674–80, doi:10.1208/s12248-014-9604-9.
89. Li, N.; Ainsworth, R. I.; Ding, B.; Hou, T.; Wang, W. Using Hierarchical Virtual Screening to Combat Drug Resistance of the HIV-1 Protease. *J. Chem. Inf. Model.* **2015**, *55*, 1400–1412, doi:10.1021/acs.jcim.5b00056.
90. Iyama, T.; Wilson, D. M. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst).* **2013**, *12*, 620–36, doi:10.1016/j.dnarep.2013.04.015.
91. Martin, L. P.; Hamilton, T. C.; Schilder, R. J. Platinum resistance: the role of DNA repair pathways. *Clin. Cancer Res.* **2008**, *14*, 1291–5, doi:10.1158/1078-0432.CCR-07-2238.
92. Albertella, M. R.; Lau, A.; O'Connor, M. J. The overexpression of specialized DNA polymerases in cancer. *DNA Repair (Amst).* **2005**, *4*, 583–93, doi:10.1016/j.dnarep.2005.01.005.
93. Zaremba, T.; Ketzer, P.; Cole, M.; Coulthard, S.; Plummer, E. R.; Curtin, N. J. Poly(ADP-ribose) polymerase-1 polymorphisms, expression and activity in selected human tumour cell lines. *Br. J. Cancer* **2009**, *101*, 256–262, doi:10.1038/sj.bjc.6605166.
94. Abbotts, R.; Madhusudan, S. Human AP endonuclease 1 (APE1): from mechanistic insights to druggable target in cancer. *Cancer Treat. Rev.* **2010**, *36*, 425–35, doi:10.1016/j.ctrv.2009.12.006.
95. Lord, C. J.; Ashworth, A. The DNA damage response and cancer therapy. *Nature* **2012**, *481*, 287–94, doi:10.1038/nature10760.
96. Jackson, S. P.; Bartek, J. The DNA-damage response in human biology and disease. *Nature* **2009**, *461*, 1071–8, doi:10.1038/nature08467.
97. Kraemer, K. H.; Patronas, N. J.; Schiffmann, R.; Brooks, B. P.; Tamura, D.; DiGiovanna, J. J. Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: A complex

- genotype–phenotype relationship. *Neuroscience* **2007**, *145*, 1388–1396, doi:10.1016/j.neuroscience.2006.12.020.
98. Hoeijmakers, J. H. J. DNA damage, aging, and cancer. *N. Engl. J. Med.* **2009**, *361*, 1475–85, doi:10.1056/NEJMra0804615.
  99. Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, *8*, 193–204, doi:10.1038/nrc2342.
  100. Barakat, K.; Gajewski, M.; Tuszynski, J. A. DNA repair inhibitors: the next major step to improve cancer therapy. *Curr. Top. Med. Chem.* **2012**, *12*, 1376–90.
  101. Kunz, B. A. Mutagenesis and deoxyribonucleotide pool imbalance. *Mutat. Res.* **1988**, *200*, 133–47.
  102. Holliday, R. Aspects of DNA repair and nucleotide pool imbalance. *Basic Life Sci.* **1985**, *31*, 453–60.
  103. Miura, G. Cancer therapeutics: Cleaning the nucleotide pool. *Nat. Chem. Biol.* **2014**, *10*, 409–409, doi:10.1038/nchembio.1543.
  104. Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; Saleh, A.; Göktürk, C.; Baranczewski, P.; Svensson, R.; Berntsson, R. P.-A.; Gustafsson, R.; Strömberg, K.; Sanjiv, K.; Jacques-Cordonnier, M.-C.; Desroses, M.; Gustavsson, A.-L.; Olofsson, R.; Johansson, F.; Homan, E. J.; Loseva, O.; Bräutigam, L.; Johansson, L.; Höglund, A.; Hagenkort, A.; Pham, T.; Altun, M.; Gaugaz, F. Z.; Vikingsson, S.; Evers, B.; Henriksson, M.; Vallin, K. S. A.; Wallner, O. A.; Hammarström, L. G. J.; Wiita, E.; Almlöf, I.; Kalderén, C.; Axelsson, H.; Djureinovic, T.; Puigvert, J. C.; Häggblad, M.; Jeppsson, F.; Martens, U.; Lundin, C.; Lundgren, B.; Granelli, I.; Jensen, A. J.; Artursson, P.; Nilsson, J. A.; Stenmark, P.; Scobie, M.; Berglund, U. W.; Helleday, T. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508*, 215–21, doi:10.1038/nature13181.

105. Huber, K. V. M.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J. M.; Stukalov, A.; Jemth, A.-S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; Pham, T.; Berglund, U. W.; Colinge, J.; Bennett, K. L.; Loizou, J. I.; Helleday, T.; Knapp, S.; Superti-Furga, G. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **2014**, *508*, 222–227, doi:10.1038/nature13194.
106. Zomaya, A. Y. *Parallel Computing for Bioinformatics and Computational Biology: Models, Enabling Technologies, and Case Studies*; John Wiley & Sons, 2006; ISBN 0471756490.
107. Sloot, P. M. A.; Hoekstra, A. G. Multi-scale modelling in computational biomedicine. *Brief. Bioinform.* **2010**, *11*, 142–52, doi:10.1093/bib/bbp038.
108. Guo, G.-S.; Zhang, F.-M.; Gao, R.-J.; Delsite, R.; Feng, Z.-H.; Powell, S. N. DNA repair and synthetic lethality. *Int. J. Oral Sci.* **2011**, *3*, 176–9, doi:10.4248/IJOS11064.
109. Ashworth, A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J. Clin. Oncol.* **2008**, *26*, 3785–90, doi:10.1200/JCO.2008.16.0812.
110. Martin, S. A.; Lord, C. J.; Ashworth, A. DNA repair deficiency as a therapeutic target in cancer. *Curr. Opin. Genet. Dev.* **2008**, *18*, 80–6, doi:10.1016/j.gde.2008.01.016.
111. Sultana, R.; Abdel-Fatah, T.; Abbotts, R.; Hawkes, C.; Albarakati, N.; Seedhouse, C.; Ball, G.; Chan, S.; Rakha, E. A.; Ellis, I. O.; Madhusudan, S. Targeting XRCC1 deficiency in breast cancer for personalized therapy. *Cancer Res.* **2013**, *73*, 1621–34, doi:10.1158/0008-5472.CAN-12-2929.
112. Fousteri, M.; Mullenders, L. H. F. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* **2008**, *18*, 73–84, doi:10.1038/cr.2008.6.
113. Gillet, L. C. J.; Schärer, O. D. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev.* **2006**, *106*, 253–76, doi:10.1021/cr040483f.

114. Masutani, C.; Sugasawa, K.; Yanagisawa, J.; Sonoyama, T.; Ui, M.; Enomoto, T.; Takio, K.; Tanaka, K.; van der Spek, P. J.; Bootsma, D. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J.* **1994**, *13*, 1831–43.
115. Compe, E.; Egly, J.-M. TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 343–54, doi:10.1038/nrm3350.
116. Oksenysh, V.; Coin, F. The long unwinding road: XPB and XPD helicases in damaged DNA opening. *Cell Cycle* **2010**, *9*, 90–6.
117. Fuss, J. O.; Tainer, J. A. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair (Amst)*. **2011**, *10*, 697–713, doi:10.1016/j.dnarep.2011.04.028.
118. Li, L.; Lu, X.; Peterson, C. A.; Legerski, R. J. An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. *Mol. Cell. Biol.* **1995**, *15*, 5396–402.
119. Patrick, S. M.; Turchi, J. J. Xeroderma pigmentosum complementation group A protein (XPA) modulates RPA-DNA interactions via enhanced complex stability and inhibition of strand separation activity. *J. Biol. Chem.* **2002**, *277*, 16096–101, doi:10.1074/jbc.M200816200.
120. Orelli, B.; McClendon, T. B.; Tsodikov, O. V; Ellenberger, T.; Niedernhofer, L. J.; Schärer, O. D. The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *J. Biol. Chem.* **2010**, *285*, 3705–12, doi:10.1074/jbc.M109.067538.
121. Tsodikov, O. V; Enzlin, J. H.; Schärer, O. D.; Ellenberger, T. Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11236–41, doi:10.1073/pnas.0504341102.
122. Shell, S. M.; Chazin, W. J. XPF-ERCC1: on the bubble. *Structure* **2012**, *20*, 566–8,

doi:10.1016/j.str.2012.03.004.

123. O'Donovan, A.; Davies, A. A.; Moggs, J. G.; West, S. C.; Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **1994**, *371*, 432–5, doi:10.1038/371432a0.
124. Zotter, A.; Luijsterburg, M. S.; Warmerdam, D. O.; Ibrahim, S.; Nigg, A.; van Cappellen, W. A.; Hoeijmakers, J. H. J.; van Driel, R.; Vermeulen, W.; Houtsmuller, A. B. Recruitment of the nucleotide excision repair endonuclease XPG to sites of UV-induced dna damage depends on functional TFIIH. *Mol. Cell. Biol.* **2006**, *26*, 8868–79, doi:10.1128/MCB.00695-06.
125. Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
126. Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **2014**, *42*, D199-205, doi:10.1093/nar/gkt1076.
127. Rabik, C. A.; Dolan, M. E. Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treat. Rev.* **2007**, *33*, 9–23, doi:10.1016/j.ctrv.2006.09.006.
128. Köberle, B.; Masters, J. R. W.; Hartley, J. A.; Wood, R. D. Defective repair of cisplatin-induced DNA damage caused by reduced XPA protein in testicular germ cell tumours. *Curr. Biol.* **1999**, *9*, 273–278, doi:10.1016/S0960-9822(99)80118-3.
129. Mendoza, J.; Martínez, J.; Hernández, C.; Pérez-Montiel, D.; Castro, C.; Fabián-Morales, E.; Santibáñez, M.; González-Barrios, R.; Díaz-Chávez, J.; Andonegui, M. A.; Reynoso, N.; Oñate, L. F.; Jiménez, M. A.; Núñez, M.; Dyer, R.; Herrera, L. A. Association between ERCC1 and XPA expression and polymorphisms and the response to cisplatin in testicular germ cell tumours. *Br. J. Cancer* **2013**, *109*, 68–75, doi:10.1038/bjc.2013.303.
130. Neher, T. M.; Bodenmiller, D.; Fitch, R. W.; Jalal, S. I.; Turchi, J. J. Novel Irreversible Small Molecule Inhibitors of Replication Protein A Display Single-Agent Activity and

- Synergize with Cisplatin. *Mol. Cancer Ther.* **2011**, *10*, 1796–1806, doi:10.1158/1535-7163.MCT-11-0303.
131. Alekseev, S.; Ayadi, M.; Brino, L.; Egly, J.-M.; Larsen, A. K.; Coin, F. A Small Molecule Screen Identifies an Inhibitor of DNA Repair Inducing the Degradation of TFIIH and the Chemosensitization of Tumor Cells to Platinum. *Chem. Biol.* **2014**, *21*, 398–407, doi:10.1016/j.chembiol.2013.12.014.
  132. Kelley, M. R.; Logsdon, D.; Fishel, M. L. Targeting DNA repair pathways for cancer treatment: what's new? *Future Oncol.* **2014**, *10*, 1215–37, doi:10.2217/fon.14.60.
  133. Brosey, C. A.; Yan, C.; Tsutakawa, S. E.; Heller, W. T.; Rambo, R. P.; Tainer, J. A.; Ivanov, I.; Chazin, W. J. A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **2013**, *41*, 2313–27, doi:10.1093/nar/gks1332.
  134. Nuss, J. E.; Sweeney, D. J.; Alter, G. M. Reactivity-based analysis of domain structures in native replication protein A. *Biochemistry* **2006**, *45*, 9804–18, doi:10.1021/bi052397v.
  135. Pretto, D. I.; Tsutakawa, S.; Brosey, C. A.; Castillo, A.; Chagot, M.-E.; Smith, J. A.; Tainer, J. A.; Chazin, W. J. Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. *Biochemistry* **2010**, *49*, 2880–9, doi:10.1021/bi9019934.
  136. Brosey, C. A.; Soss, S. E.; Brooks, S.; Yan, C.; Ivanov, I.; Dorai, K.; Chazin, W. J. Functional dynamics in replication protein A DNA binding and protein recruitment domains. *Structure* **2015**, *23*, 1028–38, doi:10.1016/j.str.2015.04.008.
  137. Liu, Y.; Vaithiyalingam, S.; Shi, Q.; Chazin, W. J.; Zinkel, S. S. BID binds to replication protein A and stimulates ATR function following replicative stress. *Mol. Cell. Biol.* **2011**, *31*, 4298–309, doi:10.1128/MCB.05737-11.
  138. Carra, C.; Cucinotta, F. A. Accurate prediction of the binding free energy and analysis of the mechanism of the interaction of replication protein A (RPA) with ssDNA. *J. Mol. Model.* **2012**, *18*, 2761–83, doi:10.1007/s00894-011-1288-9.

139. Carra, C.; Saha, J.; Cucinotta, F. A. Theoretical prediction of the binding free energy for mutants of replication protein A. *J. Mol. Model.* **2012**, *18*, 3035–49, doi:10.1007/s00894-011-1313-z.
140. Pavani, R. S.; Fernandes, C.; Perez, A. M.; Vasconcelos, E. J. R.; Siqueira-Neto, J. L.; Fontes, M. R.; Cano, M. I. N. RPA-1 from *Leishmania amazonensis* (LaRPA-1) structurally differs from other eukaryote RPA-1 and interacts with telomeric DNA via its N-terminal OB-fold domain. *FEBS Lett.* **2014**, *588*, 4740–8, doi:10.1016/j.febslet.2014.11.005.
141. Mishra, G.; Levy, Y. Molecular determinants of the interactions between proteins and ssDNA. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 5033–8, doi:10.1073/pnas.1416355112.
142. Ali, S. I.; Shin, J.-S.; Bae, S.-H.; Kim, B.; Choi, B.-S. Replication protein A 32 interacts through a similar binding interface with TIPIN, XPA, and UNG2. *Int. J. Biochem. Cell Biol.* **2010**, *42*, 1210–5, doi:10.1016/j.biocel.2010.04.011.
143. Ikegami, T.; Kuraoka, I.; Saijo, M.; Kodo, N.; Kyogoku, Y.; Morikawa, K.; Tanaka, K.; Shirakawa, M. Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nat. Struct. Biol.* **1998**, *5*, 701–6, doi:10.1038/1400.
144. Shell, S. M.; Li, Z.; Shkriabai, N.; Kvaratskhelia, M.; Brosey, C.; Serrano, M. A.; Chazin, W. J.; Musich, P. R.; Zou, Y. Checkpoint kinase ATR promotes nucleotide excision repair of UV-induced DNA damage via physical interaction with xeroderma pigmentosum group A. *J. Biol. Chem.* **2009**, *284*, 24213–22, doi:10.1074/jbc.M109.000745.
145. Barakat, K. H.; Torin Huzil, J.; Luchko, T.; Jordheim, L.; Dumontet, C.; Tuszynski, J. Characterization of an inhibitory dynamic pharmacophore for the ERCC1-XPA interaction using a combined molecular dynamics and virtual screening approach. *J. Mol. Graph. Model.* **2009**, *28*, 113–30, doi:10.1016/j.jmglm.2009.04.009.
146. Fadda, E. Conformational determinants for the recruitment of ERCC1 by XPA in the nucleotide excision repair (NER) Pathway: structure and dynamics of the XPA binding motif. *Biophys. J.* **2013**, *104*, 2503–11, doi:10.1016/j.bpj.2013.04.023.

147. Fadda, E. The role of conformational selection in the molecular recognition of the wild type and mutants XPA67-80 peptides by ERCC1. *Proteins* **2015**, *83*, 1341–1351, doi:10.1002/prot.24825.
148. Nagasundaram, N.; Priya Doss, C. G. Exploration of structural stability in deleterious nsSNPs of the XPA gene: A molecular dynamics approach. *J. Carcinog.* **2011**, *10*, 26, doi:10.4103/1477-3163.90441.
149. Newman, M.; Murray-Rust, J.; Lally, J.; Rudolf, J.; Fadden, A.; Knowles, P. P.; White, M. F.; McDonald, N. Q. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J.* **2005**, *24*, 895–905, doi:10.1038/sj.emboj.7600581.
150. Choi, Y.-J.; Ryu, K.-S.; Ko, Y.-M.; Chae, Y.-K.; Pelton, J. G.; Wemmer, D. E.; Choi, B.-S. Biophysical characterization of the interaction domains and mapping of the contact residues in the XPF-ERCC1 complex. *J. Biol. Chem.* **2005**, *280*, 28644–52, doi:10.1074/jbc.M501083200.
151. Priya Doss, C. G.; Nagasundaram, N. Molecular docking and molecular dynamics study on the effect of ERCC1 deleterious polymorphisms in ERCC1-XPF heterodimer. *Appl. Biochem. Biotechnol.* **2014**, *172*, 1265–81, doi:10.1007/s12010-013-0592-5.
152. Shuck, S. C.; Turchi, J. J. Targeted inhibition of Replication Protein A reveals cytotoxic activity, synergy with chemotherapeutic DNA-damaging agents, and insight into cellular function. *Cancer Res.* **2010**, *70*, 3189–98, doi:10.1158/0008-5472.CAN-09-3422.
153. Glanzer, J. G.; Liu, S.; Oakley, G. G. Small molecule inhibitor of the RPA70 N-terminal protein interaction domain discovered using in silico and in vitro methods. *Bioorg. Med. Chem.* **2011**, *19*, 2589–2595, doi:10.1016/j.bmc.2011.03.012.
154. Glanzer, J. G.; Carnes, K. A.; Soto, P.; Liu, S.; Parkhurst, L. J.; Oakley, G. G. A small molecule directly inhibits the p53 transactivation domain from binding to replication protein A. *Nucleic Acids Res.* **2013**, *41*, 2047–59, doi:10.1093/nar/gks1291.
155. Neher, T. M.; Shuck, S. C.; Liu, J.-Y.; Zhang, J.-T.; Turchi, J. J. Identification of novel

- small molecule inhibitors of the XPA protein using in silico based screening. *ACS Chem. Biol.* **2010**, *5*, 953–65, doi:10.1021/cb1000444.
156. Wiseman, H.; Kaur, H.; Halliwell, B. DNA damage and cancer: measurement and mechanism. *Cancer Lett.* **1995**, *93*, 113–20, doi:10.1016/0304-3835(95)03792-U.
  157. Wiseman, H.; Halliwell, B. Damage to DNA by reactive oxygen and nitrogen species: role in inflammatory disease and progression to cancer. *Biochem. J.* **1996**, *313* (Pt 1), 17–29.
  158. Cadet, J.; Sage, E.; Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. *Mutat. Res. Mol. Mech. Mutagen.* **2005**, *571*, 3–17, doi:10.1016/j.mrfmmm.2004.09.012.
  159. Xue, W.; Warshawsky, D. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol. Appl. Pharmacol.* **2005**, *206*, 73–93, doi:10.1016/j.taap.2004.11.006.
  160. Nospikel, T. DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cell. Mol. Life Sci.* **2009**, *66*, 994–1009, doi:10.1007/s00018-009-8737-y.
  161. Ferry, K. V; Hamilton, T. C.; Johnson, S. W. Increased nucleotide excision repair in cisplatin-resistant ovarian cancer cells. *Biochem. Pharmacol.* **2000**, *60*, 1305–1313, doi:10.1016/S0006-2952(00)00441-X.
  162. Rosell, R.; Taron, M.; Barnadas, A.; Scagliotti, G.; Sarries, C.; Roig, B. Nucleotide excision repair pathways involved in Cisplatin resistance in non-small-cell lung cancer. *Cancer Control* **2003**, *10*, 297–305.
  163. Selvakumaran, M.; Pisarcik, D. A.; Bao, R.; Yeung, A. T.; Hamilton, T. C. Enhanced Cisplatin Cytotoxicity by Disturbing the Nucleotide Excision Repair Pathway in Ovarian Cancer Cell Lines. *Cancer Res.* **2003**, *63*, 1311–1316.
  164. Tsodikov, O. V; Ivanov, D.; Orelli, B.; Staresincic, L.; Shoshani, I.; Oberman, R.; Schärer, O. D.; Wagner, G.; Ellenberger, T. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J.* **2007**, *26*, 4768–76, doi:10.1038/sj.emboj.7601894.

165. Welsh, C.; Day, R.; McGurk, C.; Masters, J. R. W.; Wood, R. D.; Köberle, B. Reduced levels of XPA, ERCC1 and XPF DNA repair proteins in testis tumor cell lines. *Int. J. Cancer* **2004**, *110*, 352–61, doi:10.1002/ijc.20134.
166. Weaver, D.; Crawford, E.; Warner, K.; Elkhairi, F.; Khuder, S.; Willey, J. ABCC5, ERCC2, XPA and XRCC1 transcript abundance levels correlate with cisplatin chemoresistance in non-small cell lung cancer cell lines. *Mol. Cancer* **2005**, *4*, 18, doi:10.1186/1476-4598-4-18.
167. Jiang, H.; Yang, L.-Y. Cell Cycle Checkpoint Abrogator UCN-01 Inhibits DNA Repair: Association with Attenuation of the Interaction of XPA and ERCC1 Nucleotide Excision Repair Proteins. *Cancer Res.* **1999**, *59*, 4529–4534.
168. Barakat, K.; Mane, J.; Friesen, D.; Tuszynski, J. Ensemble-based virtual screening reveals dual-inhibitors for the p53–MDM2/MDMX interactions. *J. Mol. Graph. Model.* **2010**, *28*, 555–568.
169. Barakat, K.; Tuszynski, J. Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta. *J. Mol. Graph. Model.* **2011**, *29*, 702–16, doi:10.1016/j.jmglm.2010.12.003.
170. Zhao, H.; Caflisch, A. Molecular dynamics in drug design. *Eur. J. Med. Chem.* **2015**, *91*, 4–14, doi:10.1016/j.ejmech.2014.08.004.
171. Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model* **2011**, *51*, 69–82, doi:10.1021/ci100275a.
172. Delaney, J. S. Assessing the ability of chemical similarity measures to discriminate between active and inactive compounds. *Mol. Divers.* **1996**, *1*, 217–22.
173. PubChem Substructure Fingerprint Available online: [ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem\\_fingerprints.txt](ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt) (accessed on Apr 8, 2015).

174. Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Enhancing the effectiveness of similarity-based virtual screening using nearest-neighbor information. *J. Med. Chem.* **2005**, *48*, 7049–54, doi:10.1021/jm050316n.
175. Gardiner, E. J.; Gillet, V. J.; Haranczyk, M.; Hert, J.; Holliday, J. D.; Malim, N.; Patel, Y.; Willett, P. Turbo similarity searching: Effect of fingerprint and dataset on virtual-screening performance. *Stat. Anal. Data Min.* **2009**, *2*, 103–114, doi:10.1002/sam.10037.
176. Simulations Plus Inc. ADMET Predictor 7.1 Available online: <http://www.simulations-plus.com/Products.aspx?grpID=1&cID=10&pID=13> (accessed on Jul 20, 2015).
177. Schrödinger LigPrep Available online: <http://www.schrodinger.com/productpage/14/10/> (accessed on Jan 24, 2014).
178. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today* **2010**, *15*, 444–50, doi:10.1016/j.drudis.2010.03.013.
179. Sun, H. Pharmacophore-Based Virtual Screening. *Curr. Med. Chem.* **2008**, *15*, 1018–1024, doi:10.2174/092986708784049630.
180. Böhm, H.-J.; Flohr, A.; Stahl, M. Scaffold hopping. *Drug Discov. Today. Technol.* **2004**, *1*, 217–24, doi:10.1016/j.ddtec.2004.10.009.
181. Chemical Computing Group Inc. Molecular Operating Environment (MOE). *Chem. Comput. Gr. Inc., Canada* 2017.
182. Lin, A. Overview of Pharmacophore Applications in MOE Available online: <https://www.chemcomp.com/journal/ph4.htm> (accessed on Jul 20, 2015).
183. Gilson, M. K.; Sharp, K. A.; Honig, B. H. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **1988**, *9*, 327–335, doi:10.1002/jcc.540090407.
184. Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput. Aided.*

- Mol. Des.* **2000**, *14*, 251–64.
185. Philippopoulos, M.; Lim, C. Exploring the dynamic information content of a protein NMR structure: comparison of a molecular dynamics simulation with the NMR and X-ray structures of Escherichia coli ribonuclease HI. *Proteins* **1999**, *36*, 87–110.
186. Ishima, R.; Torchia, D. A. Protein dynamics from NMR. *Nat. Struct. Biol.* **2000**, *7*, 740–3, doi:10.1038/78963.
187. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–41, doi:10.1093/nar/gks375.
188. AmberTools13 Reference Manual Available online: <http://ambermd.org/doc12/AmberTools13.pdf>.
189. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, B.; Wu, X.; Kollman, P. A. Amber 14 Available online: <http://ambermd.org/> (accessed on Oct 31, 2015).
190. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555, doi:10.1021/ct200909j.
191. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888, doi:10.1021/ct400314y.
192. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–74, doi:10.1002/jcc.20035.

193. Wang, J.; Wang, W.; Kollmann, P.; Case, D. Antechamber, An Accessory Software Package For Molecular Mechanical Calculation. *J. Comput. Chem.* **2005**, *25*, 1157–1174.
194. Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–41, doi:10.1002/jcc.10128.
195. Xu, L.; Sun, H.; Li, Y.; Wang, J.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J. Phys. Chem. B* **2013**, *117*, 8408–21, doi:10.1021/jp404160y.
196. Su, P.-C.; Tsai, C.-C.; Mehboob, S.; Hevener, K. E.; Johnson, M. E. Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of *F. tularensis* enoyl-ACP reductase (FabI). *J. Comput. Chem.* **2015**, *36*, 1859–73, doi:10.1002/jcc.24011.
197. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321, doi:10.1021/ct300418h.
198. Tan, C.; Yang, L.; Luo, R. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. *J. Phys. Chem. B* **2006**, *110*, 18680–7, doi:10.1021/jp063479b.
199. Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–606, doi:10.1021/jm0608210.
200. Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins* **2011**, *79*, 1940–51, doi:10.1002/prot.23018.
201. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, doi:10.1016/0263-7855(96)00018-5.

202. O'Grady, S.; Finn, S. P.; Cuffe, S.; Richard, D. J.; O'Byrne, K. J.; Barr, M. P. The role of DNA repair pathways in cisplatin resistant lung cancer. *Cancer Treat. Rev.* **2014**, *40*, 1161–1170, doi:10.1016/j.ctrv.2014.10.003.
203. McNeil, E. M.; Melton, D. W. DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res.* **2012**, *40*, 9990–10004, doi:10.1093/nar/gks818.
204. Chapman, T. M.; Wallace, C.; Gillen, K. J.; Bakrania, P.; Khurana, P.; Coombs, P. J.; Fox, S.; Bureau, E. A.; Brownlees, J.; Melton, D. W.; Saxty, B. N-Hydroxyimides and hydroxypyrimidinones as inhibitors of the DNA repair complex ERCC1-XPF. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4104–8, doi:10.1016/j.bmcl.2015.08.024.
205. Chapman, T. M.; Gillen, K. J.; Wallace, C.; Lee, M. T.; Bakrania, P.; Khurana, P.; Coombs, P. J.; Stennett, L.; Fox, S.; Bureau, E. A.; Brownlees, J.; Melton, D. W.; Saxty, B. Catechols and 3-hydroxypyridones as inhibitors of the DNA repair complex ERCC1-XPF. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4097–103, doi:10.1016/j.bmcl.2015.08.031.
206. Arora, S.; Heyza, J.; Zhang, H.; Kalman-Maltese, V.; Tillison, K.; Floyd, A. M.; Chalfin, E. M.; Bepler, G.; Patrick, S. M. Identification of small molecule inhibitors of ERCC1-XPF that inhibit DNA repair and potentiate cisplatin efficacy in cancer cells. *Oncotarget* **2016**, *7*, 75104–75117, doi:10.18632/oncotarget.12072.
207. Nishino, T.; Komori, K.; Ishino, Y.; Morikawa, K. X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure* **2003**, *11*, 445–57, doi:10.1016/S0969-2126(03)00046-7.
208. Ciccia, A.; McDonald, N.; West, S. C. Structural and Functional Relationships of the XPF/MUS81 Family of Proteins. *Annu. Rev. Biochem.* **2008**, *77*, 259–287, doi:10.1146/annurev.biochem.77.070306.102408.
209. Henikoff, S.; Henikoff, J. G. Performance evaluation of amino acid substitution matrices. *Proteins* **1993**, *17*, 49–61, doi:10.1002/prot.340170108.

210. Pearson, W. R. Comparison of methods for searching protein sequence databases. *Protein Sci.* **1995**, *4*, 1145–60, doi:10.1002/pro.5560040613.
211. The Statistics of Sequence Similarity Scores - National Center for Biotechnology Information Available online: <http://www.ncbi.nlm.nih.gov/BLAST/tutorial> (accessed on Dec 8, 2015).
212. Enzlin, J. H.; Schärer, O. D. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J.* **2002**, *21*, 2045–53, doi:10.1093/emboj/21.8.2045.
213. Yang, W.; Lee, J. Y.; Nowotny, M. Making and breaking nucleic acids: two-Mg<sup>2+</sup>-ion catalysis and substrate specificity. *Mol. Cell* **2006**, *22*, 5–13, doi:10.1016/j.molcel.2006.03.013.
214. Gwon, G. H.; Jo, A.; Baek, K.; Jin, K. S.; Fu, Y.; Lee, J.-B.; Kim, Y.; Cho, Y. Crystal structures of the structure-selective nuclease Mus81-Eme1 bound to flap DNA substrates. *EMBO J.* **2014**, *33*, 1061–72, doi:10.1002/emboj.201487820.
215. UniProt Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204-12, doi:10.1093/nar/gku989.
216. Kelly, K. 3D Bioinformatics and Comparative Modeling in MOE Available online: <https://www.chemcomp.com/journal/bio1999.htm> (accessed on Dec 8, 2015).
217. Kelly, K. Exhaustive and Iterative Clustering of the Protein Databank Available online: <https://www.chemcomp.com/journal/families.htm> (accessed on Mar 7, 2018).
218. Henikoff, S.; Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10915–9.
219. Gonnet, G. H.; Cohen, M. A.; Benner, S. A. Exhaustive matching of the entire protein sequence database. *Science* **1992**, *256*, 1443–5.
220. Dayhoff, M. O.; Schwartz, R.; Orcutt, B. C. A model of evolutionary change in proteins.

In *Atlas of protein sequence and structure*; 1978; pp. 345–352.

221. Labute, P. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* **2008**, *29*, 1693–8, doi:10.1002/jcc.20933.
222. Labute, P. Protonate 3D: Assignment of Macromolecular Protonation State and Geometry Available online: <http://www.ccl.net/ccl/documents/proton/> (accessed on Mar 31, 2014).
223. AmberTools 12 Reference Manual Available online: <http://ambermd.org/doc12/AmberTools12.pdf>.
224. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, doi:10.1021/acs.jctc.5b00255.
225. Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748, doi:10.1021/ct400146w.
226. Li, P.; Song, L. F.; Merz, K. M. Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* **2015**, *11*, 1645–1657, doi:10.1021/ct500918t.
227. Ryckaert, J.-P. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341, doi:10.1016/0021-9991(77)90098-5.
228. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334, doi:10.1021/ct700119m.
229. Nussinov, R.; Wolfson, H. J. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. *Proc. Natl. Acad. Sci.* **1991**, *88*, 10495–10499, doi:10.1073/pnas.88.23.10495.

230. Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in docking success rates due to dataset preparation. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 775–86, doi:10.1007/s10822-012-9570-1.
231. Sijbers, A. M.; De Laat, W. L.; Ariza, R. R.; Biggerstaff, M.; Wei, Y. F.; Moggs, J. G.; Carter, K. C.; Shell, B. K.; Evans, E.; De Jong, M. C.; Rademakers, S.; De Rooij, J.; Jaspers, N. G. J.; Hoeijmakers, J. H. J.; Wood, R. D. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **1996**, *86*, 811–822, doi:10.1016/S0092-8674(00)80155-5.
232. Roger, A. S.; Wolfram, S.; Graham, C. W.; Tom, E.; Richard, D. W.; Errol, C. F. *DNA Repair and Mutagenesis, Second Edition*; American Society of Microbiology, 2006; ISBN 9781555813192.
233. Klein Douwel, D.; Hoogenboom, W. S.; Boonen, R. A.; Knipscheer, P. Recruitment and positioning determine the specific role of the XPF-ERCC1 endonuclease in interstrand crosslink repair. *EMBO J.* **2017**, *36*, 2034–2046, doi:10.15252/embj.201695223.
234. Yagi, T.; Katsuya, A.; Koyano, A.; Takebe, H. Sensitivity of group F xeroderma pigmentosum cells to UV and mitomycin C relative to levels of XPF and ERCC1 overexpression. *Mutagenesis* **1998**, *13*, 595–599, doi:10.1093/mutage/13.6.595.
235. McHugh, P. J.; Spanswick, V. J.; Hartley, J. A. Repair of DNA interstrand crosslinks: Molecular mechanisms and clinical relevance. *Lancet Oncol.* **2001**, *2*, 483–490, doi:10.1016/S1470-2045(01)00454-5.
236. Ahmad, A.; Robinson, A. R.; Duensing, A.; van Drunen, E.; Beverloo, H. B.; Weisberg, D. B.; Hasty, P.; Hoeijmakers, J. H. J.; Niedernhofer, L. J. ERCC1-XPF Endonuclease Facilitates DNA Double-Strand Break Repair. *Mol. Cell. Biol.* **2008**, *28*, 5082–5092, doi:10.1128/MCB.00293-08.
237. de Laat, W. L.; Appeldoorn, E.; Jaspers, N. G.; Hoeijmakers, J. H. DNA structural elements required for ERCC1-XPF endonuclease activity. *J. Biol. Chem.* **1998**, *273*, 7835–42, doi:10.1074/JBC.273.14.7835.

238. Gaillard, P. H.; Wood, R. D. Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res.* **2001**, *29*, 872–9.
239. Ahmad, A.; Enzlin, J. H.; Bhagwat, N. R.; Wijgers, N.; Raams, A.; Appeldoorn, E.; Theil, A. F.; J Hoeijmakers, J. H.; Vermeulen, W.; J Jaspers, N. G.; Schärer, O. D.; Niedernhofer, L. J. Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet.* **2010**, *6*, e1000871, doi:10.1371/journal.pgen.1000871.
240. de Laat, W. L.; Sijbers, A. M.; Odijk, H.; Jaspers, N. G.; Hoeijmakers, J. H. Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res.* **1998**, *26*, 4146–52.
241. Tripsianes, K.; Folkers, G.; AB, E.; Das, D.; Odijk, H.; Jaspers, N. G. J.; Hoeijmakers, J. H. J.; Kaptein, R.; Boelens, R. The Structure of the Human ERCC1/XPF Interaction Domains Reveals a Complementary Role for the Two Proteins in Nucleotide Excision Repair. *Structure* **2005**, *13*, 1849–1858, doi:10.1016/j.str.2005.08.014.
242. Yang, L.; Ritchie, A.-M.; Melton, D. W. Disruption of DNA repair in cancer cells by ubiquitination of a destabilising dimerization domain of nucleotide excision repair protein ERCC1. *Oncotarget* **2017**, *8*, 55246–55264, doi:10.18632/oncotarget.19422.
243. Wildman, S. A.; Crippen, G. M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873, doi:10.1021/ci9903071.
244. Kassem, S.; Ahmed, M.; El-Sheikh, S.; Barakat, K. H. Entropy in bimolecular simulations: A comprehensive review of atomic fluctuations-based methods. *J. Mol. Graph. Model.* **2015**, *62*, 105–117, doi:10.1016/j.jmglm.2015.09.010.
245. Bowles, M.; Lally, J.; Fadden, A. J.; Mouilleron, S.; Hammonds, T.; McDonald, N. Q. Fluorescence-based incision assay for human XPF-ERCC1 activity identifies important elements of DNA junction recognition. *Nucleic Acids Res.* **2012**, *40*, e101, doi:10.1093/nar/gks284.
246. Mirzayans, R.; Pollock, S.; Scott, A.; Gao, C. Q.; Murray, D. Metabolic labeling of human

- cells with tritiated nucleosides results in activation of the ATM-dependent p53 signaling pathway and acceleration of DNA repair. *Oncogene* **2003**, *22*, 5562–5571, doi:10.1038/sj.onc.1206514.
247. Mani, R. S.; Karimi-Busheri, F.; Fanta, M.; Caldecott, K. W.; Cass, C. E.; Weinfeld, M. Biophysical characterization of human XRCC1 and its binding to damaged and undamaged DNA. *Biochemistry* **2004**, *43*, 16505–16514, doi:10.1021/bi048615m.
248. Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–94, doi:10.1002/prot.20033.
249. Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217–230, doi:10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A.
250. Takeda, K.; Akira, S. Toll-like receptors in innate immunity. *Int. Immunol.* **2005**, *17*, 1–14, doi:10.1093/intimm/dxh186.
251. Kawasaki, T.; Kawai, T. Toll-Like Receptor Signaling Pathways. *Front. Immunol.* **2014**, *5*, 461, doi:10.3389/fimmu.2014.00461.
252. Zhang, Z.; Ohto, U.; Shibata, T.; Krayukhina, E.; Taoka, M.; Yamauchi, Y.; Tanji, H.; Isobe, T.; Uchiyama, S.; Miyake, K.; Shimizu, T. Structural Analysis Reveals that Toll-like Receptor 7 Is a Dual Receptor for Guanosine and Single-Stranded RNA. *Immunity* **2016**, *45*, 737–748, doi:10.1016/j.immuni.2016.09.011.
253. Kaczanowska, S.; Joseph, A. M.; Davila, E. TLR agonists: our best frenemy in cancer immunotherapy. *J. Leukoc. Biol.* **2013**, *93*, 847–63, doi:10.1189/jlb.1012501.
254. Smits, E. L. J. M.; Ponsaerts, P.; Berneman, Z. N.; Van Tendeloo, V. F. I. The use of TLR7 and TLR8 ligands for the enhancement of cancer immunotherapy. *Oncologist* **2008**, *13*, 859–75, doi:10.1634/theoncologist.2008-0097.
255. Forsbach, A.; Müller, C.; Montino, C.; Kritzler, A.; Nguyen, T.; Weeratna, R.; Jurk, M.;

- Vollmer, J. Negative regulation of the type I interferon signaling pathway by synthetic Toll-like receptor 7 ligands. *J. Interferon Cytokine Res.* **2012**, *32*, 254–68, doi:10.1089/jir.2011.0091.
256. Lee, J.; Chuang, T.-H.; Redecke, V.; She, L.; Pitha, P. M.; Carson, D. A.; Raz, E.; Cottam, H. B. Molecular basis for the immunostimulatory activity of guanine nucleoside analogs: activation of Toll-like receptor 7. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6646–51, doi:10.1073/pnas.0631696100.
257. Schön, M. P.; Schön, M. TLR7 and TLR8 as targets in cancer therapy. *Oncogene* **2008**, *27*, 190–9, doi:10.1038/sj.onc.1210913.
258. Petricevic, B.; Wessner, B.; Sachet, M.; Vrbanec, D.; Spittler, A.; Bergmann, M. CL097, a TLR7/8 ligand, inhibits TLR-4--dependent activation of IRAK-M and BCL-3 expression. *Shock* **2009**, *32*, 484–90, doi:10.1097/SHK.0b013e3181a5ac8a.
259. Meyer, T.; Surber, C.; French, L. E.; Stockfleth, E. Resiquimod, a topical drug for viral skin lesions and skin cancer. *Expert Opin. Investig. Drugs* **2013**, *22*, 149–59, doi:10.1517/13543784.2013.749236.
260. Connolly, D. J.; O'Neill, L. A. J. New developments in Toll-like receptor targeted therapeutics. *Curr. Opin. Pharmacol.* **2012**, *12*, 510–8, doi:10.1016/j.coph.2012.06.002.
261. Koski, G. K.; Czerniecki, B. J. Combining Innate Immunity With Radiation Therapy for Cancer Treatment. *Clin. Cancer Res.* **2005**, *11*, 7–11.
262. Hotz, C.; Bourquin, C. Systemic cancer immunotherapy with Toll-like receptor 7 agonists: Timing is everything. *Oncoimmunology* **2012**, *1*, 227–228, doi:10.4161/onci.1.2.18169.
263. Kauffman, E. C.; Liu, H.; Schwartz, M. J.; Scherr, D. S. Toll-like receptor 7 agonist therapy with imidazoquinoline enhances cancer cell death and increases lymphocytic infiltration and proinflammatory cytokine production in established tumors of a renal cell carcinoma mouse model. *J. Oncol.* **2012**, *2012*, 103298, doi:10.1155/2012/103298.
264. Ochi, A.; Graffeo, C. S.; Zambirinis, C. P.; Rehman, A.; Hackman, M.; Fallon, N.; Barilla,

- R. M.; Henning, J. R.; Jamal, M.; Rao, R.; Greco, S.; Deutsch, M.; Medina-Zea, M. V.; Bin Saeed, U.; Ego-Osuala, M. O.; Hajdu, C.; Miller, G. Toll-like receptor 7 regulates pancreatic carcinogenesis in mice and humans. *J. Clin. Invest.* **2012**, *122*, 4118–29, doi:10.1172/JCI63606.
265. Eigenbrod, T.; Dalpke, A. H. TLR7 inhibition: A novel strategy for pancreatic cancer treatment? *JAK-STAT* **2013**, *2*, e23011, doi:10.4161/jkst.23011.
266. Lu, Y.; Chen, J.; Xiao, M.; Li, W.; Miller, D. D. An overview of tubulin inhibitors that interact with the colchicine binding site. *Pharm. Res.* **2012**, *29*, 2943–71, doi:10.1007/s11095-012-0828-z.
267. Dong, M.; Liu, F.; Zhou, H.; Zhai, S.; Yan, B. Novel Natural Product- and Privileged Scaffold-Based Tubulin Inhibitors Targeting the Colchicine Binding Site. *Molecules* **2016**, *21*, 1375, doi:10.3390/molecules21101375.
268. Johnson, L.; Goping, I. S.; Rieger, A.; Mane, J. Y.; Huzil, T.; Banerjee, A.; Luduena, R.; Hassani, B.; Winter, P.; Tuszynski, J. A. Novel Colchicine Derivatives and their Anti-cancer Activity. *Curr. Top. Med. Chem.* **2017**, *17*, 2538–2558, doi:10.2174/1568026617666170104143618.
269. Courbet, A.; Bec, N.; Constant, C.; Larroque, C.; Pugniere, M.; El Messaoudi, S.; Zghaib, Z.; Khier, S.; Deleuze-Masquefa, C.; Gattacceca, F. Imidazoquinoxaline anticancer derivatives and imiquimod interact with tubulin: Characterization of molecular microtubule inhibiting mechanisms in correlation with cytotoxicity. *PLoS One* **2017**, *12*, e0182022, doi:10.1371/journal.pone.0182022.
270. De Martino, G.; Edler, M. C.; La Regina, G.; Coluccia, A.; Barbera, M. C.; Barrow, D.; Nicholson, R. I.; Chiosis, G.; Brancale, A.; Hamel, E.; Artico, M.; Silvestri, R. New arylthioindoles: Potent inhibitors of tubulin polymerization. 2. Structure-activity relationships and molecular modeling studies. *J. Med. Chem.* **2006**, *49*, 947–954, doi:10.1021/jm050809s.
271. Gentile, F.; Deriu, M. A.; Licandro, G.; Prunotto, A.; Danani, A.; Tuszynski, J. A.;

- Gentile, F.; Deriu, M. A.; Licandro, G.; Prunotto, A.; Danani, A.; Tuszynski, J. A. Structure Based Modeling of Small Molecules Binding to the TLR7 by Atomistic Level Simulations. *Molecules* **2015**, *20*, 8316–8340, doi:10.3390/molecules20058316.
272. Maschalidi, S.; Hässler, S.; Blanc, F.; Sepulveda, F. E.; Tohme, M.; Chignard, M.; van Endert, P.; Si-Tahar, M.; Descamps, D.; Manoury, B. Asparagine endopeptidase controls anti-influenza virus immune responses through TLR7 activation. *PLoS Pathog.* **2012**, *8*, e1002841, doi:10.1371/journal.ppat.1002841.
273. Wei, T.; Gong, J.; Jamitzky, F.; Heckl, W. M.; Stark, R. W.; Rössle, S. C. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci.* **2009**, *18*, 1684–91, doi:10.1002/pro.186.
274. Wei, T.; Gong, J.; Rössle, S. C.; Jamitzky, F.; Heckl, W. M.; Stark, R. W. A leucine-rich repeat assembly approach for homology modeling of the human TLR5-10 and mouse TLR11-13 ectodomains. *J. Mol. Model.* **2011**, *17*, 27–36, doi:10.1007/s00894-010-0697-5.
275. TSeng, C.-Y.; Gajewski, M.; Danani, A.; Tuszynski, J. A. Homology and Molecular Dynamics Models of Toll-Like Receptor 7 Protein and Its Dimerization. *Chem. Biol. Drug Des.* **2014**, doi:10.1111/cbdd.12278.
276. Tanji, H.; Ohto, U.; Shibata, T.; Miyake, K.; Shimizu, T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* **2013**, *339*, 1426–9, doi:10.1126/science.1229159.
277. Kelly, K. Multiple Sequence and Structure Alignment in MOE Available online: <http://www.chemcomp.com/journal/align.htm> (accessed on Jun 10, 2014).
278. Hoffmann, R. An Extended Hückel Theory. I. Hydrocarbons. *J. Chem. Phys.* **1963**, *39*, 1397, doi:10.1063/1.1734456.
279. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926, doi:10.1063/1.445869.

280. Colovos, C.; Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511–9, doi:10.1002/pro.5560020916.
281. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291, doi:10.1107/S0021889892009944.
282. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752–1780, doi:10.1002/jcc.20292.
283. Onufriev, A. The generalized Born model: its foundation, applications, and limitations Available online: <http://people.cs.vt.edu/~onufriev/PUBLICATIONS/gbreview.pdf> (accessed on Jun 9, 2014).
284. Garnett, M. J.; Edelman, E. J.; Heidorn, S. J.; Greenman, C. D.; Dastur, A.; Lau, K. W.; Greninger, P.; Thompson, I. R.; Luo, X.; Soares, J.; Liu, Q.; Iorio, F.; Surdez, D.; Chen, L.; Milano, R. J.; Bignell, G. R.; Tam, A. T.; Davies, H.; Stevenson, J. A.; Barthorpe, S.; Lutz, S. R.; Kogera, F.; Lawrence, K.; McLaren-Douglas, A.; Mitropoulos, X.; Mironenko, T.; Thi, H.; Richardson, L.; Zhou, W.; Jewitt, F.; Zhang, T.; O'Brien, P.; Boisvert, J. L.; Price, S.; Hur, W.; Yang, W.; Deng, X.; Butler, A.; Choi, H. G.; Chang, J. W.; Baselga, J.; Stamenkovic, I.; Engelman, J. A.; Sharma, S. V.; Delattre, O.; Saez-Rodriguez, J.; Gray, N. S.; Settleman, J.; Futreal, P. A.; Haber, D. A.; Stratton, M. R.; Ramaswamy, S.; McDermott, U.; Benes, C. H. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012**, *483*, 570–575, doi:10.1038/nature11005.
285. Vogelstein, B.; Papadopoulos, N.; Velculescu, V. E.; Zhou, S.; Diaz, L. A.; Kinzler, K. W. Cancer genome landscapes. *Science* **2013**, *339*, 1546–58, doi:10.1126/science.1235122.
286. Lawrence, M. S.; Stojanov, P.; Mermel, C. H.; Robinson, J. T.; Garraway, L. A.; Golub, T. R.; Meyerson, M.; Gabriel, S. B.; Lander, E. S.; Getz, G. Discovery and saturation

analysis of cancer genes across 21 tumour types. *Nature* **2014**, *505*, 495–501, doi:10.1038/nature12912.

287. Uitdehaag, J. C. M.; de Roos, J. A. D. M.; van Doornmalen, A. M.; Prinsen, M. B. W.; de Man, J.; Tanizawa, Y.; Kawase, Y.; Yoshino, K.; Buijsman, R. C.; Zaman, G. J. R. Comparison of the Cancer Gene Targeting and Biochemical Selectivities of All Targeted Kinase Inhibitors Approved for Clinical Use. *PLoS One* **2014**, *9*, e92146, doi:10.1371/journal.pone.0092146.

## Appendix A - Supplementary Materials of Chapter 4 and 5

### A.1 Chapter 4<sup>6</sup>

Table A.1. Conformational entropy changes ( $T\Delta S$ ), calculated using normal mode analysis for a set of diverse hit structures. The ligands were selected to possess different features (for example the number of rings and flexible bonds). The unfavorable contributions oscillate between ~20 and 24 kcal/mol and they overlap within the standard deviations. The calculations were run on 100 snapshots extracted from the last 4 ns of 10 ns-long molecular dynamics simulations of each complex.

Compound	Entropy (kcal/mol)	Standard deviation (kcal/mol)
6210903	-24.20	3.26
7324126	-23.82	3.13
1696060	-22.25	3.93
1098945	-21.11	2.66
53684246	-22.91	2.83
6971912	-20.29	2.49
Ligand 1	-19.70	2.68
Ligand 63	-21.47	2.88
Ligand 80	-19.48	3.95

---

<sup>6</sup> This section has been published as Supplementary Data of Gentile, F.; Tuszynski, J. A.; Barakat, K. H. New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy. *J. Mol. Graph. Model.* 2016, *1*, 71–82.



conserved residues are reported in the Consensus line in normal font. Residue columns are colored based on their degree of conservation when compared with the XPF nuclease sequence, ranging from red (not conserved) to blue (fully conserved).

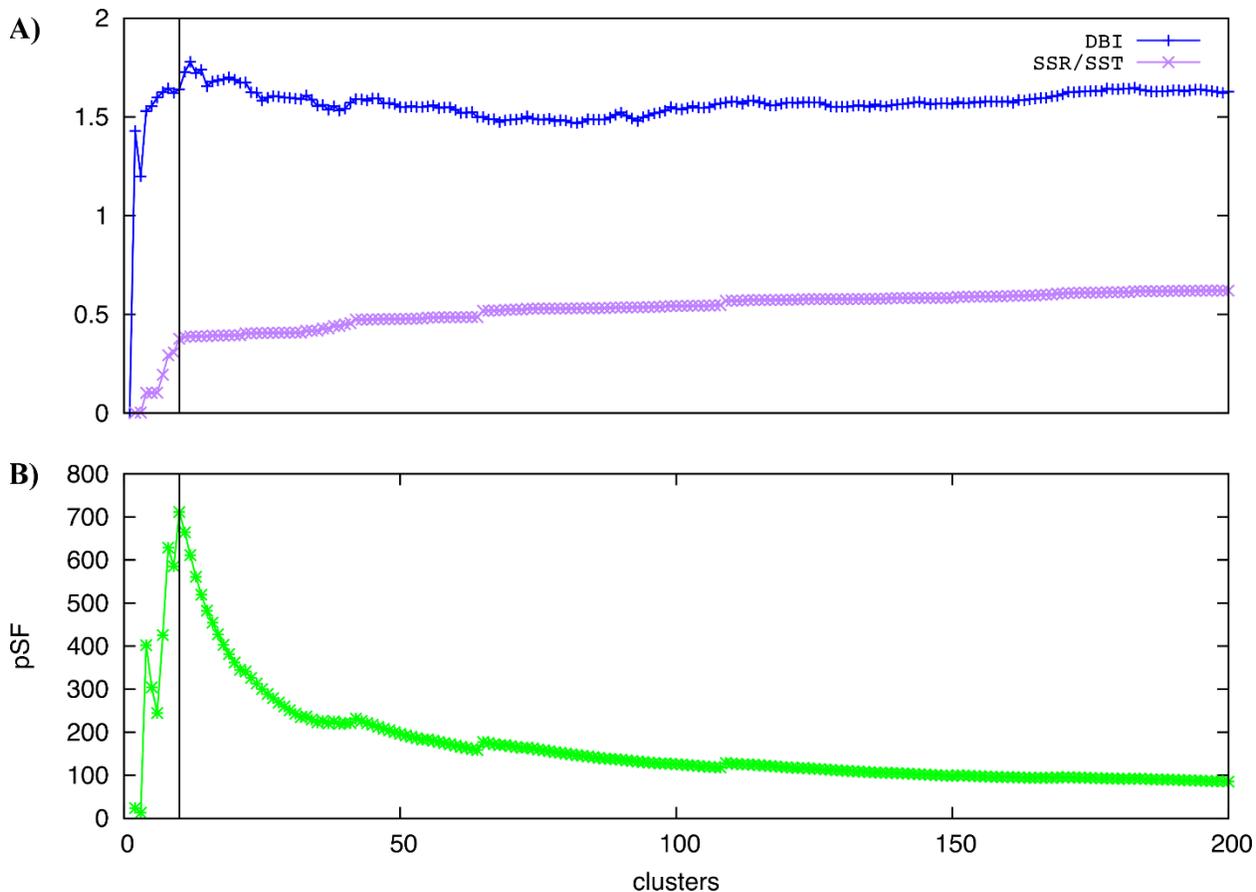
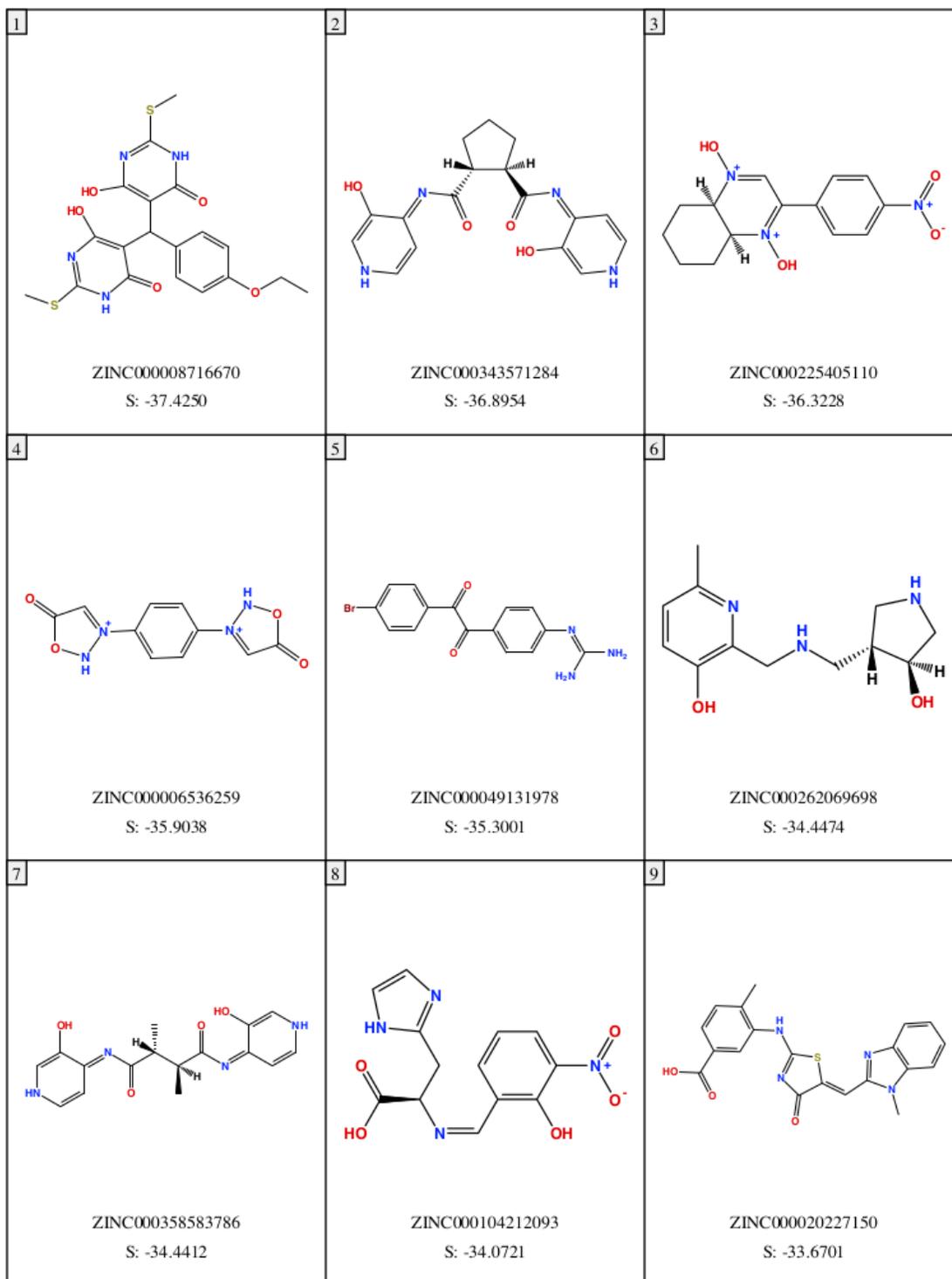
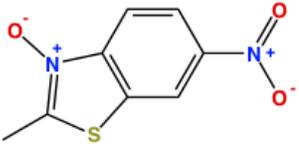
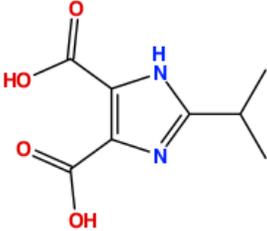
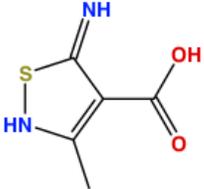
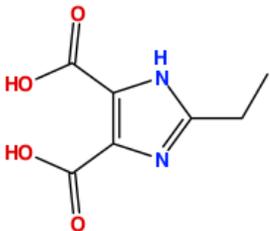
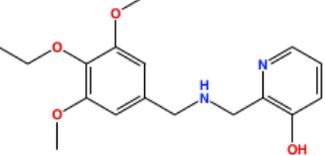
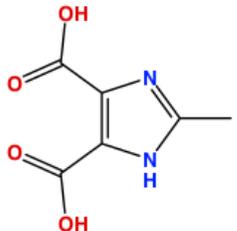
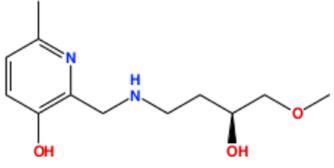
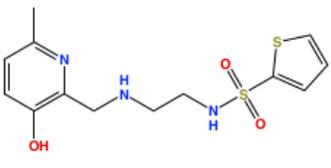
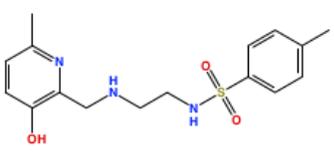


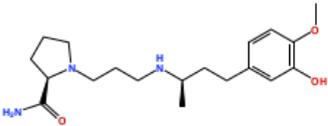
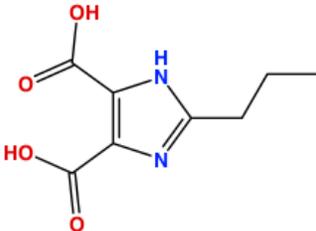
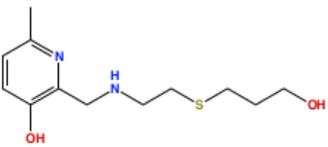
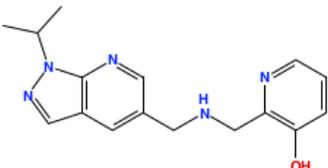
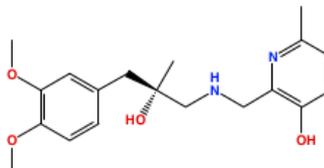
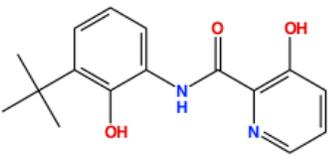
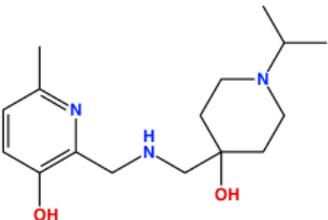
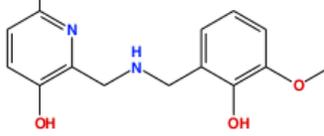
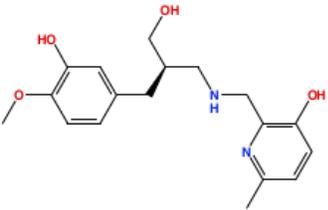
Figure A.3. A) DBI (blue) and SSR/SST ratio (purple) and B) pSF (green) trends obtained by varying the number of clusters of human XPF active site conformations. At 10 clusters (black vertical line), we observed a local minimum of the DBI, a kink in the SSR/SST curve and a maximum of the pSF, indicating optimal cluster counting.

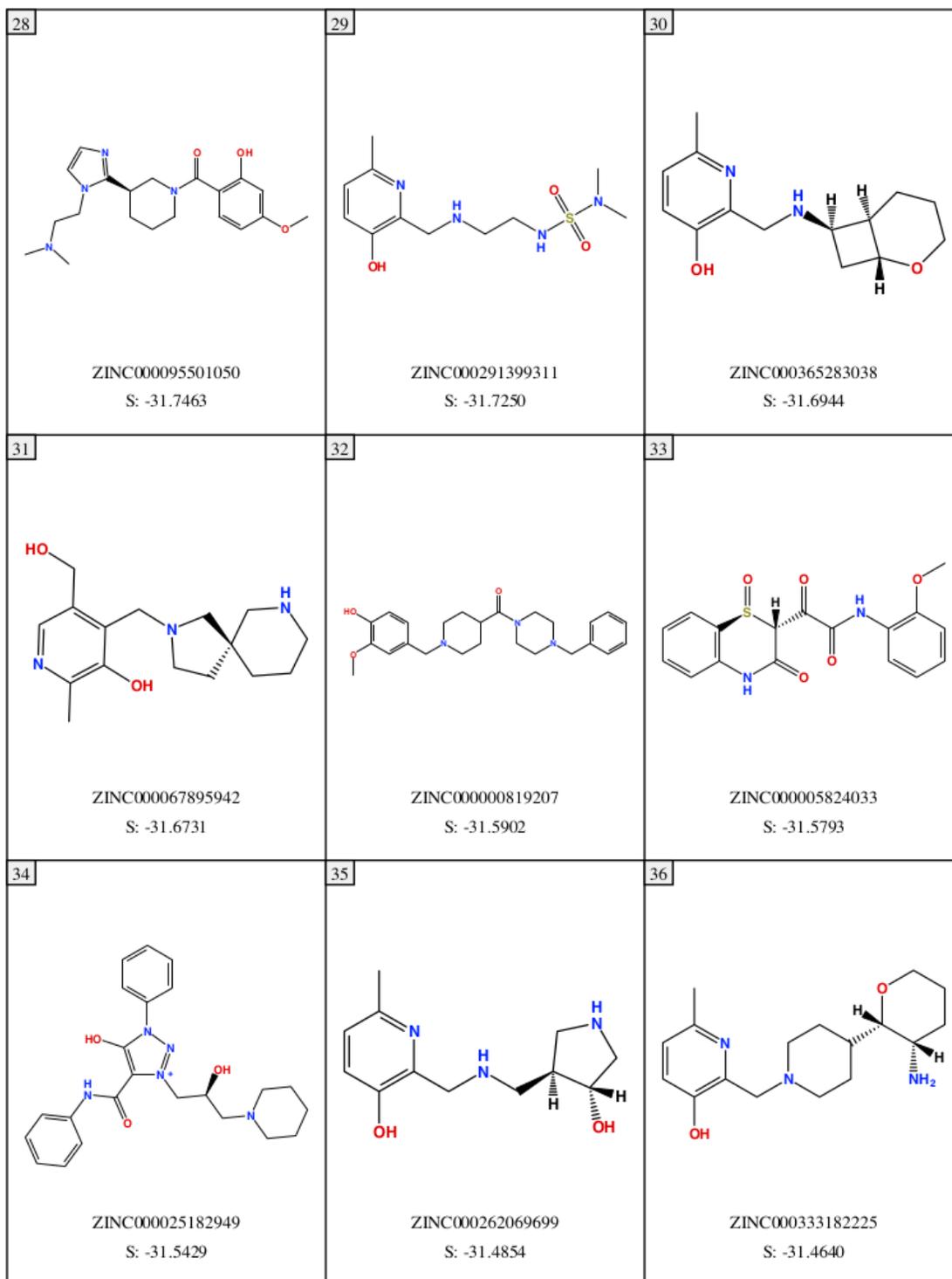
Table A.2. Composition of the 10 clusters resulting from clustering of the last 106 ns of MD trajectory. The ratio is calculated as number of frames in the cluster divided by the total number of frames (10601). Selected refers to the representative structure of a cluster being included in the molecular docking simulation as target or not.

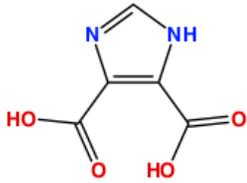
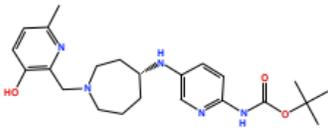
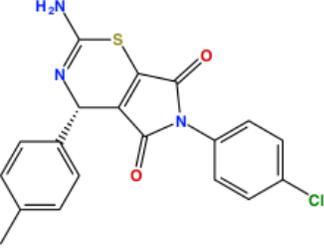
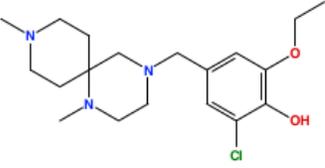
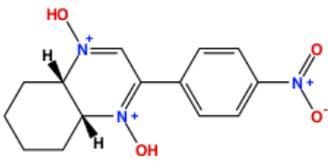
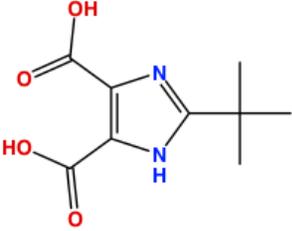
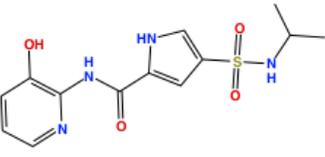
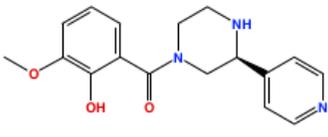
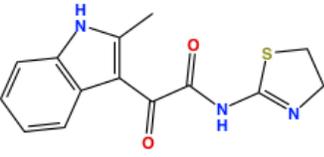
<b>Cluster</b>	<b>Number of frames</b>	<b>Ratio</b>	<b>Selected</b>
1	4160	0.392	yes
2	2125	0.200	yes
3	5	0.000	no
4	1	0.000	no
5	1370	0.129	yes
6	156	0.015	yes
7	15	0.001	no
8	1496	0.141	yes
9	1264	0.119	yes
10	9	0.001	no



<p>10</p>  <p>ZINC000045237915 S: -33.6330</p>	<p>11</p>  <p>ZINC000071774011 S: -33.5102</p>	<p>12</p>  <p>ZINC000032714687 S: -33.0514</p>
<p>13</p>  <p>ZINC000038550857 S: -32.8777</p>	<p>14</p>  <p>ZINC000347902752 S: -32.8049</p>	<p>15</p>  <p>ZINC000015886568 S: -32.7778</p>
<p>16</p>  <p>ZINC000231608350 S: -32.3570</p>	<p>17</p>  <p>ZINC000122283098 S: -32.3530</p>	<p>18</p>  <p>ZINC000103344220 S: -32.3287</p>

<p>19</p>  <p>ZINC000277276522 S: -32.2163</p>	<p>20</p>  <p>ZINC000021299679 S: -32.1401</p>	<p>21</p>  <p>ZINC000105643133 S: -32.1053</p>
<p>22</p>  <p>ZINC000348012855 S: -32.0948</p>	<p>23</p>  <p>ZINC000289920009 S: -31.9850</p>	<p>24</p>  <p>ZINC000361793164 S: -31.9330</p>
<p>25</p>  <p>ZINC000268407008 S: -31.9033</p>	<p>26</p>  <p>ZINC000265089061 S: -31.7699</p>	<p>27</p>  <p>ZINC000289505839 S: -31.7633</p>



<p>37</p>  <p>ZINC000015852947 S: -31.4342</p>	<p>38</p>  <p>ZINC000170618062 S: -31.3782</p>	<p>39</p>  <p>ZINC000002219244 S: -31.3611</p>
<p>40</p>  <p>ZINC000067886560 S: -31.3346</p>	<p>41</p>  <p>ZINC000225405065 S: -31.2085</p>	<p>42</p>  <p>ZINC000038550858 S: -31.2061</p>
<p>43</p>  <p>ZINC000341887368 S: -31.1700</p>	<p>44</p>  <p>ZINC000374391981 S: -31.1451</p>	<p>45</p>  <p>ZINC000030601464 S: -31.0931</p>

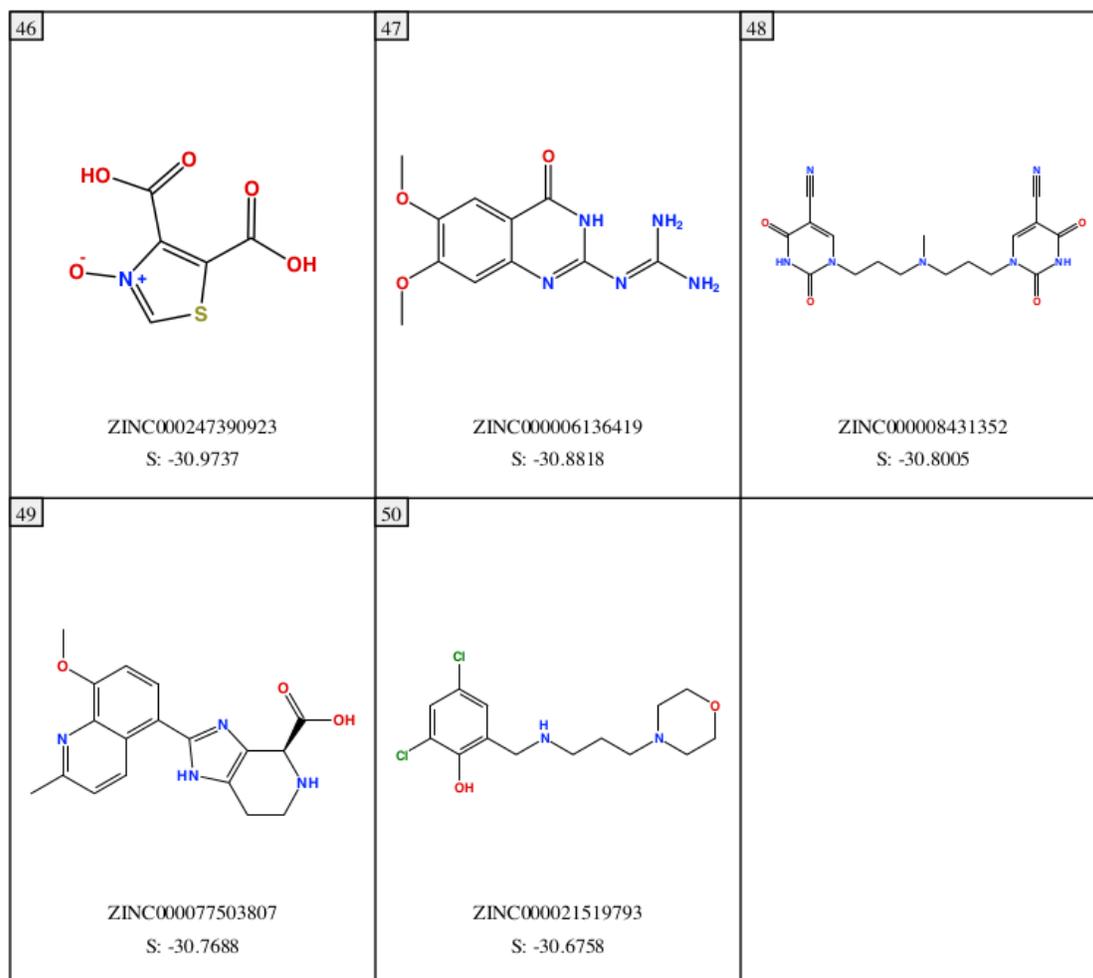


Figure A.4. Chemical structures, ZINC IDs and London dG scores (in kcal/mol) of the top fifty hits obtained by screening the ZINC15 compound database.

## Appendix B – Supplementary Materials to Chapter 6<sup>8</sup>

### B.1 Materials and Methods

#### B.1.1 Molecular Docking of F06

##### Target Preparation

We prepared the twenty NMR structures of the ERCC1-XPF complex (PDB entry 1Z00). We initially removed the ERCC1 terminus, as our previous study identified the best suitable binding pocket being on the XPF dimerization surface [84]. We then assigned the protonation states of the XPF residues the H<sup>++</sup> server [187] using pH 7, a salinity of 0.15 M, a protein and solvent dielectric constants of 10 and 80, respectively. In AmberTools14 tleap [189], Amber ff14SB force field parameters were assigned to the protein [224], whereas the Li, Song and Merz 's 12-6-4 parameters for mono and divalent ions in TIP3P water have been used for the ions [225,226]. The protein was solvated with an octahedral box of TIP3P explicit water molecules with 15 Å of buffer (minimal distance between any protein atom and the edge of the box). Na<sup>+</sup> and Cl<sup>-</sup> ions have been added to the system to neutralize it and reach a 0.15 M physiological ionic concentration. We then performed a two steps minimization in Amber pmemd.cuda [191] with the following procedure: 1000 steps of conjugate gradients method keeping harmonically restrained the whole protein (force constant of 500 kcal/mol/Å<sup>2</sup>) to relax ion and water positions; followed by 1000 steps of conjugate gradients method keeping restrained the backbone atoms with a force constant of 2 kcal/mol/Å<sup>2</sup>. A cutoff of 9 Å was used during this relaxation process.

---

<sup>8</sup> This appendix is under consideration for publication as Supplementary Materials to Elmenoufy A.H.; Gentile F.; Jay D.; Karimi-Busheri F.; Yang X.; Soueidan O.M.; Weilbeer C.; Mani R.S.; Barakat K.H.; Tuszynski J.A.; Weinfeld M.; West F.G. Novel ERCC1-XPF Inhibitors for Resistant Colorectal Cancer Treatment Based on Targeting DNA Repair in *Mol. Cancer Ther.*

## Docking Simulations

The 2D structure of F06 was downloaded from the PubChem repository [65] (entry 421105). Different protonation states were obtained using Epik, tautomers and up to three low energy ring conformations at pH 7.0  $\pm$  2.0 of the compound have been calculated in Schrödinger LigPrep, using the OPLS-2005 force field [177]. The docking box for each XPF structure was defined as 56X50X56 points spaced by 0.375 Å and centered in the geometric center of the residues Tyr833, Asn834, Pro837, Gln838, Met856, Lys860, Asn861 and Ile862 [84], as calculated in Visual Molecular Dynamics (VMD) [201]. Affinity, electrostatic and desolvation maps were obtained using Autogrid4 [56]. The Lamarckian genetic algorithm (LGA) of Autodock4 [55] was used for the docking simulations, with a population size of 300, 25000000 maximal energetic evaluations, 27000 maximal generations and one survival for each generation with the rates of mutation and crossover set to 0.02 and 0.8, respectively. For each simulation, 100 LGA runs were performed and the results were clustered based on the root-mean squared deviation (RMSD) metric using 2 Å as tolerance; just the simulations resulting in at least a cluster with more than 25 conformations were considered. The Autodock scoring function [47] was used to rank the results. Hydrogen bonds were detected in VMD with the following parameters: distance cutoff of 3.5 Å and angle cutoff of 120° (60° of deviation from linearity).

### *B.1.2 MD Simulations and MM/GBSA Rescoring of the Analogues*

#### Molecular dynamics simulations of top hit complexes

Antechamber [193] was used to assign the General Amber Force Field (GAFF) [192] parameters to the compounds. The ff14SB force field parameters were used for the protein. The same explicit solvation and ionic concentration setup introduced previously was employed. Molecular

dynamics (MD) simulations of the systems were performed on pmemd.cuda using the following procedure: 1) relaxation of ions and water molecule positions using 1000 steps of steepest descent and 1000 steps of conjugate gradients minimization, keeping all the complex atoms fixed through an harmonic restraint (force constant of 500 kcal/mol/Å<sup>2</sup>). 2) 2000 steps of steepest descent followed by 3000 steps of conjugate gradients methods without restraints. 3) Gradual heating of the system from 0 to 300 K in 100 ps using Langevin dynamics, keeping restrained the backbone atoms (2 kcal/mol/Å<sup>2</sup>) and using a time step of 0.5 fs and NVT conditions. 4) Gradual release of the restraint in four phases of 50 ps each at constant pressure (1 atm). The time step was set to 2 fs. 5) 2 ns of production simulation, recording the coordinates every 2 ps, similar to what was done in our previous work [82].

The enthalpic contribution to the change in free energy due to the binding of the ligands has been estimated using the MM/GBSA method using the script MMPBSA.py [197]; the change of the energy of a complex configuration is calculated as

$$\Delta G_{bind,solv} = \Delta G_{MM,vac} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,protein}) - T\Delta S \quad (B.1)$$

where  $\Delta G_{MM,vac}$  is calculated as the sum of electrostatic and van der Waals interactions occurring between the protein and the ligand.  $T\Delta S$  is the term modeling the change in conformational entropy due to the binding. The solvation terms are modeled as

$$\Delta G_{solv} = \Delta G_{solv,polar} + \Delta G_{solv,npolar} \quad (B.2)$$

where the polar contribution of the solvent is calculated solving the Generalized Born equation [28]. The ionic concentration was set to 0.15 M. igb flag was set to 5 [248] and mbondi2 radii were used. The hydrophobic contribution to the solvation free energy has been calculated as

$$\Delta G_{\text{solv, npolar}} = \gamma \cdot \text{SASA} \quad (\text{B.3})$$

where  $\gamma$  (surface tension) was  $0.005 \text{ kcal/mol/\AA}^2$  and A the solvent accessible surface area (SASA) calculated using the linear combinations of pairwise overlaps (LCPO) model method [249].

### B.1.3 Synthesis of F06-Based Analogues

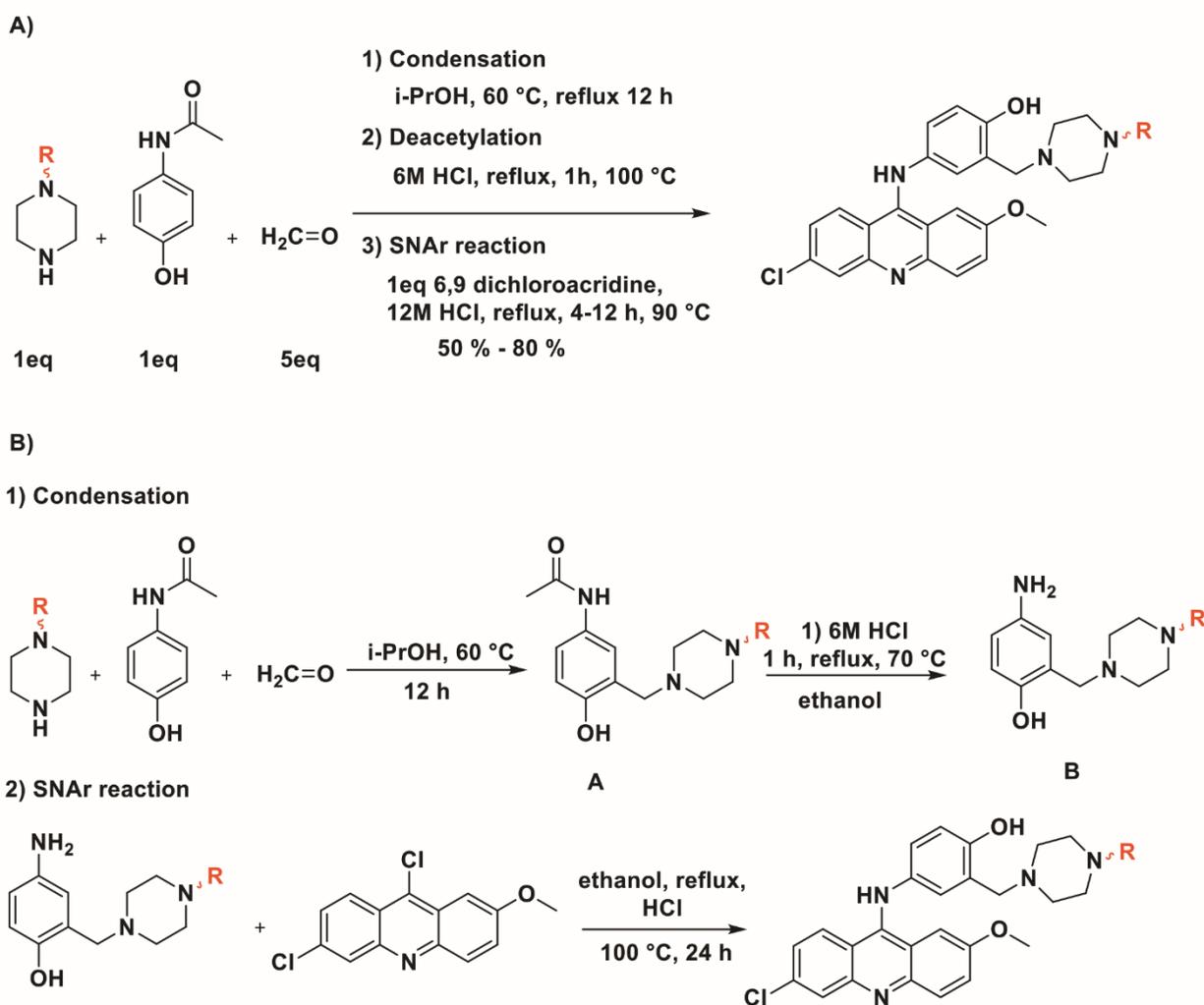


Figure B.1. A) One-pot sequential addition reaction for synthesis of compounds **1-8**. B) Main steps and intermediates included in the one pot sequential addition reaction.

### B.1.4 Cell Proliferation Assay

The Cell Titer 96® Aqueous One Solution Cell Proliferation Assay (MTS) (Promega, Madison, WI) was used to measure the proliferation of HCT-116 colorectal cancer cells. Briefly, 3,000 cells were seeded in 100 µl of culture medium. After 24 hours, the medium was aspirated and new culture medium without (control) and with inhibitory compounds were added to each well. After 72 hours 20 µl of MTS [3-(4, 5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium] solution was added to each well and plates were incubated for 2 hours at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> followed by reading the absorbance of treated and untreated cells at 490 nm wavelength with a microplate reader, and plotting the corrected absorbance at 490 nm versus concentration of drug as shown in Figure B.3 in Appendix B.

## **B.2 Results**

### *B.2.1 Virtual Screening of F06 Analogues*

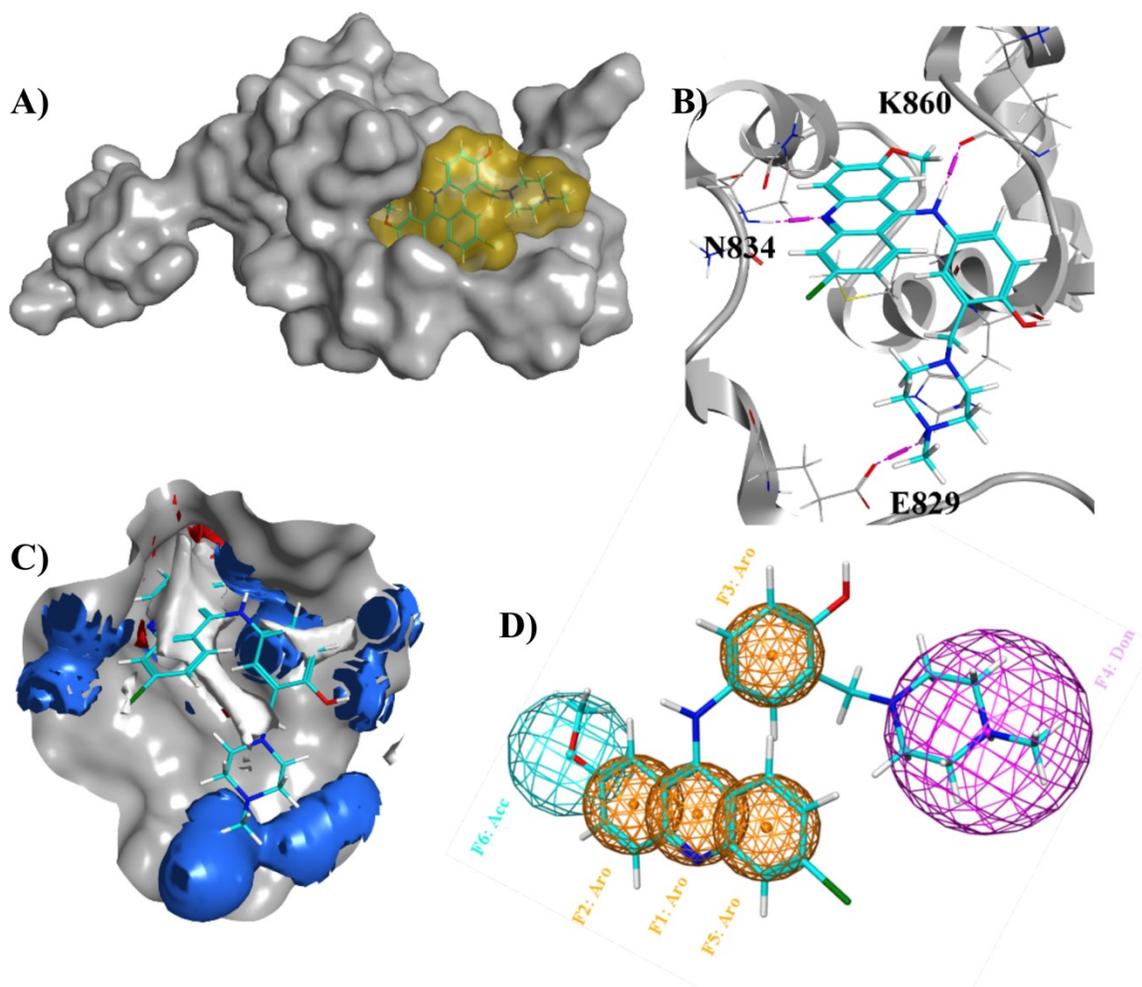


Figure B.2. Visual analysis of the F06 docking pose. A) Shape complementarity between the docked pose of F06 and the XPF binding site. The compound was docked to the XPF pocket that wraps around residue Phe293 of ERCC1 in the dimerization complex. B) Non-bonded interactions. Observed hydrogen bonds are represented in purple. Residues involved in hydrogen bonding or hydrophobic interactions with the ligand are explicitly represented. See text for more details. C) Affinity mapping of the F06 binding site on the XPF HhH2 domain. Blue color indicates zones of interaction where a donor probe atom shows a potential of -2 kcal/mol, therefore where a ligand donor atom will contribute favourably to the free energy of binding. Red indicates acceptor favorable zones, and white hydrophobic favorable zones. D)

Pharmacophore model designed for the lead compound binding to the XPF binding site. Aro features are aromatic, Acc are hydrogen bond acceptors and Don hydrogen bond donors.

Table B.1. Pharmacophore features built from the docking pose of F06 to the XPF site.

<b>Feature</b>	<b>Type</b>	<b>Radius (Å)</b>
F1	Aromatic	1.5
F2	Aromatic	1.5
F3	Aromatic	1.5
F4	Donor	3
F5	Aromatic	1.5
F6	Acceptor	2

Table B.2. Computational results for the subset of analogues chosen for synthesis. The ligand efficiency was calculated using the GBVI/WSA MOE score divided for the number of heavy atoms of each molecule. The analogues are ranked based on their MM/GBSA score which includes the entropy contributions. The lead compound F06 is reported as the last entry.

<b>Compound</b>	<b>GBVI/WSA (kcal/mol)</b>	<b>Ligand efficiency ((kcal/mol)/h.a.)</b>	<b>logP</b>	<b>MM/GBSA (kcal/mol)</b>
3	-7.06	-0.19	5.80	-21.73
4	-7.52	-0.20	2.61	-13.12
2	-6.83	-0.17	6.24	-11.60
7	-6.93	-0.18	5.50	-11.40
5	-7.57	-0.19	6.98	-9.62
6	-7.13	-0.15	7.63	-9.03
8	-6.44	-0.26	5.50	-4.46

1 (F06)	-6.70	-0.20	4.10	-17.78
---------	-------	-------	------	--------

### B.2.2 Synthesis and Characterization of Inhibitors (1-8)

#### **4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-methylpiperazin-1-yl)methyl)phenol (1)**

Synthesis of compound **1** was achieved through one pot sequential addition reaction in 3 steps. A mixture of *N*-methyl piperazine (0.12 mL, 1.0 mmol), acetaminophen (0.15 g, 1.0 mmol) and 37% formaldehyde (0.10 mL, 5 mmol) in isopropyl alcohol (2 ml) was taken in 10 mL single neck round bottom flask. Then, the reaction mixture was stirred and heated under reflux at 65 °C for 6 h. Upon completion of the reaction (monitored by TLC with 15% MeOH/DCM eluent system), the solvent was removed on a rotatory evaporator. Next, the residue was dissolved in 3 mL ethanol and 3 drops of 12 M HCl were added. Subsequently, the reaction mixture was heated at 90 °C under reflux for 90 min. Afterwards, 6,9-dichloro-methoxyacridine (0.28 g, 1.0 mmol) was added to the mixture and further stirred at 90 °C under reflux and the course of reaction followed by TLC until little or no starting material was detected (around 12 h). On cooling to room temperature, the reaction mixture was diluted with cold water and neutralized to pH of 8-9 with 28% v/v ammonia solution. The alkaline solution was extracted with dichloromethane. The organic layer was washed with brine, concentrated in vacuum, and purified by column chromatography (gradient elution with 5% to 10% MeOH:DCM system) to afford compound **1** as orange reddish semisolid (0.08 g) in 78 % yield;  $R_f$  0.50 (2:8, MeOH:DCM); IR (cast film)  $\nu_{\max}$  = 3272, 2923, 1629, 1254, 1231, 1032, 926, 816, 815, 775  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.99 (s, 1H), 7.91 (d,  $J$  = 8.7 Hz, 1H), 7.80 (d,  $J$  = 9.2 Hz, 1H), 7.29 (dd,  $J$  = 9.4, 2.6 Hz, 1H), 7.11 (d,  $J$  = 10.7 Hz, 1H), 7.03 (s, 1H), 6.77 – 6.70 (m, 2H), 6.49 (d,  $J$  = 2.5 Hz, 1H), 3.63 (s, 3H), 3.53 (s, 2H), 2.75 – 2.32 (m, 8H), 2.27 (s, 3H), OH and NH protons were not

observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  155.9, 153.2, 147.8, 143.9, 137.0, 135.0, 125.0, 124.9, 122.0, 121.8, 120.0, 119.9, 119.7, 117.61, 116.8, 116.5, 116.1, 115.9, 100.3, 61.1, 55.2, 54.8 (2C), 52.4 (2C), 45.8; HRMS (ESI) calcd for  $\text{C}_{26}\text{H}_{28}\text{ClN}_4\text{O}_2$   $[\text{M} + \text{H}]^+$  463.1895; found 463.1890.

***4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-(3-methylbenzyl)piperazin-1-yl)methyl)phenol (2)***

The previous method was employed to synthesize compound 2 with the following stoichiometric amounts; 1-(3-methylbenzyl)piperazine (0.19 g, 1.0 mmol) and 6,9-dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.14 g) in 73 % yield;  $R_f$  0.44 (1:9, MeOH:DCM); IR (cast film)  $\nu_{\text{max}}$  = 3258, 2921, 1629, 1560, 1493, 1253, 1133, 1007, 927, 826, 775  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.05 (s, 1H), 7.95 (d,  $J$  = 9.2 Hz, 1H), 7.83 (d,  $J$  = 9.2 Hz, 1H), 7.32 (d,  $J$  = 7.6 Hz, 1H), 7.21 (t,  $J$  = 7.5 Hz, 1H), 7.17 – 7.06 (m, 5H), 6.87 (d,  $J$  = 7.1 Hz, 1H), 6.77 (d,  $J$  = 8.6 Hz, 1H), 6.62 (d,  $J$  = 1.9 Hz, 1H), 3.70 (s, 3H), 3.59 (s, 2H), 3.49 (s, 2H), 2.79 – 2.40 (m, 8H), 2.35 (s, 3H). OH and NH protons were not observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  156.1, 153.9, 150.4, 146.8, 144.5, 137.9, 137.6, 136.3, 135.5, 129.9, 128.2, 127.9, 126.2, 125.4, 125.0, 122.1, 121.8, 120.6, 120.3, 119.2, 116.9, 116.5, 116.1, 115.9, 100.4, 62.8, 61.2, 55.4, 52.8 (2C), 52.5 (2C), 21.4; HRMS (ESI) calcd for  $\text{C}_{33}\text{H}_{34}\text{ClN}_4\text{O}_2$   $[\text{M} + \text{H}]^+$  553.2357; found 553.2365.

***4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-cyclohexylpiperazin-1-yl)methyl)phenol (3)***

The previous method was employed to synthesize compound 3 with the following stoichiometric amounts; 1-cyclohexyl piperazine (0.17 g, 1.0 mmol) and 6,9- dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.13 g) in 81 % yield;  $R_f$  0.52 (0.5:9.5, MeOH:

EtOAc); IR (cast film)  $\nu_{\max}$  = 3313, 2918, 1736, 1560, 1468, 1255, 1235, 1181, 1032, 929, 829, 722  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.03 (d,  $J$  = 1.4 Hz, 1H), 7.92 (d,  $J$  = 9.6 Hz, 1H), 7.80 (d,  $J$  = 9.5 Hz, 1H), 7.28 (d,  $J$  = 5.7 Hz, 1H), 7.11 (d,  $J$  = 9.1 Hz, 2H), 6.91 (d,  $J$  = 9.3 Hz, 1H), 6.77 (d,  $J$  = 8.5 Hz, 1H), 6.69 (s, 1H), 3.71 (s, 3H), 3.60 (s, 2H), 2.73 – 2.22 (m, 8H), 1.83 (m, 4H), 1.63 (d,  $J$  = 12.8 Hz, 1H), 1.26 – 1.18 (m, 6H). OH and NH protons were not observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  156.1, 153.8, 146.9, 144.5, 136.3, 135.5, 125.4 – 116.8 (10 C), 101.3, 100.4, 98.8, 63.4, 61.3, 52.9 (2 C), 48.7 (2 C), 29.7, 28.9 (2 C), 26.2, 25.8 (2 C); HRMS (ESI) calcd for  $\text{C}_{31}\text{H}_{35}\text{ClN}_4\text{O}_2$  [ $\text{M} + \text{H}$ ] $^+$  531.2521; found 531.2527.

***4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-(2-(dimethylamino)ethyl) piperazin-1-yl) methyl) phenol (4)***

The previous method was employed to synthesize compound 4 with the following stoichiometric amounts; 1-[2-(dimethylamino)ethyl] piperazine (0.16 g, 1.0 mmol) and 6,9-dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.13 g) in 80 % yield;  $R_f$  0.40 (1.3: 8.5: 0.2, MeOH: DCM:  $\text{Et}_3\text{N}$ ); IR (cast film)  $\nu_{\max}$  = 3361, 2918, 1736, 1562, 1467, 1295, 1255, 1143, 1032, 945, 828, 763, 722  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.10 (s, 1H), 8.00 (s, 1H), 7.89 (d,  $J$  = 9.1 Hz, 1H), 7.40 (d,  $J$  = 9.8 Hz, 1H), 7.24 (d,  $J$  = 8.0 Hz, 1H), 7.09 (s, 1H), 6.85 (dd,  $J$  = 8.6, 2.7 Hz, 1H), 6.78 (d,  $J$  = 8.6 Hz, 1H), 6.56 (d,  $J$  = 2.3 Hz, 1H), 3.74 (s, 3H), 3.59 (s, 2H), 2.70 – 2.40 (m, 12H), 2.27 (s, 6H). OH and NH protons were not observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  156.2, 153.6, 147.5, 144.0, 136.6, 135.2, 125.3, 125.1, 124.8, 122.0, 120.3, 120.2, 120.0, 119.8, 119.6, 117.5, 116.9, 116.3, 100.2, 61.2, 56.7, 56.7, 55.3, 53.3 (2C), 52.4 (2C), 45.8 (2C); HRMS (ESI) calcd for  $\text{C}_{29}\text{H}_{35}\text{ClN}_5\text{O}_2$  [ $\text{M} + \text{H}$ ] $^+$  520.2474; found 520.2473.

***4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-(3,4-dichlorophenyl)piperazin-1-yl)methyl)phenol (5)***

The previous method was employed to synthesize compound 5 with the following stoichiometric amounts; 1-(3,4-dichlorophenyl) piperazine (0.23 g, 1.0 mmol), acetaminophen (0.15 g, 1.0 mmol), 37% formaldehyde (0.10 mL, 5 mmol) and 6,9- dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.16 g) in 71 % yield;  $R_f$  0.56 (0.5:9.5, MeOH:DCM); IR (cast film)  $\nu_{\max}$  = 3271, 2918, 1736, 1563, 1227, 1180, 1080, 1031, 803, 721  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.09 (s, 1H), 8.00 (s, 1H), 7.88 (d,  $J$  = 9.2 Hz, 1H), 7.40 (d,  $J$  = 9.3 Hz, 1H), 7.28 (d,  $J$  = 8.7 Hz, 1H), 7.24 (d,  $J$  = 9.0 Hz, 1H), 7.09 (s, 1H), 6.95 (d,  $J$  = 2.8 Hz, 1H), 6.87 (dd,  $J$  = 8.6, 2.7 Hz, 1H), 6.80 (d,  $J$  = 8.6 Hz, 1H), 6.73 (dd,  $J$  = 8.9, 2.8 Hz, 1H), 6.54 (d,  $J$  = 2.6 Hz, 1H), 3.74 (s, 3H), 3.63 (s, 2H), 3.27 – 3.15 (m, 4H), 2.72 – 2.62 (m, 4H), OH and NH protons were not observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  156.3, 153.1, 150.2, 147.9, 143.3, 137.0, 135.0, 132.9, 130.5, 125.2, 124.7, 122.9, 121.6, 121.4, 120.2, 120.0, 119.7, 117.9, 117.6, 117.0, 116.7, 116.2, 115.6, 115.5, 100.0, 61.2, 55.4 (2C), 52.2 (2C), 48.7; HRMS (ESI) calcd for  $\text{C}_{31}\text{H}_{28}\text{Cl}_3\text{N}_4\text{O}_2$   $[\text{M} + \text{H}]^+$  593.1272; found 593.1272.

***2-((4-(Bis(4-fluorophenyl)methyl)piperazin-1-yl)methyl)-4-((6-chloro-2-methoxyacridin-9-yl)amino)phenol (6)***

The previous method was employed to synthesize compound 6 with the following stoichiometric amounts; 1-bis (4-fluorophenyl)methylpiperazine (0.29 g, 1.0 mmol), acetaminophen (0.15 g, 1.0 mmol), 37% formaldehyde (0.10 mL, 5 mmol) and 6,9- dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.19 g) in 65 % yield;  $R_f$  0.56 (0.5:9.5, MeOH:DCM); IR (cast film)  $\nu_{\max}$  = 3275, 2919, 1737, 1566, 1255, 1217, 1181, 1032, 828, 722  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500

MHz, CDCl<sub>3</sub>) δ 8.07 (s, 1H), 7.98 (d, *J* = 9.0 Hz, 1H), 7.84 (d, *J* = 9.2 Hz, 1H), 7.34 (m, 6H), 7.18 (d, *J* = 8.0 Hz, 1H), 7.06 (d, *J* = 1.8 Hz, 1H), 6.97 (m, 5H), 6.83 (dd, *J* = 8.5, 2.6 Hz, 1H), 6.75 (d, *J* = 8.6 Hz, 1H), 6.53 (d, *J* = 2.2 Hz, 1H), 4.23 (s, 1H), 3.69 (s, 3H), 3.57 (s, 2H), 2.74 – 2.21 (m, 8H); <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 161.9 (2C), 156.2, 153.5, 150.3, 147.7, 143.7, 137.8 (2C), 136.7, 135.0, 129.2 (4C), 125.2, 124.8, 121.9, 121.7, 120.2, 119.9, 119.7, 117.6, 116.8, 116.5, 116.1, 115.5 (4C), 100.0, 74.3, 61.1, 55.3, 52.6 (2C), 51.5 (2C); HRMS (ESI) calcd for C<sub>38</sub>H<sub>34</sub>ClF<sub>2</sub>N<sub>4</sub>O<sub>2</sub> [M + H]<sup>+</sup> 651.2333; found 651.2328.

***4-((6-Chloro-2-methoxyacridin-9-yl)amino)-2-((4-(4-hydroxyphenyl)piperazin-1-yl)methyl)phenol (7)***

The previous method was employed to synthesize compound 7 with the following stoichiometric amounts; 1-(4-hydroxyphenyl)piperazine (0.30 g, 1.0 mmol), acetaminophen (0.15 g, 1.0 mmol), 37% formaldehyde (0.10 mL, 5 mmol) and 6,9-dichloroacridine (0.28 g, 1.0 mmol) to afford it as orange reddish semisolid (0.17 g) in 59 % yield; *R<sub>f</sub>* 0.58 (1.5:8.5, MeOH:DCM); IR (cast film)  $\nu_{\text{max}}$  = 3334, 2918, 1737, 1564, 1468, 1235, 1181, 1031, 930, 830, 722 cm<sup>-1</sup>; <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.07 (d, *J* = 9.3 Hz, 1H), 7.87 (d, *J* = 1.8 Hz, 1H), 7.82 (d, *J* = 9.3 Hz, 1H), 7.51–7.46 (m, 1H), 7.41 (d, *J* = 2.5 Hz, 1H), 7.29 – 7.24 (m, 1H), 7.04 (dd, *J* = 8.7, 2.4 Hz, 1H), 6.93 – 6.89 (m, 1H), 6.88 – 6.81 (m, 3H), 6.70 (dd, *J* = 8.9, 2.3 Hz, 2H), 3.72 (s, 3H), 3.61 (s, 2H), 3.11 – 2.99 (m, 4H), 2.82 – 2.67 (m, 4H), OH and NH protons were not observed; <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 156.3, 153.5, 152.1, 150.2, 150.1, 145.3, 145.2, 144.0, 125.3, 124.7, 122.5, 121.9, 121.5, 120.4, 120.0, 118.8 (2C), 118 (2C), 118.4, 118.1, 117.0, 116.5, 116.1, 115.9, 61.3, 55.4, 52.7 (2C), 50.7 (2C); HRMS (ESI) calcd for C<sub>31</sub>H<sub>30</sub>ClN<sub>4</sub>O<sub>3</sub> [M + H]<sup>+</sup> 541.2001; found 541.2010.

#### **4-((6-Chloro-2-methoxyacridin-9-yl)amino)phenol (8)**

Synthesis of compound 7 was carried out by mixing 4-aminophenol (0.11 g, 1.0 mmol) and 6,9-dichloroacridine (0.28 g, 1.0 mmol) in 10 mL single neck round bottom flask and stirred at 90 °C under reflux. The course of reaction followed by TLC until little or no starting material was detected (around 12 h). On cooling to room temperature, the reaction mixture was diluted with cold water and neutralized to pH of 8-9 with 28% v/v ammonia solution. The alkaline solution was extracted with dichloromethane. The organic layer was washed with brine, concentrated in vacuum, and purified by column chromatography (gradient elution with 1% to 5% MeOH:DCM system) to afford it as orange reddish semisolid in 73 % yield;  $R_f$  0.47 (1.5:8.5, MeOH:DCM); IR (cast film)  $\nu_{max}$  = 3325, 2918, 1737, 1562, 1468, 1236, 1180, 1103, 1050, 722  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  8.12 (d,  $J$  = 9.4 Hz, 1H), 7.84 – 7.83 (m, 1H), 7.78 (d,  $J$  = 9.3 Hz, 1H), 7.61 (dd,  $J$  = 9.3, 2.6 Hz, 1H), 7.48 (d,  $J$  = 1.5 Hz, 1H), 7.35 (dd,  $J$  = 9.4, 2.1 Hz, 1H), 7.27 – 7.22 (m, 2H), 6.95 – 6.91 (m, 2H), 3.88 (s, 1H), 3.69 (s, 3H). OH proton was not observed;  $^{13}\text{C}$  NMR (125 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  158.9, 157.7, 141.7, 137.9, 133.2, 129.8, 128.4, 127.8 (2C), 126.6, 125.7, 122.2, 120.1, 119.3, 117.8 (2C), 116.1, 113.3, 104.5, 56.1.; HRMS (ESI) calcd for  $\text{C}_{20}\text{H}_{16}\text{ClN}_2\text{O}_2$   $[\text{M} + \text{H}]^+$  351.0908; found 351.0896.

Table B.3. Half-maximum inhibitory concentrations for the compounds with the highest inhibitory potentials. The data was obtained from at least three different experiments of  $V_o$  versus compound concentration.

<b>Compound</b>	<b><math>\text{IC}_{50} \pm \text{SD}</math> (<math>\mu\text{M}</math>)</b>	<b><math>\text{K}_d \pm \text{SD}</math> (nM)</b>
1 (F06)	$1.86 \pm 0.25$	$140 \pm 5$
3	$0.38 \pm 0.10$	$145 \pm 5$
4	$0.33 \pm 0.07$	$100 \pm 5$

B.2.3 Cytotoxicity Profile of Compound 4

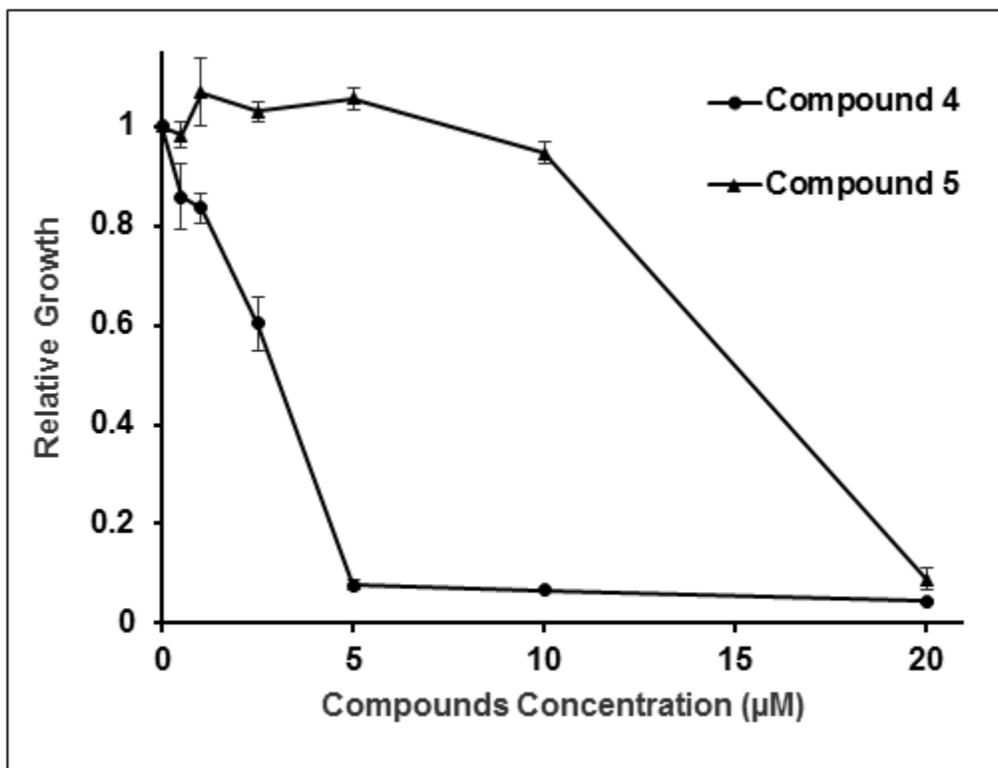


Figure B.3. Cell proliferation of HCT-116 cells treated with compounds 4 (active inhibitor) and 5 (non-active inhibitor) using MTS assay.

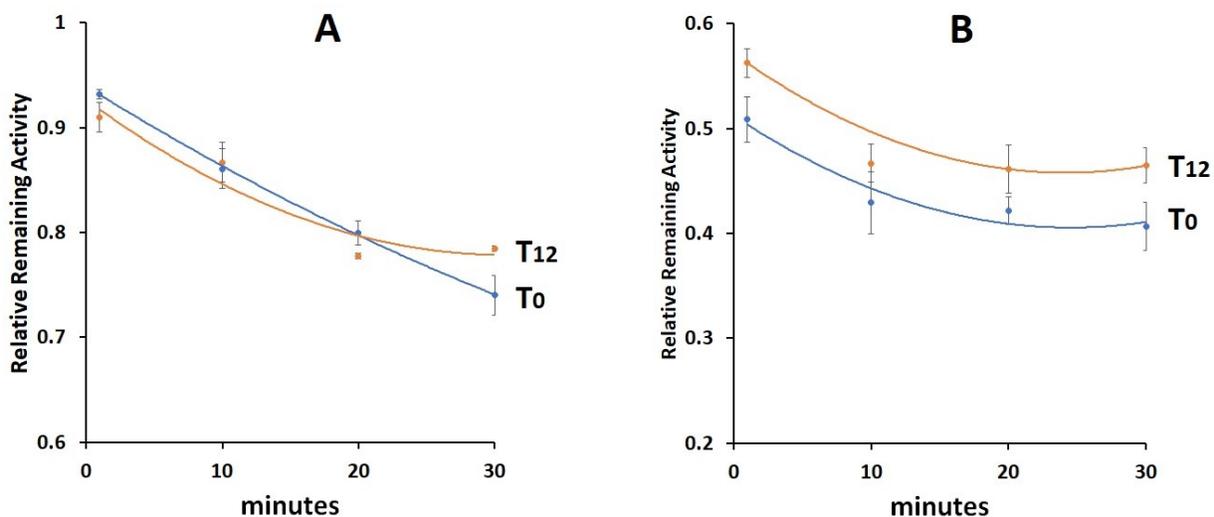


Figure B.4. The inactivation experiments were carried out at 25°C in the reaction medium indicated in the “Materials and Methods” section which also included 10% or 5% DMSO final concentration (A or B, respectively). The reaction was started (T0) by the addition of ERCC1-XPF (2.5 µg protein/ml) followed by the addition of 3 µM compound 4. 10 µl aliquots from these media (25 ng protein) were withdrawn at the indicated times and the activity ( $\Delta$ RFU/time) was measured for 6 minutes by adding 200 nM stem-loop substrate and additional compound 4 diluted in DMSO (10% or 5% DMSO final concentration, A and B, respectively) to render final concentrations of 100 nM stem-loop substrate and 3 µM compound 4 in 10 µl for a final total reaction volume of 20 µl. In T12, the activity was measured by the addition of 200 nM stem-loop substrate which had already been pre-incubated with 3 µM compound 4 for 12 minutes to give the same final concentrations of reagents as indicated for T0.

## Appendix C - Commentary of Chapter 4<sup>9</sup>

We recently published a scientific article [82] on the discovery of small molecule inhibitors for the nucleotide excision repair (NER) protein complex, XPA-ERCC1. In this paper, we reported on the computational workflow we adopted to screen for compounds that bind to ERCC1 and block their interactions with XPA. Following this workflow, we identified promising scaffolds with the potential of modulating the NER pathway. In this commentary, we discuss the relevant findings of this study as well as its limitations and future directions.

In our recent study, we identified small molecule compounds with the potential of regulating the NER DNA repair pathway. The compounds were selected to disrupt the XPA-ERCC1 protein-protein interaction. Combining such drugs with DNA damaging agents can improve their effectiveness and allow for the use at a reduced dose in cancer therapy [99]. We employed a computational workflow to screen the entire PubChem [65] and National Cancer Institute (NCI) small molecule [68] repositories to identify potential binders to the XPA-binding site within the ERCC1 protein. This binding site lies within the central domain of ERCC1.

The rationale for our study was to build upon an earlier study by our group [80] which identified two small molecule structures that target the XPA-ERCC1 interaction. These compounds were used as a starting point for our virtual screening (VS) campaign. In this context, we followed two independent screening approaches, namely a similarity-based approach and a pharmacophore-based approach.

---

<sup>9</sup> This appendix has been published as commentary as has been published as Gentile, F.; Tuszynski, J. A.; Barakat, K. H. Commentary: New design of nucleotide excision repair (NER) inhibitors for combination cancer therapy. *J Cancer Treat Diagnosis* 2017, 1, 1–3.

For the similarity-based approach, we filtered the PubChem database for compounds that are structurally similar to these two lead molecules. Using a similarity Tanimoto score between chemical fingerprints, we retained only around 22,000 small molecules out of the 68 million compounds in the database. In addition, we performed an *in-silico* filtering step to retain only the molecules with drug-like properties, based on their Adsorption, Distribution, Metabolism, Elimination and Toxicity (ADMET) properties. This approach provided a significant improvement over the original study [80] as it focused on compounds with drug-like properties. This strategy reduced the enormous number of compounds in PubChem and built a focused library of compounds.

As a second screening approach, we used molecular docking simulations to study the binding modes of the originally identified compounds [80] and build a consensus pharmacophore model based on their docked poses. We then employed this model to filter four different subsets of the NCI databases for compounds that satisfy conditions imposed by this pharmacophore. For this approach, we aimed to identify new scaffolds that are able to inhibit the protein-protein interaction, with different chemical structures other than the original lead compounds. As a result, we used a lead-like filter to retain only the structures that can be used as lead compounds for further optimization. These two ligand-based strategies complemented each other. On the one hand, we identified improved drug-like analogues over the original lead compounds and on the other hand, we provided a set of diverse lead scaffolds.

Compounds coming from the two different screening funnels (i.e. similarity-based and pharmacophore-based) were ranked based on their binding energies within ERCC1. Therefore, we adopted a target-based VS protocol. For our docking simulations, we used seven optimized ERCC1 NMR conformations in order to accommodate the flexibility of both the side chains and

backbone atoms of the binding site. After the docking simulation for each compound, we ran short molecular dynamics (MD) simulations for the top hits, followed by molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) rescoring [28]. We concluded our study by providing not only a more reliable rank of the top hits compounds, but also to quantitatively represent the binding mode of the molecules to the key residues of the ERCC1 binding site.

Although the two screening approaches provided a comprehensive way to identify new scaffolds with drug-like properties, there are two critical points that we discuss next in this commentary. First concerns the relatively short MD simulations (2 nanoseconds) that were used to relax the top hits from docking simulations. Second is the neglect of the conformational entropy contribution when calculating the binding energies using the MM/PBSA method. For the first aspect, we decided to use this short simulation time scale, because of a deterioration of the ranking power of the MM/PBSA method was reported when applied to longer simulations [195,196]. Moreover, our group has used short timescales in a number of successful drug design studies [80,84]. These considerations, together with the limits derived by the availability of computational resources, drove our choice to use 2 nanoseconds of simulation for the docked complexes. We are aware of the fact that 2 nanoseconds is an insufficient amount of time to simulate any major conformational change of a molecule or a complex. However, we believe this was not relevant to our study. Regarding the entropy contribution, we performed the computationally-expensive calculations only for a subset of structurally-diverse hits. We found entropy values of the same order of magnitude for all the complexes, and hence we neglected that for the remaining hits. It is noteworthy that entropy contributions were shown to introduce large fluctuations to the binding energies in different studies. Also, these contributions can be

safely ignored to obtain a relative rank of compounds with similar size and flexibility binding to the same target, as in our case.

In our work, we have adopted a multi-step VS protocol to identify XPA-ERCC1 inhibitors. We initially selected compounds that are either similar in structure or pharmacophore features to the known active inhibitors. In our opinion, this step should reduce the number of false positive results obtained from target-based screening methodologies. We then performed a first target-based screening retaining compounds which showed better binding energies than the lead structures, as calculated with the docking scoring function. Being aware of the limitations of such kind of methods regarding the ranking, we ran MD simulations followed by MM/PBSA rescoring. We believe that our approach was able to identify active inhibitors, having taken into account the physicochemical similarities with the lead compounds and also the binding energy values calculated with two different methods (docking scoring function followed by MM/PBSA rescoring). In conclusion, the outcome of our study was a set of seventy-two small molecules with a binding energy value more favorable than the two lead compounds. The top eighteen hits are reported in Table C.1. The details of the remaining fifty-four compounds are available upon request. Our hits showed similar interaction patterns among the residues constituting the XPA binding site of ERCC1. Although it was our strong belief that our hits can be the key to the next generation of NER inhibitors, we recognized that the lack of experimental evidence constitutes a critical limitation of our findings. Hence, we strongly recommend to test these compounds in protein and cell-based assays for their ability to bind to the ERCC1 central domain and to inhibit the interaction with the XPA protein and, consequently, inhibit the NER pathway in cancer cells. Also, further experiments will be required to test the ability of these compounds to act synergistically with DNA damaging cancer therapy such as platinum-based drugs.

Table C.1. Top eighteen hits of the VS experiment, scored according to their MM/PBSA binding energy. All the compounds showed better values of the energy than the lead structures (rank 19 and 20). In total, seventy-two molecules showed lower binding energy values. The ID column indicates the ID associated to the compound in its original database. The MM/PBSA energies do not include the entropy contribution. The VS method column indicates the ligand-based screening technique used to include the compound in the target-based VS step from the indicated database (last column).

<b>Rank</b>	<b>ID</b>	<b>MM/PBSA binding energy (kcal/mol)</b>	<b>VS method</b>	<b>Database</b>
1	6210903	-33.84	Similarity search	PubChem
2	7324126	-33.41	Similarity search	PubChem
3	8486248	-31.49	Similarity search	PubChem
4	7730851	-31.25	Similarity search	PubChem
5	1696060	-30.29	Similarity search	PubChem
6	1161060	-30.11	Similarity search	PubChem
7	1098945	-30.11	Similarity search	PubChem
8	53684246	-29.75	Similarity search	PubChem
9	7260552	-29.48	Similarity search	PubChem
10	6971912	-29.47	Similarity search	PubChem
11	5105640	-29.33	Similarity search	PubChem
12	2645792	-28.56	Similarity search	PubChem
13	8013886	-28.23	Similarity search	PubChem
14	1705254	-28.16	Similarity search	PubChem
15	24539908	-28.06	Similarity search	PubChem

16	5113	-28.24	Pharmacophore	NCI Natural Products
17	106408	-25.41	Pharmacophore	NCI Mechanistic
18	107582	-25.25	Pharmacophore	NCI Diversity
19	Compound 10	-24.86	Lead compound	-
20	NERI01	-23.24	Lead compound	-

## **Appendix D - A Novel Interaction Between the TLR7 and a Colchicine Derivative Revealed Through a Computational and Experimental Study<sup>10</sup>**

### **D.1 Abstract**

The Toll-Like Receptor 7 (TLR7) is an endosomal membrane receptor involved in the innate immune system response. Its best-known small molecule activators are imidazoquinoline derivatives such as imiquimod (R-837) and resiquimod (R-848). Recently, an interaction between R-837 and the colchicine binding site of tubulin was reported. To investigate the possibility of an interaction between structural analogues of colchicine and the TLR7, a recent computational model for the dimeric form of the TLR7 receptor was used to determine a possible interaction with a colchicine derivative called CR42-24, active as a tubulin polymerization inhibitor. The estimated values of the binding energy of this molecule with respect to the TLR7 receptor were comparable to the energies of known binders as reported in a previous study. The binding to the TLR7 was further assessed by introducing genetic transformations in the TLR7 gene in cancer cell lines and exposing them to the compound. A negative shift of the IC<sub>50</sub> value in terms of cell growth was observed in cell lines carrying the mutated TLR7 gene. The reported study suggests a possible interaction between TLR7 and a colchicine derivative, which can be explored for rational design of new drugs acting on this receptor by using a colchicine scaffold for additional modifications.

### **D.2 Introduction**

---

<sup>10</sup> This appendix has been published as has been published as Gentile, F.; Deriu, M. A.; Barakat, K.; Danani, A.; Tuszynski, J. A Novel Interaction Between the TLR7 and a Colchicine Derivative Revealed Through a Computational and Experimental Study. *Pharmaceuticals* 2018, 11, 22.

In humans, Toll-like receptors (TLRs) are a family of ten receptors (TLR1-10), which are part of the innate immune system. The innate immune system's main function is to recognize pathogen-associated molecular patterns (PAMPs), which belong to microbial pathogens such as bacteria, viruses, fungi and protozoa, in addition to damage-associated molecular patterns (DAMPs) coming from damaged cells [250,251]. It is commonly assumed that each TLR receptor recognizes specific molecular patterns, and all of them are multi-domain, trans-membrane proteins localized in the cellular membrane, except for TLR7, 8 and 9 which are found on the endosomal membrane [250]. The common structure of TLRs involves an ectodomain comprising leucine-rich repeats (LRR), a transmembrane domain and a Toll/IL-1 receptor (TIR) domain, which initiates the cascade of signaling after a PAMP/DAMP recognition event by engaging TIR domain-containing proteins such as myeloid differentiation primary response 88 (MyD88) and TIR-domain-containing adapter-inducing interferon- $\beta$  (TRIF) proteins. This transduction pathway results in the production of cytokines, chemokines and type I interferons to protect the cells from microbial invasion [251].

The TLR7 is an endosomal transmembrane protein composed of 1049 amino acid residues, divided between an endosomal ectodomain with twenty-seven LLR arranged in a horseshoe structure, a trans-membrane domain and a TIR domain located in the cellular compartment. The functional form of TLR7 is a homodimer complex [252]. The natural ligands recognized by TLR7 are viral single-strand RNA (ssRNA). The signaling cascade is mediated through the interaction with MyD88 [253]. Synthetic ligands that bind to TLR7 are synthetic ssRNA and nucleoside analogues, such as imidazoquinoline, adenosine and guanosine derivatives [254,255]. Recently, different TLR7-targeting small molecules were developed as immune response modifiers for anti-cancer and anti-viral therapy. Imiquimod (R-837) (see Figure D.1A) was the

first Food and Drug Administration (FDA)-approved TLR7 agonist, and its formulation containing 5% of cream (Aldara®) is used to treat genital warts caused by human Papilloma virus (HPV) infection, and malignant skin cancers. This imidazoquinoline derivative induces the production of pro-inflammatory cytokines and other molecules upon TLR7 activation [256,257]. Resiquimod (R-848) and CL097 are two dual TLR7/8 agonists with a similar effect on R-837 [257–259]. Interestingly, efforts were made for developing both activators and inhibitors of TLR7. While activation of the receptor is sought to enhance the immune response to viral agents, its inhibition gains interest in the treatment of autoimmune disease such as lupus [260]. In the context of cancer treatment, again the interest is in both directions. First, TLR7 activation is an assessed strategy for cancer immunotherapy, where its stimulation leads to the secretion of anti-tumor cytokines [261–263]. However, it has been reported that the receptor expression is upregulated in human pancreatic tumor cells, its signaling regulates carcinogenesis, TLR7 ligation promotes tumor progression and TLR7 blockade protects against cancer development [264]. This suggests a possible role of TLR7 inhibitors in the treatment of pancreatic cancer [265].

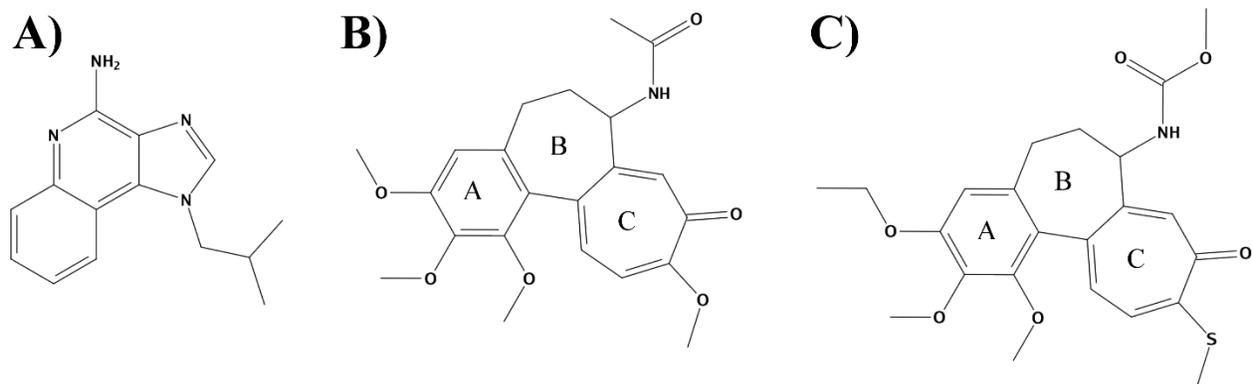


Figure D.1. Chemical structure of A) R-837, B) colchicine and C) CR42-24. For the latter two, the rings are named according to the convention used in the article (A, B, C).

Colchicine is a natural compound extracted from the meadow saffron (*Colchicum autumnale*). It was identified as the first microtubules-destabilizing agent, and it is an FDA-approved drug for the treatment of familial Mediterranean fever and gout. Colchicine acts by binding to a binding site in the  $\beta$ -tubulin monomer, resulting in a curved tubulin dimer conformation, which prevents microtubule assembly, thus inhibiting mitosis. Although it has a cytotoxic effect against cancer cells due to their accelerated rate of mitosis, the use of colchicine in chemotherapy is limited by high toxicity affecting normal cells and multidrug resistance [266]. To overcome these problems, numerous colchicine derivatives were synthesized and tested in the past few years, mainly by modifying the B and C rings of the original structure (Figure D.1B) [267]. Through an iterative process of computational screening and subsequent cytotoxicity testing the Tuszynski lab identified a number of compounds that showed enhanced binding affinity for  $\beta$ III tubulin. A compound coded CR42-24 was selected based on its high affinity to bind  $\beta$ III tubulin, and showing improved cytotoxicity over other derivatives. CR42-24 is a colchicine analogue with the following modifications of the original structure (see Figure D.1C): one of the methoxy groups of the A ring is replaced by an ethoxy group, the acetamide in B ring is replaced by a methyl carbamate, and the methoxy group of the C ring is substituted by a methyl sulfanyl group. CR42-24 had shown great promise in its ability to treat cancer and it had been shown to be highly effective against breast cancer, leukemia, and lung cancer cell lines. An extensive review of this compound and a number of similar structures has been recently published where detailed information about their profiles can be found [268].

Recently, it has been reported for the first time that R-837 might bind to the colchicine-binding site of tubulin, hence inhibiting tubulin polymerization [269]. Molecular docking of R-837 to this site revealed a binding mode similar to the experimentally obtained colchicine one, showing

agreement with the colchicine site pharmacophores published earlier for active compounds [270]. In this research, we explored the reverse process, i.e. we aimed to understand if compounds binding to the colchicine binding site in tubulin could possibly bind to the R-837 binding site at the interface between the two TLR7 monomers. Starting from our recent computational model of the human TLR7 model [271], we employed a series of *in-silico* methods in order to identify and quantitatively characterize the binding mode of CR42-24 to the TLR7 site. We also employed a cell line-based experiment to demonstrate the possible interaction between the compound and TLR7, by studying the effect that CR-24-42 has on cancer cells carrying either a wild-type or a mutant TLR7 gene, respectively.

CR42-24 is an in-house discovered and developed, potent microtubules-destabilizing agent with promising anti-cancer activities. The results reported in this article are of potential interest not only to predict possible side effects of CR42-24 on the immune system, but also to enable the design of novel small molecules targeting TLR7 with structures based on this scaffold.

### **D.3 Results and Discussion**

#### *D.3.1 Homology Modeling*

We have already reported the validation results for our TLR7 model in [271]. To summarize, the sequence identity between the TLR7 and TLR8 ectodomains was around 46%. The model that was generated includes 1514 residues equally divided in two monomers (757 residues each, hereinafter referred to as monomer a and b, respectively), excluding the residues that are removed after the proteolytic cleavage required for TLR7 activation [272]. Our model showed ERRAT scores and dihedral angle values (calculated with PROCHECK) comparable with the experimental template and the other TLR7 homology models reported in literature [273–275],

especially after the minimization step. In fact, just 1% of modeled residues showed dihedral values in the disallowed zone of the Ramachandran plot. The 3D structure of our TLR7 model is reported in Figure D.2.

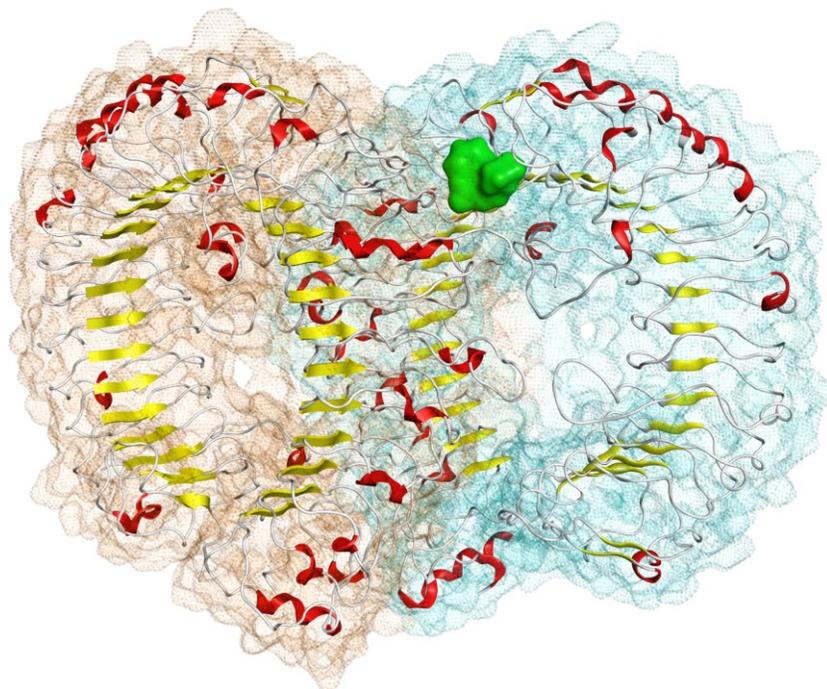


Figure D.2. Structure of the homodimeric model of the TLR7 ectodomains obtained in our previous work from the TLR8 template. The a and b monomer surfaces are represented in brown and cyan, respectively. Each monomer secondary structures are colored as follow: helices in red, strands in yellow, loops and turns in white. The R-837 binding site targeted in this study, lying at the interface of the two monomers, is represented as green surface.

### *D.3.2 Molecular Docking*

Among the top five poses obtained from the docking simulations, we identified one which showed a binding geometry similar to the R-837 pose obtained and validated in our previous study [271] (Figure D.3A). The remaining four poses are reported in Figure D.7. The binding energy of the selected pose, calculated with the Generalized Born Volume Integral/Weighted Surface Area (GBVI/WSA) scoring function [221], was -8.194 kcal/mol. The binding energy of

R-837, calculated by rescoring the docking pose with the GBVI/WSA method, was -4.991 kcal/mol. Regarding the specific interactions with the residues of the binding pocket, R-837 was involved in hydrophobic contacts mainly with residue aL557 and bF408 (where a and b refer to the first and second monomer, respectively), having the three-ring structure interposed between the two hydrophobic residues. The same contacts, in addition to others, were observed for the three rings of CR42-24, although the higher flexibility of the B ring resulted in a less planar geometry, when compared with R-837. R-837 also showed hydrogen bonding with the charged side chain of aD555 and the backbone oxygen of aT586. Although we did not observe any hydrogen bonds being established in the CR42-24-TLR7 complex, the amine of the acetamide group of the compound was positioned in the same zone of the amine group of R-837, a polar-favorable site of the binding pocket surrounding the negatively charged side chain of residue aD555. Also, the carbonyl group of the C ring of CR42-24 was positioned in a polar-favorable zone lying around the positively charged side chain of residue bK432

Finally, we observed some similarities between the docking pose of CR42-24 and the colchicine conformation co-crystallized in the tubulin binding site as reported by Courbet *et al* [269], especially in terms of hydrophobic contacts. Indeed, in tubulin, the C ring of colchicine interacted through van der Waals interactions with a first hydrophobic pocket composed by  $\alpha$ V181,  $\alpha$ S178, and  $\beta$ V315, where  $\alpha$  and  $\beta$  indicate the respective tubulin monomer. In our TLR7-bound model of CR42-24, the corresponding C ring was embedded in a hydrophobic cleft composed by aL557 and the aromatic rings of bY356 and bF408 (labelled in purple in Figure D.3B). In tubulin, the A ring of colchicine was embedded in a second hydrophobic pocket, constituted by several residues. In TLR7, similarly, the corresponding A ring in CR42-24 is also embedded in a second hydrophobic zone, constituted by residues aI585, aT586, bF351 and

bV381 (in orange in Figure D.3B). The superposition between the CR42-24 docked pose in TLR7 and the *N*-deacetyl-*N*-(2-mercaptoacetyl)-colchicine (DAMA-colchicine) binding pose to tubulin found in the pdb structure 1SA0 [28] is reported in Figure D.8.

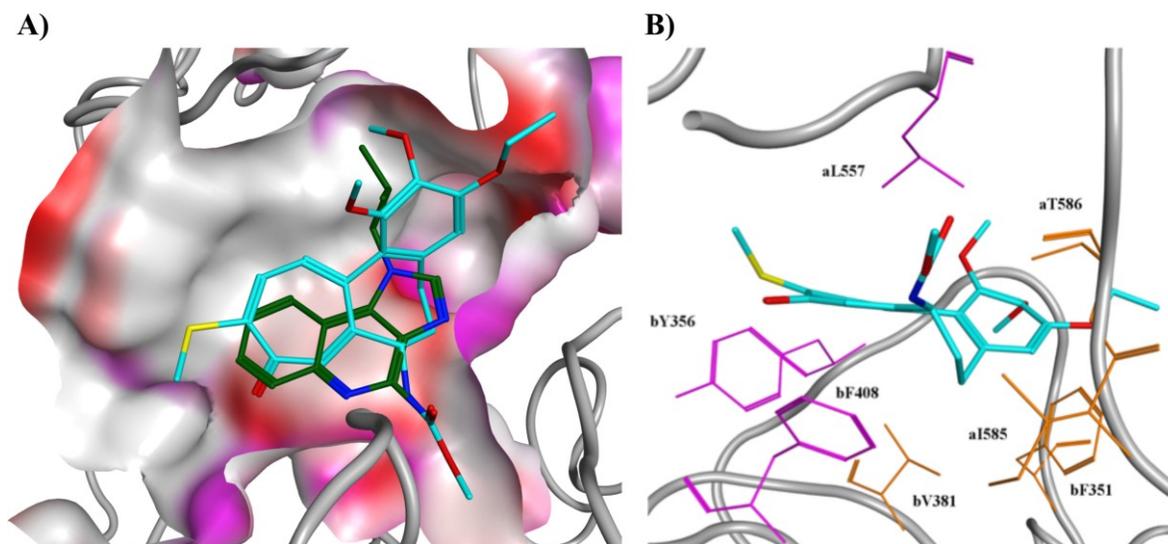


Figure D.3. Analysis of the predicted binding pose of CR42-24 to the R-837 binding site of TLR7. A) Comparison between the docking poses of R-837 (dark green carbons) and CR42-24 (cyan carbons) to the TLR7 binding site. Red surface indicates solvent-exposed zones, white surface indicate hydrophobic zones and purple surface indicates polar zones of the site. Key common interactions are the position of the three-ring moieties in the hydrophobic zone of the binding pocket and the orientation of amine groups in both the compounds towards the polar zone surrounding aD555. In addition, the carbonyl group of the C ring of CR42-24 was oriented in proximity of bK432. B) Hydrophobic contacts of CR42-24 with the TLR7 residues of the binding pocket. From the reported docking pose, we observed two distinct hydrophobic zones interacting with the compound, one constituted by aL557, bY356 and bF408 (represented in purple) and the other one by aI585, aT586, bF351 and bV381 (represented in orange). Two hydrophobic clefts, interacting with the same compound rings (A and C) as in our model, were observed for colchicine bound to the tubulin site. The pose is represented in an orientation similar to Figure S6 from [269] for comparison with the colchicine binding pose to the tubulin site.

### *D.3.3 Molecular Dynamics Simulations and MM/GBSA Calculations*

Although the assessed power of docking to identify near-native poses of ligands within binding sites, some limitations arise when using these methods. Firstly, the binding site flexibility is at least fully neglected, although it is possible to introduce some degrees of flexibility for its side chains. Also, docking simulations provide a static picture of the binding interactions, and a binding energy calculated from on single pose rather than an average made over an ensemble. Lastly, the water molecules around the site are usually treated implicitly during docking. In the light of these considerations, we performed explicit solvent MD simulations of the ligand-receptor complex obtained from the docking. The root-mean square deviation (RMSD) trends of the protein backbone, the backbone and side chain atoms of the binding site and the heavy atoms of CR42-24 are reported in Figure D.4. We observed RMSD values consistent with a stable, active ligand-receptor complex. Indeed, CR42-24 stabilized around 1 Å of RMSD value after about 4 ns of production simulation, and maintained the same fluctuation trend for the remaining simulated time. In our previous study, we observed a similar trend for the RMSDs of activators of TLR7, and consistently more unstable trends for inactive compounds in complex with the receptor [271]. The equilibrated RMSD trends of the protein backbone and the binding pocket were also similar to the trends observed in our earlier study. Stable RMSD trends were also observed when the average structure of the last 5 ns was used as reference (Figure D.9).

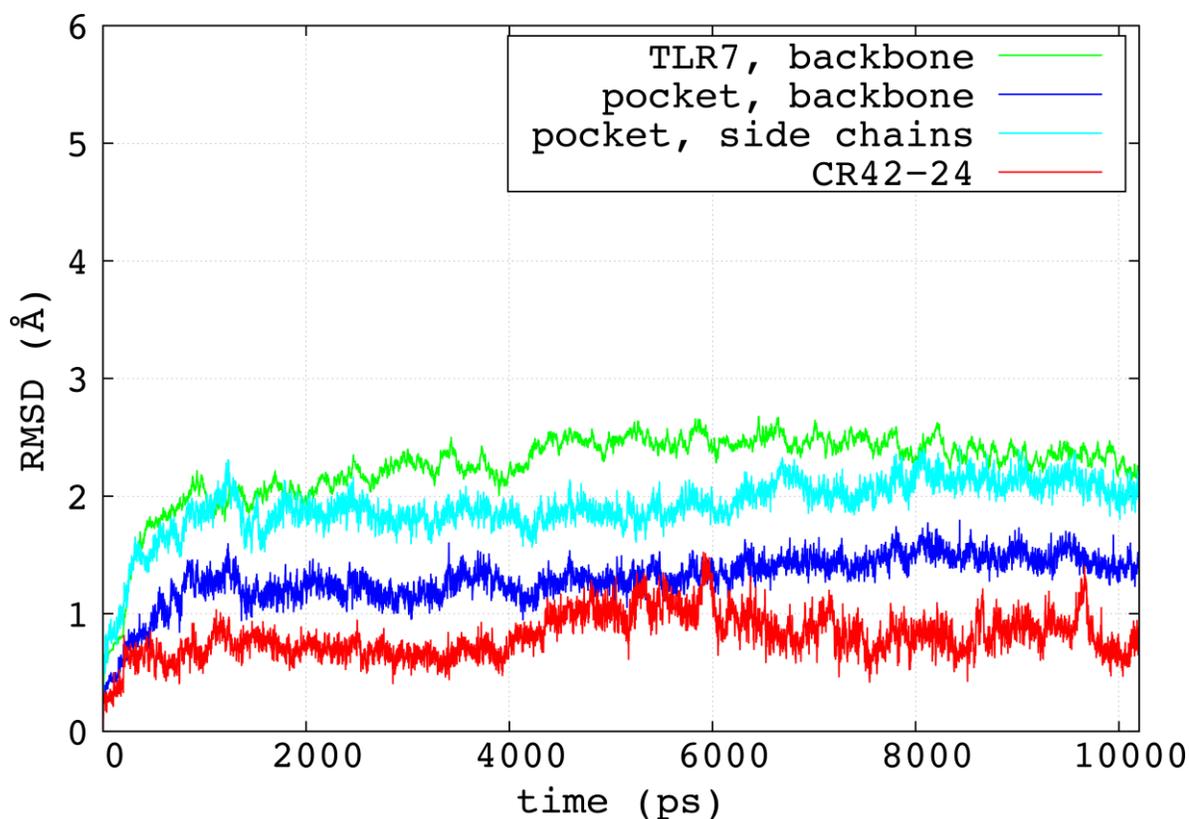


Figure D.4. RMSD trends for the TLR7 backbone atoms (green), the backbone atoms of the binding pocket residues (blue), the side chain atoms of the binding pocket (cyan) and the heavy atoms of the CR42-24 molecule (green), calculated between the initial phase of restraint release and the end of the 10 ns-long production simulation (10.2 ns of total time).

The results of the molecular mechanics-generalized Born surface area (MM/GBSA) calculations are reported in Table A.1. The total binding energies of the two complexes (TLR7 bound to R-837, from the previous study, and CR42-24) are comparable, with the CR42-24 complex showing a slightly more favorable energy. The highest differences observed in the single terms were for the van der Waals interactions ( $\sim 11$  kcal/mol less for the CR42-24 complex) and the polar solvation terms, calculated with the GB method. For these terms, the change in free energy for CR42-24 binding to the TLR7 was considerably larger ( $\sim 14$  kcal/mol more for the CR42-24 complex) than the R-837 one. This result suggests a binding of CR42-24 to the TLR7 dimer,

which involves stronger hydrophobic interactions than the R-837 binding. Also, CR42-24 showed a more favorable energy value for the polar ligand-receptor interactions (~6 kcal/mol less than R-837). The two non-polar terms were comparable.

Table D.1. Results from the MM/GBSA calculations over the last 5 ns of MD production trajectories of the R-837-TLR7 and CR42-24-TLR7 complexes. The total binding energies of the two compounds are in a comparable range, with the more favorable van der Waals and electrostatic contributions of the CR42-24-TLR7 complex being partially compensated by a larger, unfavorable polar desolvation term, when compared to R-837. The total values reported here do not include the entropic terms. The value of the standard deviation is reported in brackets for each term.

	<b>MM/GBSA <math>\Delta G</math> (kcal/mol)</b>				
<b>Compound</b>	<b>Total</b>	<b>Van der Waals</b>	<b>Electrostatic</b>	<b>Polar solvation</b>	<b>Non-polar solvation</b>
R-837	-31.081 (2.531)	-35.028 (2.600)	-18.251 (2.912)	27.331 (2.653)	-5.133 (0.216)
CR42-24	-36.055 (3.090)	-46.841 (2.719)	-24.050 (5.273)	41.089 (5.105)	-6.252 (0.239)

In addition, we performed the decomposition of the MM/GBSA energies among the residues constituting the binding site of TLR7, in order to quantitatively assess the pattern of intermolecular interactions established between CR42-24 and the TLR7 dimer, and to compare it to the pattern observed for the R-837-TLR7 complex. The decompositions of the total binding energy between the residues within 5 Å of the docking poses, as well as the van der Waals and electrostatic terms, are reported in Figure D.5. The residues aL557, aI585, bF351 and bF408 showed the most similar contribution to the binding energies of the two compounds.

Interestingly, these residues showed similar contributions also for the hydrophobic terms, whereas the electrostatic terms were either negligible or different between the two complexes. Also, CR42-24 showed a considerably more favorable total interaction energy with bY356 and bK432. For the first residue, the hydrophobic contribution was clearly more negative than the R-837 complex. Regarding bK432, we observed a very favorable electrostatic interaction energy for CR42-24 (~-11 kcal/mol), not present in the R-837-TLR7 complex. As already mentioned, this was due to a polar interaction involving the carbonyl group of the C ring of CR42-24, which established a stable, strong hydrogen bond with the side chain of bK432 during the simulation. This compensated the unfavorable total and electrostatic interactions between CR42-24 and aD555, which was reported as a key residue for the binding of imidazoquinoline and adenine derivatives to the TLR7 binding pocket [271], as in the case of R-837. By visually inspecting the MD production trajectory, this unfavorable contribution was observed to be due to a salt bridge established between aD555 and bK432, which locked the location of the former in close proximity to the carbonyl group of the CR42-24 C ring and far from the amine of the acetamide group (Figure D.10).

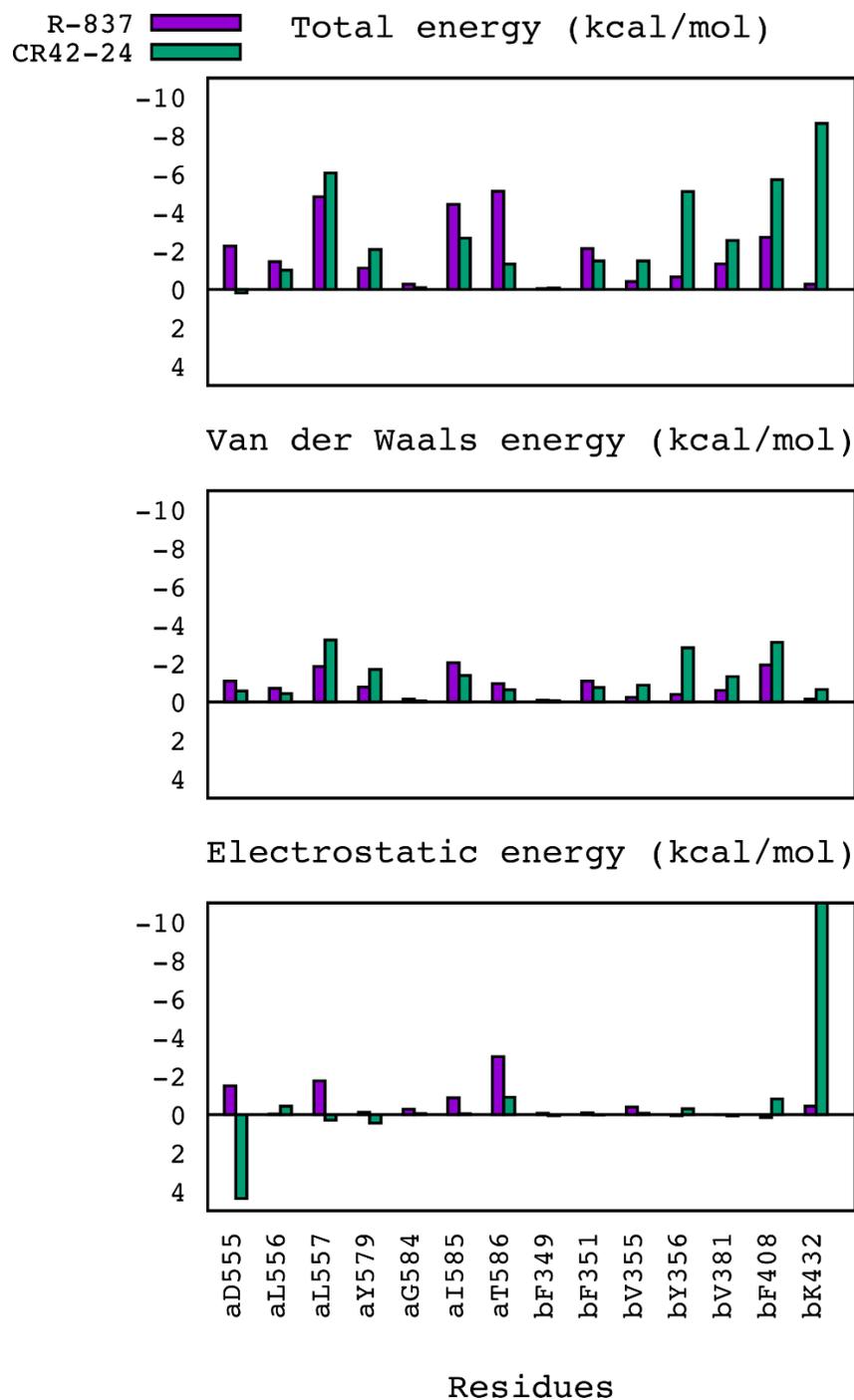


Figure D.5. Decomposition of total, van der Waals and electrostatic interaction energies between the residues of the TLR7 pocket and the two compounds R-837 (violet) and CR42-24 (green). Similar patterns were observed for the total and van der Waals interactions of the two compounds with the TLR7 dimer, while the electrostatic interaction pattern was different. The unfavorable interaction of CR42-24 with the key residue aD555 was compensated by a

favorable, electrostatic-driven interaction with bK432, which was not observed for other TLR7 ligands in our previous study.

#### D.3.4 Experiments

The results from the type II Anova test performed over the subset of cancer genes, including TLR7, are reported in the volcano plot in Figure D.6. The cell lines carrying the mutated TLR7 gene were more sensitive to CR42-24, as it is clearly visible from the reported negative  $IC_{50}$  shift.

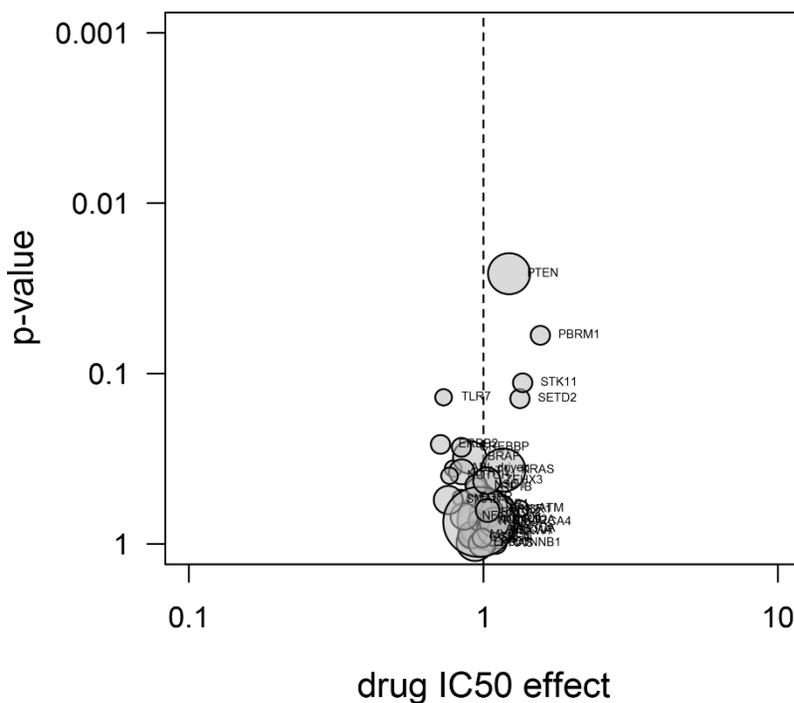


Figure D.6. Volcano plot reporting the Anova analysis performed over thirty-nine cancer genes, including TLR7. The x-axis represents the factor of  $IC_{50}$  shift, while the y-axis reports the p-values, i.e. the confidence level for genetic association of mutations in a particular gene with a  $IC_{50}$  shift. The areas in the circles are proportional to the number of mutants present in the cell panel (each reported gene was present in at least three Oncolines<sup>TM</sup> cell lines).

#### D.4 Materials and Methods

#### D.4.1 Model Building and Validation

The model for the endosomal domain of the human TLR7 heterodimer was the same described in our previous work [271]. Briefly, the x-ray structure of the human TLR8 dimer (Protein Data Bank [32] ID 3W3J) in complex with the TLR7/8 dual agonist CL097 [276] was selected as template for a homology modeling procedure in Molecular Operating Environment (MOE) 2013 [181]. *MOE-Align* [277] was used to align the two amino acid sequences. The modeling step was performed using the automated homology modeling tool of MOE, choosing the final model based on the Generalized Born Integration/Volume Integral (GB/VI) scoring function [221]. The Amber ff12SB force field parameters [223] were used for the protein. The two CL097 molecules present in the template were conserved as environment for the Induced Fit during the modeling protocol, assigning them the Extended Hückel Theory (EHT) parameters [278]. The protonation states of the model residues were assigned using *MOE Protonate3D* [222]. The ff12SB parameters were used for the protein, while the General Amber Force Field (GAFF) [192] parameters were assigned to the two ligand molecules using Antechamber [193]. An octahedral box of TIP3P water molecules [279] was built around the model, with a buffer of at least 12 Å between any complex's atom and the edge of the box. The system was neutralized by adding Cl<sup>-</sup> ions. The TLR7 dimer was then minimized using the Amber *pmemd.cuda* engine [191]. The cutoff for long range interactions was set to 9 Å. Water molecule and ion positions were relaxed using a sequence of 1,000 steps of steepest descent and 1,000 steps of conjugate gradients minimization methods, keeping the protein and ligand atoms highly restrained through a harmonic potential with a force constant of 500 kcal/mol/Å<sup>2</sup>. Then a second minimization procedure was performed, with 2,000 and 3,000 steps of steepest descent and conjugate gradients methods, respectively. For this step, we restrained the protein backbone and the ligand heavy

atoms only using a harmonic potential with a force constant of 4 kcal/mol/Å<sup>2</sup>. The optimized structure was then evaluated using the ERRAT [280] and PROCHECK [281] programs.

#### *D.4.2 Ligand Preparation and Molecular Docking*

The CR42-24 structure was prepared using the *LigPrep* module from Schrödinger's Maestro [177], in order to obtain possible protonation states, tautomers and low-energy ring conformations at pH 5.5, which corresponds to the endosomal compartment value. The OPLS 2005 force field [282] was used for this process.

The docking was performed using *MOE Dock*. The target structure was the R-837-TLR7 dimer complex obtained and validated in our previous work [271]. The choice of the target structure was made based on the evidence of a possible interaction of R-837 with the colchicine-binding site of tubulin [269], where CR42-24 is also likely to bind. The site to which performs the simulations was defined as the one occupied by the R-837 atoms, a cleft present at the interface of the two TLR7 monomers around residue aD555. The placement method was chosen as Triangle Matcher [229], returning thirty poses based on the London dG scoring [221]. The receptor was kept rigid for the refinement step, where five poses were retained based on the GBVI/WSA scoring function [221]. The five final poses were visually inspected.

#### *D.4.3 Molecular Dynamics Simulations*

The parameters required for the simulation of the small molecule were assigned from the General Amber Force Field (GAFF) [192] using Antechamber [193], whereas the parameters for the protein were chosen from the Amber ff12SB force field. Again, we inserted the system in an octahedral box of TIP3P water molecules, with at a minimum of 12 Å of buffer between any atom of the complex and the edge of the box. The system was neutralized by adding the correct

number of Cl<sup>-</sup> ions. A value of 9 Å was set as cutoff for long range interactions. Using the *pmemd.cuda* engine, the system were initially minimized with the same minimization protocol used for the homology model. They were then heated up to 310 K in a time of 100 ps, using the Langevin thermostat with a time collision frequency of 2 ps and an integration time step of 0.5 fs. The target temperature was gradually reached after 150,000 steps and maintained for the last 50,000 of the heating protocol. A restraint of 2 kcal/mol/Å<sup>2</sup> was imposed to the protein backbone and on the heavy atoms of the small molecule during this step, and constant volume conditions were applied. The restraints were slowly removed in four phases of 50 ps and constant pressure conditions, in each of which the force constant was reduced of 0.5 kcal/mol/Å<sup>2</sup>. The production NPT simulation was then run for 10 ns with a time step of 2 fs at 1 atm, using the Berendsen barostat. During the production run, the bonds which involve hydrogens were frozen with the SHAKE algorithm [227]. AmberTools12 ptraj [223] was used to compute the RMSDs of the protein, the binding site and the small molecule from the initial restraint release phase to the end of the production simulation, using the first snapshot as reference. The RMSDs were also calculated over the production trajectory, using the average structure of the last 5 ns of simulation as reference.

#### *D.4.4 MM/GBSA Binding Energy Calculations*

One of the goals of the MD simulations of the complex was to obtain an ensemble of ligand-receptor configurations over which it is possible to calculate an average binding energy. Similar to our previous work [271], we adopted the MM/GBSA method [248,283] as implemented in the AmberTools12 script *MMPBSA.py* [197]. The binding energy of a ligand-receptor complex was calculated as

$$\Delta G_{bind,solv} = \Delta G_{MM,vac} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,protein}) - T\Delta S \quad (D.1)$$

where  $\Delta G_{MM,vac}$  is the sum of electrostatic and van der Waals interactions established between the protein and the ligand. The  $T\Delta S$  term models the change in conformational entropy due to the binding, which was neglected in our study. Indeed, we did not require any ranking of the compounds in this case. When using the MM/GBSA method, the solvation terms are given by the equation

$$\Delta G_{solv} = \Delta G_{solv,polar} + \Delta G_{solv,npolar} \quad (D.2)$$

where the polar contribution of the solvent is calculated solving the Generalized Born equation [283]. All the  $\Delta S$  in the above equations are calculated as the contribution value of the complex minus the values of the ligand and the protein alone, where each contribution is computed as the averaged value over a single MD trajectory where each time the complex is maintained, the protein is removed or the ligand is removed, respectively. The salt concentration was set to 0.15 M. The igb flag was set to 5 [248]. The hydrophobic contribution was calculated as

$$\Delta G_{solv,npolar} = \gamma \cdot SASA \quad (D.3)$$

where  $\gamma$  (surface tension) was set to  $0.005 \text{ kcal/mol/\AA}^2$ , and  $SASA$  is the solvent-accessible surface area (SASA) calculated using the linear combinations of pairwise overlaps (LCPO) model method [249]. Per-residue decomposition of the computed binding energies was performed among the residues within  $5 \text{ \AA}$  of the docked pose of CR42-24. We performed all the energy calculations considering the last 5 ns of production MD simulation of the system, using 250 snapshots extracted at regular intervals of 20 ps from the trajectory.

#### *D.4.5 Cell Preparation*

All cell lines have been licensed from the American Type Culture Collection (ATCC) Manassas, Virginia (US). Master and working cell banks (MCB and WCB) were prepared by sub-culturing in ATCC-recommended media and freezing according to ATCC recommended protocols ([www.atcc.org](http://www.atcc.org)). Cell line stocks for the assays were prepared from the WCB. The MCB, WCBs and assay stocks were prepared within respectively 3, 6 and 10 passages of the ATCC vial. The complete list of the cell lines used in this study, as well as their genetic status, is reported in Table D.2.

#### *D.4.6 Compound Preparation*

Solid powders of reference compounds were stored as indicated by supplier. Compounds were weighed on a calibrated balance and dissolved in 100% DMSO. DMSO samples were stored at room temperature. At the day of the experiment, the compound stock was diluted in 3.16-fold steps in 100% DMSO to obtain a 9-point dilution series. This was further diluted 31.6 times in 20 mM sterile HEPES buffer pH 7.4. A volume of 5  $\mu$ l was transferred to the cells to generate the test concentration range in duplicate. The final DMSO concentration during incubation was 0.4% in all wells. If a compound showed very potent activity, the testing range was expanded to ensure a full dose-response curve could be measured in duplicate. If a compound can only be dissolved in an aqueous solution, the recommended buffer is used instead of 100% DMSO.

#### *D.4.7 Cell Proliferation Assay*

Cells were diluted in the corresponding ATCC recommended medium and dispensed in a 384-well plate, depending on the cell line used, at a density of 200 - 6400 cells per well in 45  $\mu$ l medium. For each used cell line the optimal cell density was used. The margins of the plate were filled with phosphate-buffered saline. Plated cells were incubated in a humidified atmosphere of

5% CO<sub>2</sub> at 37 °C. After 24 hours, 5 µl of compound dilution was added and plates were further incubated. At  $t=end$ , 24 µl of ATPlite 1Step™ (PerkinElmer, Waltham, MA, USA) solution was added to each well, and subsequently shaken for 2 minutes. After 10 minutes of incubation in the dark, the luminescence was recorded on an Envision multimode reader (PerkinElmer).

#### D.4.8 Data Analysis

IC<sub>50</sub> values were calculated by non-linear regression using IDBS XLfit 5. The percentage growth after incubation until  $t=end$  was calculated as follows:

$$\%_{growth} = 100\% \times \left( \frac{luminescence_{t=end}}{luminescence_{untreated,t=end}} \right) \quad (D.4)$$

This was fitted to the <sup>10</sup>log compound concentration ( $conc$ ) by a 4-parameter logistics curve:

$$\%_{growth,fit} = bottom + \left( \frac{top - bottom}{1 + 10^{(logIC_{50} - conc) \times hill}} \right) \quad (D.5)$$

where  $hill$  is the Hill coefficient, and  $bottom$  and  $top$  the asymptotic minimum and maximum cell growth that the compound allows in that assay.

#### D.4.9 Curve Fitting

Curves calculated automatically by the software were adjusted manually according to the following protocol: the curve bottom was fixed at 0% when the calculated curve had a bottom below zero. The  $hill$  was fixed on -6 when the software calculated a lower value. Curves were invalidated when the F-test value for fitting quality was >1.5 or when the compound was inactive (<20% maximal effect), in which cases curves were removed from the graphs. When a curve had a biphasic character, it was fitted on the most potent IC<sub>50</sub>. Incidentally, when technical failures were likely, concentration points were knocked out. This is always shown in the dose-

response graphs. The maximal effect (Max effect) was calculated as 100% (signal of untreated cells) minus the curve bottom when the dose-response curve was completely determined for more than 85%. A dose-response curve is considered 100% complete when the data points at the highest concentrations reach the curve bottom. If the completeness was smaller than 85%, Max effect was calculated as 100% minus the average of the lowest signal. In cases where the bottom of the curve was locked on 0%, the maximal effect was always calculated as 100% minus the growth inhibition at the highest concentration.

#### *D.4.10 Cell Genetics*

The mutation status of cell lines was established from a combination of public and proprietary data. Based on public data (COSMIC Cancer Genome Project, version 80) [284], we collected mutations, amplifications and deletions in established cancer driver genes that occur in Oncolines™ [285,286]. For further validation, a selection of twenty-three cancer genes were sequenced by nitrogen regulatory protein (NTRC) by targeted and full exome sequencing directly from the cell lines used in Oncolines™. As an extra filter, genetic changes were required to be observed with a preset frequency in patient tumor samples in COSMIC, depending on the type of genetic alteration. This discards sporadic, non-cancer-causing mutations. Cell lines were classified as having a wild type or a mutated genotype, where mutated means: at least one allele changed by point mutation, insertion, deletion, amplification or copy number variation. Analysis was performed on genes that were mutated in at least three different Oncolines™ cell lines (ninety-eight genes in total).

#### *D.4.11 Drug Sensitivity*

The relation between measured  $IC_{50}$ s (calculated as explained in section 3.8) and cell line genetics was determined as follow: a larger subset of the most commonly occurring and best-

known cancer genes (thirty-eight in total, plus TLR7) was analyzed with type II ANOVA analysis in the statistical program R. The results were displayed in a volcano plot. For more information about the Oncolines™ methods, refer to [287].

## D.5 Conclusions

TLR7 is an endosomal, trans-membrane, homodimeric receptor which is involved in the innate immune response, by initiating the signaling cascade upon binding to PAMPs and DAMPs. Its main activators are viral ssRNA and nucleoside analogues. R-837 (also called imiquimod) was the first small molecule to be FDA-approved for the treatment of gout and skin cancer, in its cream form. R-837 was shown to bind also to the colchicine-binding site of  $\beta$ -tubulin, inhibiting microtubule polymerization in an analogous way to colchicine and its derivatives. Due to this suggested binding site promiscuity, which leads to off-target interactions of R-837, we explored the possibility of CR42-24, a potent colchicine analogue, to bind to the R-837 binding site in TLR7. We employed a series of computational tools to identify and fully characterize the potential binding pose of CR42-24 in such site. Also, we tested the compound in a comprehensive cancer cell line panel, which includes TLR7 mutations among other common cancer mutated genes, to observe the possible TLR7-mediated anti-cancer effect of this compound.

Molecular docking simulations targeting the TLR7 site revealed a binding pose of CR42-24 which was consistent with the R-837 one, showing conserved hydrophobic contact patterns. In addition, the CR42-24 predicted pose showed similarities with the binding pose assumed by colchicine in  $\beta$ -tubulin. In particular, the A and C rings of CR42-24 were involved in van der Waals contacts with two hydrophobic zones within the TLR7 site, in an analogue way of the A

and C rings of colchicine in tubulin. MD simulations of the CR42-24-TLR7 complex revealed a stable binding conformation. Also, the averaged binding affinity calculated with the MM/GBSA method for CR42-24 was more favorable than the R-837 one. Finally, an important hydrogen bond between the carbonyl group of the C ring of CR42-24 and residue bK432 was identified by decomposing the free energy of binding between the pocket residues. Regarding the testing in cancer cell lines, we observed a considerable negative  $IC_{50}$  shift in terms of cell growth, in cancer cells carrying the TLR7 mutated gene, when compared to wild type cell lines. This suggests a possible interaction between CR42-24 and TLR7, and a TLR7-mediated anti-cancer effect of the compound.

In conclusion, we fully characterized the possible binding mode of CR42-24 to the TLR7's R-837 binding site through a sophisticated computational workflow. Taking into account the following factors: 1) the similarity between the predicted binding mode of CR42-24 and the R-837 one, which was assessed experimentally, 2) the similarities between the CR42-24 pose in TLR7 and the colchicine pose in tubulin and 3) the negative  $IC_{50}$  shift observed for TLR7-mutated cell lines, we suggest a possible off-target interaction between our colchicine derivative, CR42-24, and the TLR7 dimer, with a possible anti-cancer effect. This result is interesting not only to extend the known off-target interactions of colchicine derivatives and provide a potential mechanistic elucidation of immune system suppression by these compounds, but also to develop novel TLR7-targeting small molecules based on the CR42-24 scaffold. The results of the biological assay used in this study were not conclusive in terms of identifying the effect of CR42-24 to TLR7 as agonistic or inhibitory. However, the similarity between the binding poses of CR42-24 and R-837 may suggest a similar activating effect upon binding to the receptor. In addition, R-837 is known to inhibit cell growth especially in skin cancer [54], and we observed a

higher sensitivity of TLR7-mutated cell lines to the CR42-24 compound. Future works will include more detailed biochemical and biological testing to assess the direct binding of CR42-24 to TLR7 and the effect that this has on the receptor, i.e. if CR42-24 has an activating or inhibiting effect to the signaling process initiated by TLR7 upon binding to it.

## D.6 Supplementary Materials

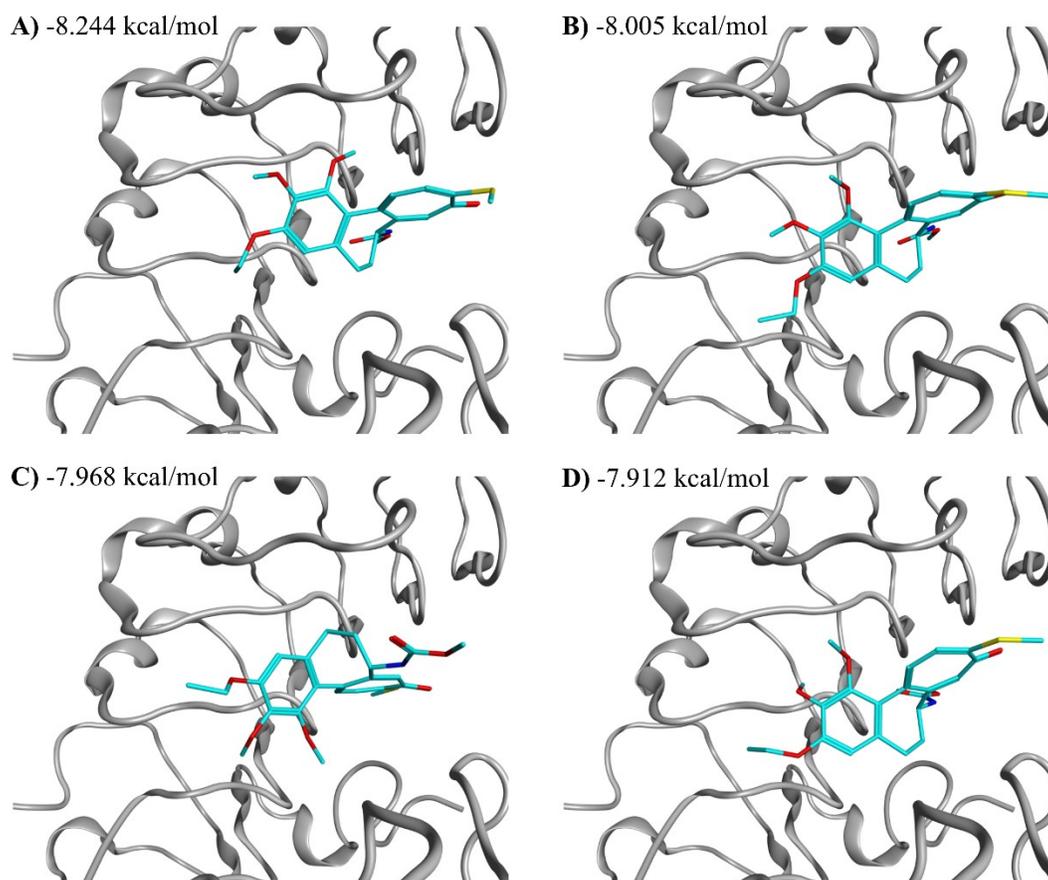


Figure D.7. The additional four poses of CR42-24 obtained with molecular docking to the R-837 binding site of TLR7. Computed binding energies are also reported.

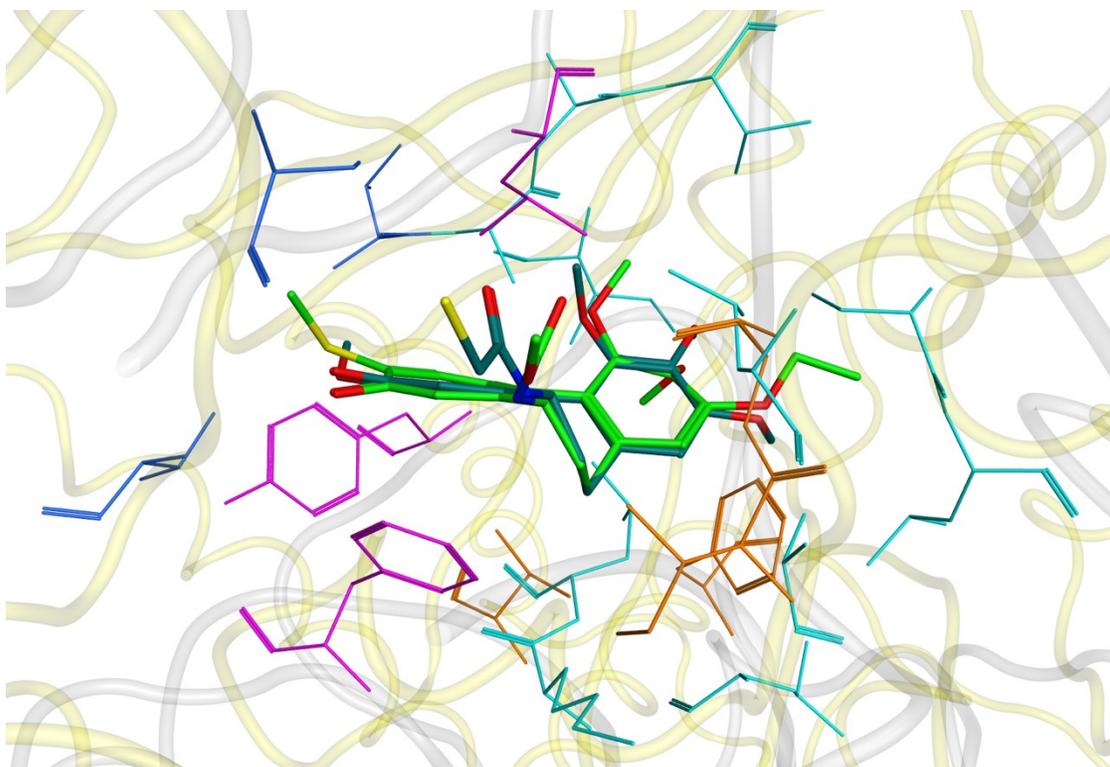


Figure D.8. Superposition of the crystallographic pose of DAMA-colchicine (dark green carbons) in tubulin from pdb ID 1SA0 and the CR42-24 docking pose in TLR7 (carbons in light green). The first hydrophobic cleft of the tubulin binding site, interacting with the C ring of the DAMA-colchicine, is reported in blue sticks. The corresponding first hydrophobic zone of TLR7 is reported in purple sticks. The second hydrophobic zone is reported in cyan sticks for tubulin and orange sticks for TLR7. Tubulin backbone is reported in yellow ribbons, TLR7 backbone in grey ribbons.

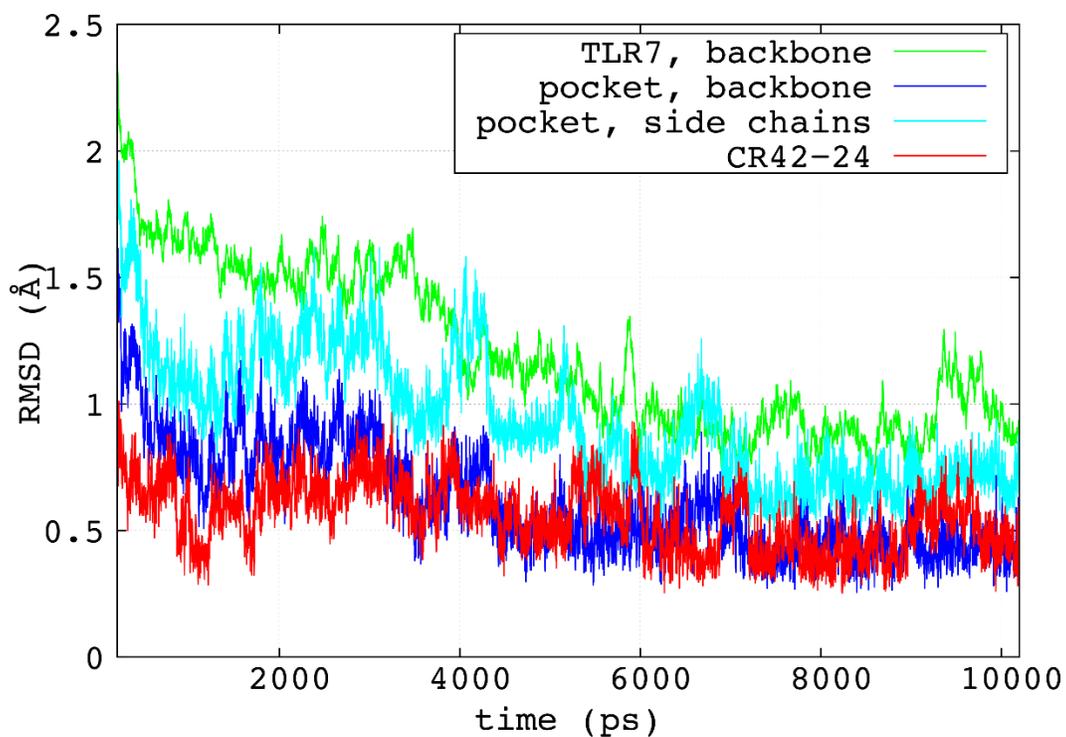


Figure D.9. RMSD trends of the MD production simulation for the TLR7 backbone atoms (green), the backbone atoms of the binding pocket (blue), the side chain atoms of the binding pocket (cyan) and the heavy atoms of CR42-24 (red), calculated using the average structure of the last 5 ns as reference. The highest fluctuations are observed for the side chain atoms of the binding pocket, as they adapted to the bound conformation upon ligand binding.

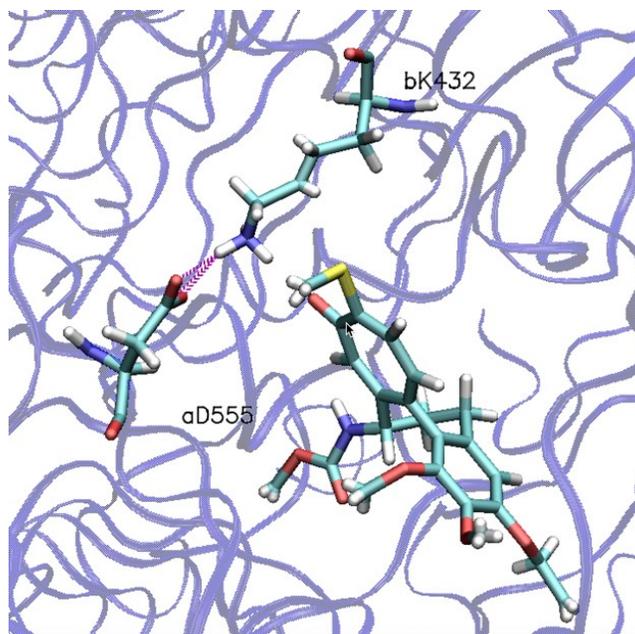


Figure D.10. Salt bridge (in purple) established between aD555 and bK432, keeping the former far from the amine of the acetamide group of CR42-24.

Table D.2. Cell lines in the Oncolines™ panel used in this study and their genetic status. Indicated in the second column are the genes that are either mutated, part of a translocation or that have altered copy numbers (*italic*) in that cell line.

Cell line	Cancer genes
5637	ERBB2, RB1, TP53
769-P	No coding alterations found in 38 cancer driver genes
786-O	<i>CDKN2A</i> , PTEN, TP53
A-172	<i>CDKN2A</i> , <i>PTEN</i>
A-204	ABL-driven
A-427	<i>CDKN2A</i> , CTNNB1, KRAS
A-498	<i>CDKN2A</i> , SETD2
A-549	<i>CDKN2A</i> , KRAS, SMARCA4, STK11

A-704	PBRM1, TP53
A375	BRAF, CDKN2A, <i>CDKN2A</i>
A388	<i>EGFR</i> , NOTCH1, PIK3CA, TP53
ACHN	<i>CDKN2A</i> , PBRM1
AN3 CA	CCND1, FBXW7, NSD1, PIK3R1, PTEN, SETD2, TLR7, TP53, ZFHX3
AsPC-1	CDKN2A, FBXW7, KRAS, TP53
AU-565	<i>ERBB2</i> , <i>MYC</i> , SMAD4, TLR7, TP53
BT-20	<i>CDKN2A</i> , <i>EGFR</i> , PIK3CA, RB1, SPEN, TP53
BT-549	PTEN, <i>RBI</i>
BxPC-3	BRAF, <i>CDKN2A</i> , EP300, <i>SMAD4</i> , TP53
C-33 A	ARID1A, FBXW7, PIK3CA, PTEN, SMARCA4, TP53, ZFHX3
CAL 27	CDKN2A, SMAD4, TP53
CCF-STTG1	PTEN
CCRF-CEM	CCND1, <i>CDKN2A</i> , FBXW7, KRAS, NRAS, NSD1, <i>PTEN</i> , TLR7, TP53
COLO 205	APC, BRAF, <i>SMAD4</i>
COLO 829	BRAF, CDKN2A, <i>PTEN</i>
Daoy	<i>CDKN2A</i> , <i>NF1</i> , PIK3R1, TP53
DB	TP53
DLD-1	APC, EP300, KRAS, NCOR1, PIK3CA, SMARCA4, TP53
DoTc2 4510	BRCA2, FAT1, NSD1
DU 145	CDKN2A, CREBBP, FAT1, <i>KRAS</i> , RB1, STK11, XIRP2
DU4475	APC, BRAF, <i>RBI</i>
ES-2	BRAF, <i>CCND1</i> , TP53

FaDu	<i>CCND1</i> , <i>FAT1</i> , <i>SMAD4</i> , <i>TP53</i>
G-361	<i>BRAF</i> , <i>CDKN2A</i> , <i>STK11</i>
HCT-15	<i>APC</i> , <i>BRCA2</i> , <i>EP300</i> , <i>KRAS</i> , <i>NCOR1</i> , <i>PIK3CA</i> , <i>TP53</i> , <i>ZFH3</i>
HCT 116	<i>BRCA2</i> , <i>CDKN2A</i> , <i>CHD4</i> , <i>CTNNB1</i> , <i>EP300</i> , <i>KRAS</i> , <i>NCOR1</i> , <i>NF1</i> , <i>PIK3CA</i> , <i>SMARCA4</i> , <i>XIRP2</i>
HL-60	<i>CDKN2A</i> , <i>MYC</i> , <i>NRAS</i>
Hs 578T	<i>MYC</i> , <i>PIK3R1</i> , <i>TP53</i>
Hs 746T	<i>TP53</i>
Hs 766T	<i>ARID1A</i> , <i>KRAS</i> , <i>LRP1B</i> , <i>SMAD4</i>
HT	<i>EP300</i> , <i>SPEN</i> , <i>TP53</i>
HT-1080	<i>CDKN2A</i> , <i>NRAS</i>
HuTu 80	<i>CTNNB1</i> , <i>SMARCA4</i>
J82	<i>ERBB2</i> , <i>PTEN</i>
JAR	No coding alterations found in 38 cancer driver genes
Jurkat E6.1	<i>APC</i> , <i>ARID1A</i> , <i>CDKN2A</i> , <i>CREBBP</i> , <i>FAT1</i> , <i>FBXW7</i> , <i>PTEN</i> , <i>SMARCA4</i> , <i>TP53</i>
K-562	<i>ABL-driven</i> , <i>CDKN2A</i> , <i>TP53</i>
KATO III	<i>NOTCH1</i>
KG-1	No coding alterations found in 38 cancer driver genes
KLE	<i>CCNE1</i> , <i>FBXW7</i> , <i>KRAS</i> , <i>TP53</i>
KU812	<i>ABL-driven</i> , <i>TP53</i>
LNCaP FGC	<i>ATM</i> , <i>BRCA2</i> , <i>PIK3R1</i> , <i>PTEN</i> , <i>SETD2</i> , <i>XIRP2</i>
LoVo	<i>APC</i> , <i>ARID1A</i> , <i>FBXW7</i> , <i>KRAS</i> , <i>SPEN</i>

LS 174T	ARID1A, CTNNB1, KRAS, PBRM1, PIK3CA, TLR7, ZFH3
LS411N	APC, ARID1A, BRAF, EP300, FBXW7, PTEN, TP53, XIRP2
MCF7	<i>CDKN2A</i> , EP300, PIK3CA
MeWo	CDKN2A, NF1, TP53, XIRP2
MG-63	<i>CDKN2A</i> , <i>MYC</i>
MIA PaCa-2	<i>CDKN2A</i> , KRAS, TP53
MOLT-4	<i>CDKN2A</i> , CHD4, EP300, NOTCH1, NRAS, PTEN, SMARCA4, TP53
NCCIT	PTEN, TP53
NCI-H460	ARID1A, <i>CDKN2A</i> , KRAS, <i>LRP1B</i> , <i>MYC</i> , PIK3CA, STK11
NCI-H661	<i>CCNE1</i> , TP53
NCI-H82	<i>MYC</i>
OVCAR-3	<i>CCNE1</i> , <i>LRP1B</i> , TP53
PA-1	NRAS
PC-3	<i>MYC</i> , <i>PTEN</i> , TP53
PFSK-1	TP53
RD	NF1, NRAS, TP53
RKO	ARID1A, BRAF, BRCA2, EP300, FAT1, NCOR1, NF1, NSD1, PIK3CA, ZFH3
RL	EP300
RL95-2	ARID1A, ATM, BRCA2, EP300, PIK3R1, PTEN, SMARCA4, SPEN, TP53, ZFH3
RPMI-7951	BRAF, <i>PTEN</i> , TP53
RS4-11	<i>CDKN2A</i> , SMARCA4

RT4	<i>CDKN2A, SMAD4</i>
SHP-77	KRAS, TP53
SJCRH30	TP53
SK-N-AS	NRAS, TP53, XIRP2
SK-N-FI	<i>NF1</i> , TP53
SNU-5	<i>ARID1A</i> , CDKN2A
SNU-C2B	ERBB2, KRAS, SMARCA4, TP53
SR	<i>CDKN2A, ZFH3</i>
SU-DHL-1	CREBBP, TP53
SU-DHL-6	CREBBP, EP300, TP53
SUP-T1	<i>CDKN2A, CHD4, PBRM1, PIK3CA, TP53</i>
SW48	BRCA2, CTNNB1, EGFR, EP300, FBXW7, NCOR1, SETD2, ZFH3
SW480	APC, KRAS, TP53
SW620	APC, EP300, KRAS, <i>KRAS, MYC</i> , TP53
SW626	APC, KRAS, SMAD4
SW837	APC, FBXW7, KRAS, TP53
SW872	BRAF, CDKN2A, <i>PTEN</i>
SW900	<i>CDKN2A, FAT1, KRAS, KRAS, NF1, NSD1</i> , TP53
SW948	APC, ATM, KRAS, PIK3CA, SMAD4, TP53
SW982	BRAF, <i>CDKN2A</i>
T24	EP300, FAT1, TP53
T98G	<i>CDNK2A</i> , TP53
TCCSUP	ARID1A, PIK3CA, <i>RBI</i> , TP53

THP-1	ARID1A, <i>CDKN2A</i> , NRAS, <i>PTEN</i> , TP53
TT	TP53
U-118 MG	<i>CDNK2A</i> , TP53
U-2 OS	<i>LRP1B</i>
U-87 MG	<i>CDNK2A</i> , NF1
VA-ES-BJ	<i>CDNK2A</i>