







## Article

# Identification of a Chemotherapeutic Lead Molecule for the Potential Disruption of the FAM72A-UNG2 Interaction to Interfere with Genome Stability, Centromere Formation, and Genome Editing

Senthil Renganathan <sup>1,†</sup>, Subrata Pramanik <sup>2,†</sup>, Rajasekaran Ekambaram <sup>3</sup>, Arne Kutzner <sup>4</sup>, Pok-Son Kim <sup>5</sup>, and Klaus Heese <sup>6,\*</sup>

- <sup>1</sup> Department of Bioinformatics, Marudupandiyar College, Thanjavur 613403, India; rsenthil@mpi.edu.in  
<sup>2</sup> Department of Biology, Life Science Centre, School of Science and Technology, Örebro University, 701-82 Örebro, Sweden; subrata.pramanik@oru.se  
<sup>3</sup> Department of Chemistry, V.S.B. Engineering College, Karur 639111, India; ekambaramrajasekaran@gmail.com  
<sup>4</sup> Department of Information Systems, College of Engineering, Hanyang University, Seoul 133-791, Korea; kutzner@hanyang.ac.kr  
<sup>5</sup> Department of Information Security, Cryptology, and Mathematics, Kookmin University, Seoul 136-702, Korea; pskim@kookmin.ac.kr  
<sup>6</sup> Graduate School of Biomedical Science and Engineering, Hanyang University, Seoul 133-791, Korea  
\* Correspondence: klaus@hanyang.ac.kr  
† Both authors contributed equally to the study.



**Citation:** Renganathan, S.; Pramanik, S.; Ekambaram, R.; Kutzner, A.; Kim, P.-S.; Heese, K. Identification of a Chemotherapeutic Lead Molecule for the Potential Disruption of the FAM72A-UNG2 Interaction to Interfere with Genome Stability, Centromere Formation, and Genome Editing. *Cancers* **2021**, *13*, 5870. <https://doi.org/10.3390/cancers13225870>

Academic Editors: Klaus Podar, Shahab Uddin, Mohd W. Nasser and Martin Sattler

Received: 9 October 2021  
Accepted: 20 November 2021  
Published: 22 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Pivotal factors that contribute to tumorigenesis were subjected to analysis by molecular modeling. In particular, the FAM72A-UNG2 protein–protein interaction was modeled to predict a potential solution for the treatment of cancer. We screened chemical libraries to identify withaferin B as a lead molecule capable of interfering with the FAM72A-UNG2 interaction, thus opening new therapeutic avenues for cancer.

**Abstract:** Family with sequence similarity 72 A (FAM72A) is a pivotal mitosis-promoting factor that is highly expressed in various types of cancer. FAM72A interacts with the uracil-DNA glycosylase UNG2 to prevent mutagenesis by eliminating uracil from DNA molecules through cleaving the N-glycosylic bond and initiating the base excision repair pathway, thus maintaining genome integrity. In the present study, we determined a specific FAM72A-UNG2 heterodimer protein interaction using molecular docking and dynamics. In addition, through in silico screening, we identified withaferin B as a molecule that can specifically prevent the FAM72A-UNG2 interaction by blocking its cell signaling pathways. Our results provide an excellent basis for possible therapeutic approaches in the clinical treatment of cancer.

**Keywords:** cell cycle; centromere; DNA repair; proliferation

## 1. Introduction

Genomic uracil bases may occur from cytosine deamination or the misincorporation of dUMP residues during DNA replication [1]. The uracil-DNA glycosylase UNG physiologically functions in the base excision repair (BER) mechanism of the cell in order to replace uracil from U/G mispairs with cytosine, thus preventing genomic mutations [2–6]. It excises unwanted genomic uracil bases using an extrahelical base recognition mechanism, thus preventing possible C-to-T transition mutations that eventually arise from cytosine deamination [7–9]. The resulting apurinic/apyrimidinic site (AP-site) is considered one of the most common DNA lesions in the genome, and a persistent AP-site can have ad-

verse consequences, as the lesion disrupts many DNA and RNA transactions and leads to cytotoxic strand breaks, mutations, and other forms of genomic instability [1,10,11].

Human UNG exists in two different isoforms, mitochondrial UNG1 and nuclear UNG2, that are both encoded from the same single 13.5-kb nuclear UNG gene as a result of two separate promoters and alternative splicing [12–14]. While UNG1 and UNG2 share a common conserved catalytic domain, they contain differing N-termini sequences responsible for differential subcellular localization. Amino acid (AA) residues 1–92 make up the N-terminus of UNG2; they contain a nuclear localization signal of positively charged residue clusters (K and R residues), rendering UNG2 as the primary uracil-DNA glycosylase enzyme in the nucleus [14,15]. Interestingly, in the absence of binding partners, the N-terminal region is, for the most part, without a fixed structure [2,16–18]. UNG2 is rapidly recruited to sites of DNA damage where its N-terminus can interact with its catalytic site (which binds to uracil) and with chromatin. UNG2 colocalizes with CENP-A at centromeres and other sites of DNA damage in proliferating cells, thus implying that it is also required for chromosome segregation during mitosis [19].

Family with sequence similarity 72 A (FAM72A) is a novel gene expressed in the brain hippocampus area in proliferating neural stem cells, particularly during the G2/M-phase of the cell cycle [20,21]. Most strikingly, humans have four paralogs (FAM72 A-D), whereas all other species express just one ortholog [22,23]. Under pathophysiological conditions, FAM72A is also expressed in various proliferating cancer cells [24,25]. Notably, FAM72A interacts with UNG2 [26,27]. This denotes that the cellular role of FAM72A is as a cooperative partner for genomic BER in order to ensure genome integrity and impede the formation of cancer. Recent data show that decreased levels of FAM72A lead to hyperphysiological UNG2 levels, an increased uracil correction, and, thus, error-free DNA repair. In contrast, the binding of FAM72A with UNG2 antagonizes UNG2 activity and causes UNG2 degradation in B cells, leading to increased levels of genome-wide deoxyuracils and, therefore, mediating increased levels of U•G mispairs that engage in mutagenic mismatch repair, promoting the error-prone processing of activation-induced cytidine deaminase (AID)-induced deoxyuracils [27,28]. Thus, FAM72A bridges BER and mismatch repair in order to modulate antibody diversification during B cell and antibody maturation [27,28]. Overall, an increased FAM72A level could lead to reduced UNG2 levels and could thus shift the balance of appropriate mutagenic DNA repair, therefore making the cells more susceptible to mutations, with possible effects on tumor development [24,25,27–29].

To understand the possible disruption of the FAM72A-UNG2 interaction, the current investigation conducted an *in silico* prediction of FAM72A-UNG2 heterodimer–protein interaction and the identification of potential chemicals that interfere with the FAM72A-UNG2 heterodimer protein activity for the potential treatment of cancer.

## 2. Materials and Methods

### 2.1. Data Collection—Sequence and Structure Details

The crystallographic three-dimensional (3D) protein structure of the FAM72A protein (from Protein Data Bank (PDB) data [30]) and UNG2 protein (Gene ID: 7374, isoform-2: NP\_550433.1, 313 AAs; 1AKZ\_A PDB model; DOI: 10.2210/pdb1akz/pdb [31]) were retrieved from the PDB [32] with a resolution of 1.55 Å. Compound structures were downloaded from the PubChem database [33] as described previously [30,34,35].

### 2.2. Homology Modeling and Protein Structure Validation of UNG2 by Modeller, I-TASSER and AlphaFold

FAM72A 3D protein structure was used from previously designed PDB data [30]. Unfortunately, no suitable UNG2 3D protein structure was available, and the N-terminal UNG2 residues were called an intrinsically disordered region. Thus, the UNG2 protein sequence was checked in the National Center for Biotechnology Information (NCBI) and PDB, and the closest suggested template for a UNG2 3D protein structure model was

selected. The UNG2 3D peptide sequence was based on the UNG2 protein sequence (Gene ID: 7374, isoform-2: NP\_550433.1, 313 AAs; Uniprot-ID: P13051) and the 1AKZ\_A PDB model (DOI: 10.2210/pdb1akz/pdb) [31] was selected as the template. The obtained template for the N-terminal UNG2 3D peptide structure model (AA 1-313) was then forwarded for UNG2 3D peptide structure modeling with I-TASSER [36,37] and Modeller v9.20 [38,39] software, and Chimera software was used as a graphical interface as described previously [34,35,39]. For comparison, we also applied the UNG2 protein sequence to the state-of-the-art machine learning method, AlphaFold (<https://alphafold.ebi.ac.uk/>), accessed on 3 November 2021) [40,41].

### 2.3. Intrinsically Disordered Region in UNG2 (AA 1–92)

The N-terminal regulatory region of UNG2 is described as an intrinsically disordered region by several groups [16,17,42]. In continuation, JPred4 [43] was used for additional secondary structure prediction. A search algorithm and sequence weighting method against the given UNG2 protein sequence was applied with default parameters (hidden Markov model (HMM) and BLOSUM filter). The UNG2 AA composition was calculated to identify and justify the AA residues promoting structured or unstructured regions in the UNG2 protein. Modeled structures were visualized using Chimera software as the graphical interface to check the core, rim, and buried regions, as described previously [30,34,35].

### 2.4. Molecular Docking of FAM72A Protein and UNG2 Peptide (AA 1–45) by HPEPDOCK

Docking interactions for the FAM72A protein with the modeled UNG2 peptide (AA 1–45) as a heterodimer were performed by HPEPDOCK (default parameters were applied) [44]. The FAM72A monomer was exported as a PDB file, whereas the UNG2 peptide was submitted as a FASTA-formatted AA sequence (AA 1–45), and MODPEP and MDOCK were applied for the fine adjustments of FAM72A-UNG2 interactions [44]. Modeled structures were visualized using Chimera software as a graphical interface, as described previously [30,34,35].

### 2.5. Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) Calculation

The molecular mechanics/generalized Born surface area (MM/GBSA) free energy decomposition per AA residue in protein–protein interactions was predicted on the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer [45]. Hawkdock calculated the free binding energy for the key AA residues in the protein–protein interfaces based on the Amber16 force field [46].

### 2.6. Carbon Distribution (CARd) Analysis

The protein carbon distribution (CARd) analysis was performed to validate the specific FAM72A-UNG2 interaction sites using our recently described algorithm [35,47].

### 2.7. Amino Acid-Specific Mutations in FAM72A Protein-UNG2 (AA 1–45) Peptide Heterodimer (FAM72A F104A, F104R, F104N, F104G, and F104S)

A site-specific mutagenic approach was enabled to check the hot spot residues in the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer interaction. AA modifications in FAM72A (F104A, F104R, F104N, F104G, and F104S) and the effect on the conformational stability of the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer complex were investigated by the BIOVIA Discovery Studio software (Dassault Systems; Waltham, MA, USA) as described previously [48,49].

### 2.8. Molecular Dynamics Simulation by GROMACS

The starting coordinates of FAM72A and UNG2 were taken from the modeled structures, as described in Section 2.2. GROMOS96 43a1 force field was used in this study. Hydrogens were added to the protein molecules by using the pdb2gmw application in GROMACS (2019.2). Then, protein molecules were placed in a cubic simulation box (de-

fault parameters). A simple point charge water model was used for the solvation simulation box. To neutralize the system,  $\text{Na}^+$  and  $\text{Cl}^-$  ions were added to the simulation box. The structure was relaxed through a process called energy minimization (EM). Subsequently, energy-minimized structures were used for the system equilibration performed under constant NVT and NPT (number, volume, temperature, and pressure) ensembles. The production run was carried out using the NPT ensemble for 50 ns with a time step of 2 fs at a constant temperature of 300 K and 1 bar pressure. Simulation trajectories were visualized using Visual Molecular Dynamics (VMD) 1.9.4a42. Analysis of features, including root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration (Rg), were performed using GROMACS (2019) tools [50]. RMSD is the most commonly used metric, in which, the root mean square distance between corresponding residues is calculated [51,52]. Since the RMSD can weight the distances between all residue pairs equally, a small number of local structural deviations can result in high RMSD, even when the global topologies of compared structures are similar. The average RMSD of randomly related proteins depends on the length of compared structures, rendering the absolute magnitude of RMSD meaningless [53]. The RMSD computes the average distance between the backbone atoms of starting structure (reference structure) with simulated structures (frame by frame) when superimposed. The RMSF computes fluctuations (standard deviation) of atomic positions of each AA residue in the trajectory. The RMSD and RMSF were calculated for 50 ns using GROMACS (2019) [50] for the FAM72A-UNG2 heterodimer with the UNG2 (AA 1–45) peptide and FAM72A protein (wildtype [wt], W125A, W125R, F104A, F104R, F104G, F104N, and F104S) [50]. The trajectory files resulting from the molecular dynamics simulation were computed as RMSD, RMSF, and Rg, and were plotted by Grace (“GRaphing, Advanced Computation and Exploration of data”; a WYSIWYG 2D graph plotting tool for Unix operating systems).

### 2.9. Virtual Screening for Lead Molecule Identification against FAM72A Protein and UNG2 (AA 1–45) Peptide Mono/Heterodimer

The MTiOpenScreen [54], along with Autodock vina [55] and iScreen [56] databases, were applied for chemical library virtual screening and de novo drug design. Additionally, pharmacophore analysis was performed using the PharmMapper server [57] to detect the basic pharmacophore group of select chemicals for docking analysis. Further application of COACH, TM-SITE, S-SITE, COFACTOR, and ConCavity approaches [30,34,35] provided potential ligand-binding sites of the 3D FAM72A-UNG protein heterodimer or the FAM72A protein monomer structure model (with refinement by ModRefiner [58]), with potential molecules based on a BioLiP [59].

### 2.10. Molecular Docking of FAM72A Protein and UNG2 (AA 1–45) Peptide Heterodimer and FAM72A Monomer with Small Chemical Molecules

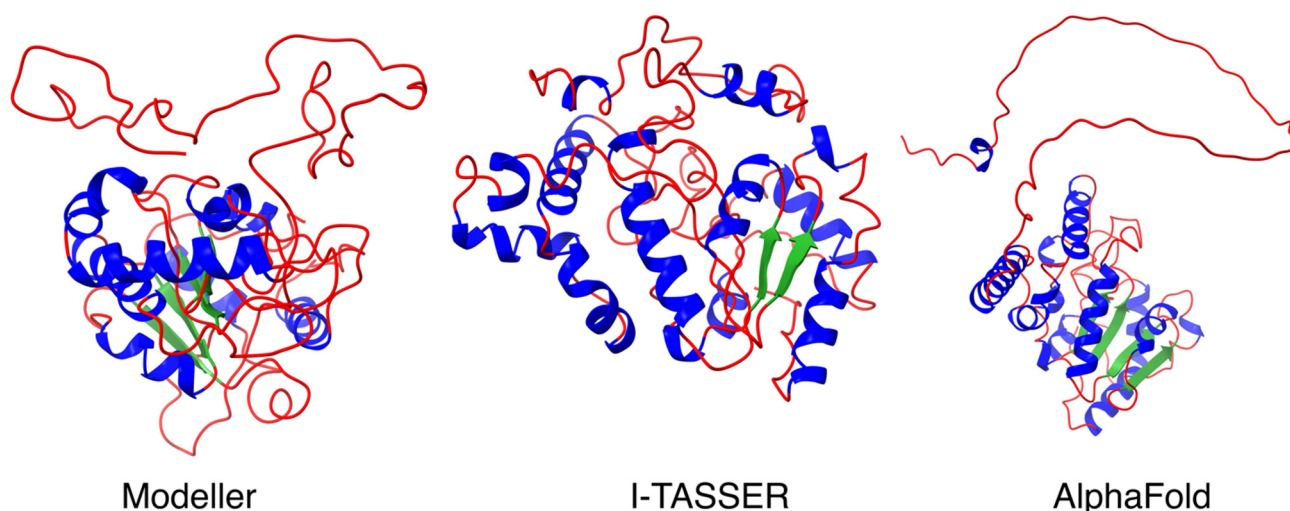
Further molecular docking studies have been undertaken in order to gain further insights into the possible FAM72A-UNG2 binding interference by molecules newly identified by protein–ligand binding site prediction and to understand their mechanisms of interaction. Identified molecules obtained by the virtual screening were docked onto the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer and/or FAM72A monomer using Schrödinger to depict binding mode and calculate binding energy [60,61]. FAM72A and UNG2 3D protein structures were prepared using the protein preparation wizard panel of the Schrödinger software package (Schrödinger, LLC, New York, NY, USA). The 3D protein crystal structures of FAM72A and UNG2 were transferred to the workspace and pre-processed, and missing loops were filled [62]. Water molecules were removed from the ligand-binding domain. H-bonds were optimized using the hydrogen bond optimizer, and the FAM72A and UNG2 protein structures were moved to the minimization process to minimize the energy in order to confirm the lowest energy conformational structure [63]. Default parameters were used for the molecular docking process, applying the Glide 4.0 XP extra precision module of the Schrödinger software package (Schrödinger, LLC). The binding affinity with FAM72A was calculated for each chemical compound and ranked by

the scoring function. Modeled structures were visualized using the same Chimera software as the graphical interface, as described previously [30,34,35,62].

### 3. Results and Discussion

#### 3.1. Homology Modeling and Protein Structure Validation of UNG2 by Modeller, I-TASSER and AlphaFold

The UNG2 protein can be functionally divided into two domains: an N-terminal regulatory region (AA 1–92) and a C-terminal catalytic region (AA 93–313). The disordered N-terminal region has been identified as interacting with several proteins, including proliferating cell nuclear antigen (PCNA) and replication protein A (RPA) (both found at DNA replication forks), as well as with FAM72A [2,16–18,26,27]. To further enlighten the FAM72A-UNG2 interaction, we investigated the N-terminus of UNG2 (AA 1–92), applying the Modeller, I-TASSER, and AlphaFold protein structure prediction analysis programs. Our predicted comparative structure analysis revealed a long protruding thread-like disordered N-terminal loop (AA 1–92) required to gather and catch more targets for molecular crowding (Figure 1).

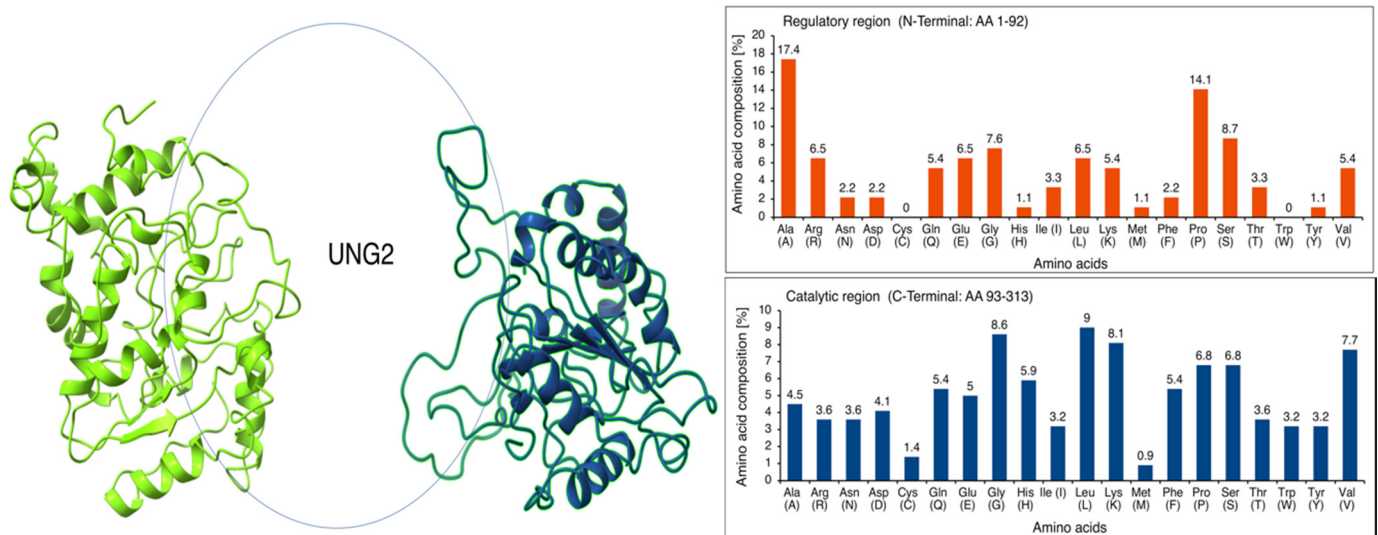


**Figure 1.** Predicted comparative 3D protein structure analysis of full-length UNG2 protein, including N-terminal UNG2 (AA 1–92) (**Left**): Modeller modeled full-length UNG2 3D protein structure output (using human UNG2 sequence from UniProtKB (P13051) and PDB template 1AKZ) implying a conformation suitable for N-terminal UNG2 protein binding to chromatin; (**Center**): I-TASSER modeled full-length UNG2 3D protein structure output implying a conformation suitable for the binding of the N-terminal UNG2 protein to its catalytic site; (**Right**): AlphaFold output of full length UNG2 (using human UNG2 sequence from UniProtKB (P13051)). All three protein structure prediction approaches revealed the N-terminal loop as a protruding thread-like disordered region suitable for interactions with multiple protein binding partners and for molecular crowding. Secondary structure color code:  $\alpha$ -helix in blue,  $\beta$ -sheet in green, and loop in red.

#### 3.2. Intrinsically Disordered Region in N-Terminal UNG2 (AA 1–92)

Intrinsically disordered proteins (IDPs) execute various functions in all kinds of cellular processes [64–66]. The N-terminal regulatory domain of UNG2 possesses such an unstructured regional IDP motif (AA 1–92). In general, the absence of a hydrophobic core is probably the reason for an unstructured region, whereby hydrophilic AAs may dominate in number. AAs have been classified as order-promoting (Asn, Cys, Ile, Leu, Phe, Trp, Tyr, and Val) and disorder-promoting (Ala, Arg, Gln, Glu, Gly, Lys, Pro, and Ser) [65,67]. The calculated AA composition revealed an abundance of disorder-promoting Ala, Pro, Ser, and Gly in the N-terminal regulatory region, which is very flexible in moving and orienting the N-terminus of UNG2. In contrast, the catalytic region is full of hydrophobic order-promoting AA residues, including Leu, Val, Ile, and Trp. Evidently, Cys and Trp are not available for a foldable secondary structure formation, as linker residues form in the N-terminus of the UNG2. The plot shows the biases in AA composition at N-terminal residues

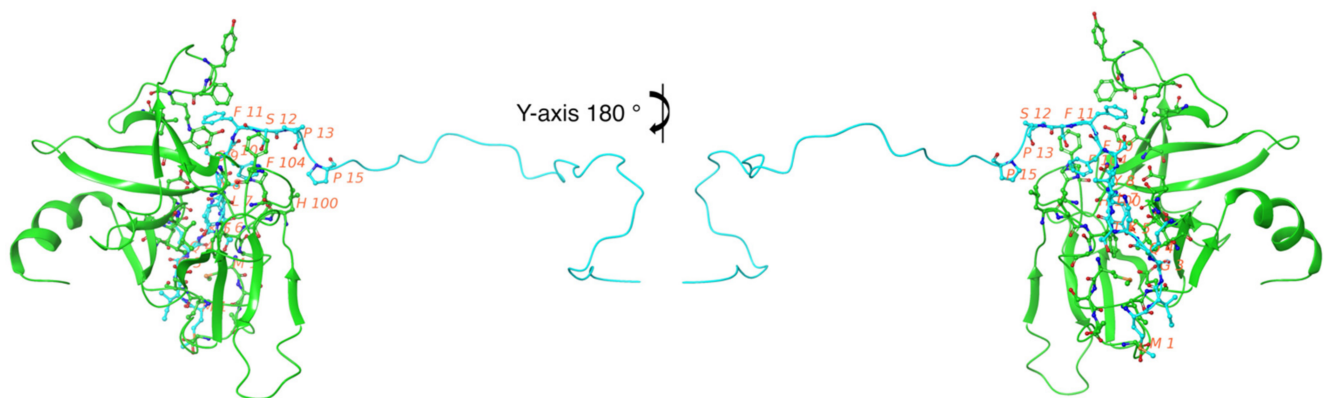
and explains the importance of sulfur-containing AAs and tryptophan at hydrophobic cores for protein rigidity. The absence of AAs, such as Cys and Trp, has been recognized by evolutionary studies about protein plasticity and disordered protein regions in previous studies (Figure 2) [66,68–70].



**Figure 2.** (Left): The intrinsically disordered proteins (IDP) regions in the I-TASSER- and Modeler-based predicted models is circled. Left-hand side: I-TASSER output; right-hand side: Modeler output. A similar N-terminal IDP was obtained with AlphaFold (Figure 1). (Right): AA composition was calculated for AA residue enrichment in promoting ordered or disordered regions in UNG2. The AA composition of UNG2 analysis clearly shows the abundance of disorder-promoting AAs Ala, Pro, Ser, and Gly within the N-terminal regulatory region, while the catalytic region is enriched by hydrophobic order-promoting AA residues, including Leu, Val, Ile, and Trp. Evidently, Cys and Trp are not available for a foldable secondary structure formation as linker residues form in the N-terminus of UNG2. The plot shows biases in AA composition at N-terminal residues and explains the importance of sulfur-containing AAs and tryptophan at hydrophobic cores for protein rigidity. The absence of AAs, such as Cys and Trp, has been recognized by evolutionary studies of protein plasticity and disordered protein regions in previous studies.

### 3.3. FAM72A-UNG2 Interaction and Molecular Docking Study of FAM72A Protein and UNG2 (AA 1–45) Peptide by HPEPDOCK

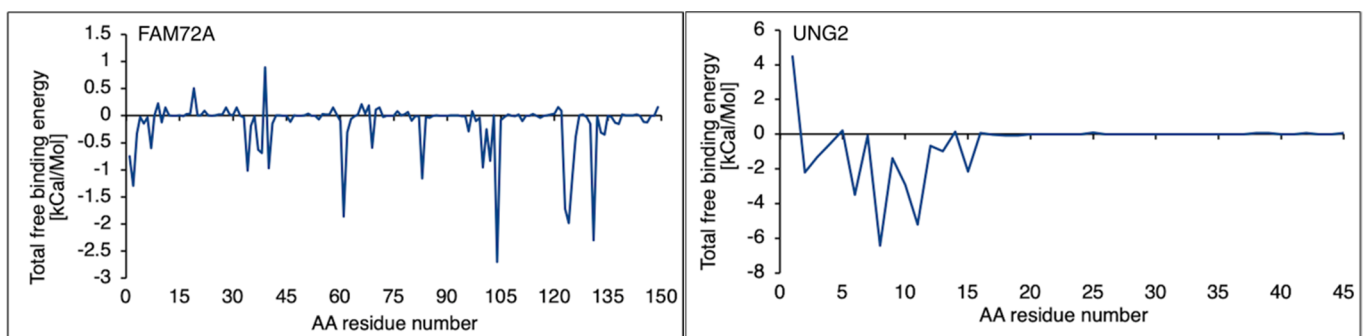
Our molecular docking study evaluated the molecular forces responsible for specific biomolecular FAM72A-UNG2 interactions. The FAM72A monomer was exported as a PDB file (FAM72A 3D protein structure was used from previously designed PDB data [30]), whereas the UNG2 (AA 1–45) peptide was submitted as a FASTA-formatted AA sequence (AA 1–45). The UNG2 (AA 1–45) peptide was used because these AAs appeared to be the pivotal interacting AAs [2,16–18,26]. The docked structure was analyzed for specific AAs contributing to the FAM72A protein and UNG2 peptide interactions. Mostly, electrostatic forces dominate other forces. Hydrogen bonding is less preferred for the FAM72A-UNG2 association because the side chain (UNG2; chain-B) moves along the diagonal portion of the other FAM72A protein (chain-A). Due to the lack of a proper quaternary structure in the N-terminal UNG2 region, the UNG2 peptide prefers surface AA residues (such as AAs 5, 7, 8, 10, 11, 12, 13, and 15) in order to make connections and to increase the prevalence rate of interactions (Figure 3).



**Figure 3.** FAM72A-UNG2 (AA 1–45) interactions. FAM72A = chain-A is shown in green and UNG2 (AA 1–45) = chain-B is shown in blue. The image on the right-hand side is a  $y$ -axis  $180^\circ$  of the FAM72A-UNG2 interaction to illustrate the interaction clearly. Key interacting AA residues are labeled in red. The prevalence rate of interface residues on the FAM72A-UNG2 interaction has been depicted by molecular rendering (cartoon model). Hydrophobicity is the major phenomenon of the FAM72A protein and UNG2 peptide interaction.

#### 3.4. Free Binding Energy Prediction on FAM72A Protein and UNG2 (AA 1–45) Peptide Heterodimer

An MM/GBSA prediction is imposed on the free binding energy calculation in the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer. The MM/GBSA analysis offered a breakthrough regarding catalytic AAs in the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer. AA residue–residue contacts in the FAM72A-UNG2 heterodimer were calculated in terms of free binding energy, considering van der Waals forces, electrostatic energy, solvent accessible surface areas, and polar and non-polar energies. Binding affinity values were plotted as a scatter plot indicating key AA residues in FAM72A (2, 37–40, 61, 83, 101–104, 123–126, and 131, respectively), while AA F104 was identified as pivotal AA. In proportion, the UNG2 (AA 1–45) peptide has been verified with only a few AAs accountable for binding contribution (AAs 2, 5, 8, 11, and 15, respectively) (Figure 4).

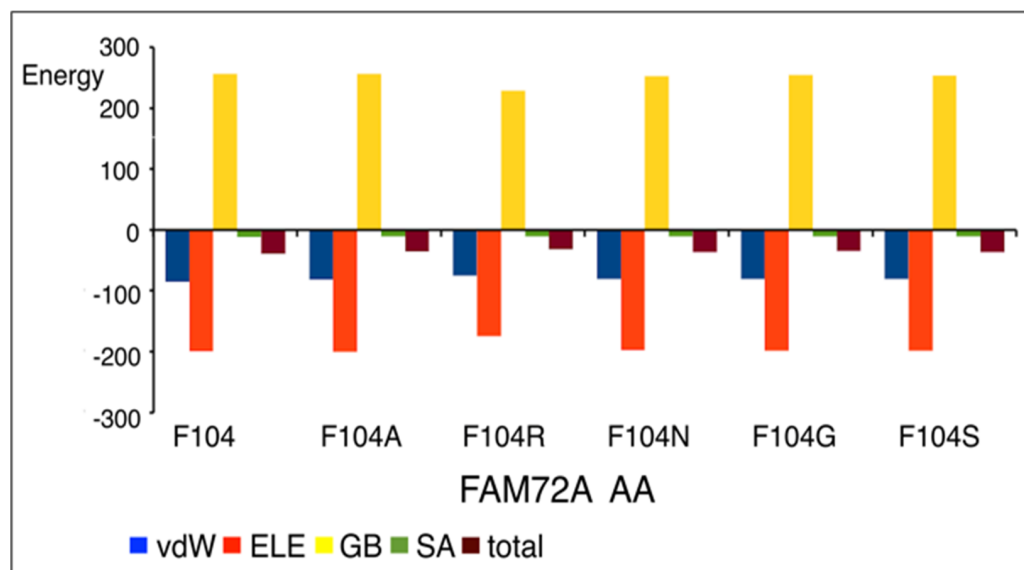


**Figure 4.** Free binding energy was calculated for the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer interaction. Specific AA residue–residue contact energy calculated the FAM72A-UNG2 (AA 1–45) interaction. Key AA residues were identified in FAM72A (2, 37–40, 61, 83, 101–104, 123–126, and 131, respectively), while AA F104 was identified as a pivotal AA with the highest binding energy contacting the Ser12 and Pro13 AAs of UNG. The UNG2 (AA 1–45) peptide has been verified with only a few AAs accountable for binding contribution (AAs 2, 5, 8, 11, and 15, respectively).

#### 3.5. AA-Specific Mutations in the FWMF Motif (AA 101–104) of FAM72A Affecting the FAM72A Protein and UNG2 (AA 1–45) Peptide Heterodimer Binding

Site-directed specific mutations in the FWMF motif (AA 101–104) of FAM72A were used (F104A, F104R, F104N, F104G, and F104S) to evaluate the rigidity and flexibility of

the interface in the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer binding (Figures 5 and 6).



**Figure 5.** FWMF motif (AA 101–104)-specific mutation-dependent FAM72A protein and UNG2 (AA 1–45) peptide heterodimer free binding energy (energy in [kcal/mol]). FAM72A mutations at AA F104 phenylalanine (F104 → 104A, F104R, F104N, F104G, and F104S). A difference in total free binding energy was observed individually at the FWMF motif (AA 101–104)-specific mutations, with F104R mutation showing the highest effect on energy-dependent conformational change affecting FAM72A-UNG2 binding. The FAM72A protein and UNG2 (AA 1–45) peptide heterodimer free binding energy was calculated by the MM/GBSA approach. ELE, electrostatic; GB, polar accessible (a grid-based surface GB model is used for estimating the polar component of solvation free energy, coupled with water and atomic radii); SA, solvent area; total, sum of total energy; vdW, van der Waals; MM/GBSA, molecular mechanics/generalized Born surface area.

We modeled the FAM72A and UNG2 (AA 1–45) peptide and the interaction of FAM72A with the UNG2 (AA 1–45) peptide. The FWMF motif (AA 101–104) appears to be key for the FAM72A protein structure and its binding to UNG2 (Figure 6). A mutation in the FWMF motif from wt F104 to F104R had the largest effect, turning the binding energy from negative (strong binding/hydrophobic core) to positive (strong binding/hydrogen bonding).

### 3.6. Molecular Dynamic Simulation by GROMACS Validates AA-specific Mutations in the FWMF Motif (AA 101–104) of FAM72A Affecting FAM72A-UNG2 Heterodimer Binding

Since phenylalanine F104 appeared to be the key AA within the FWMF motif (AA 101–104) at the interface of the FAM72A-UNG2 interaction, we further investigated the effect of FAM72A mutations at wt AA F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S) within the FWMF motif (AA 101–104) on the dynamic nature of FAM72A-UNG2 binding. Dynamic conformation changes in FAM72A-UNG2 (AA 1–45) binding were simulated by GROMACS and plotted by Grace (Figures 7–9).



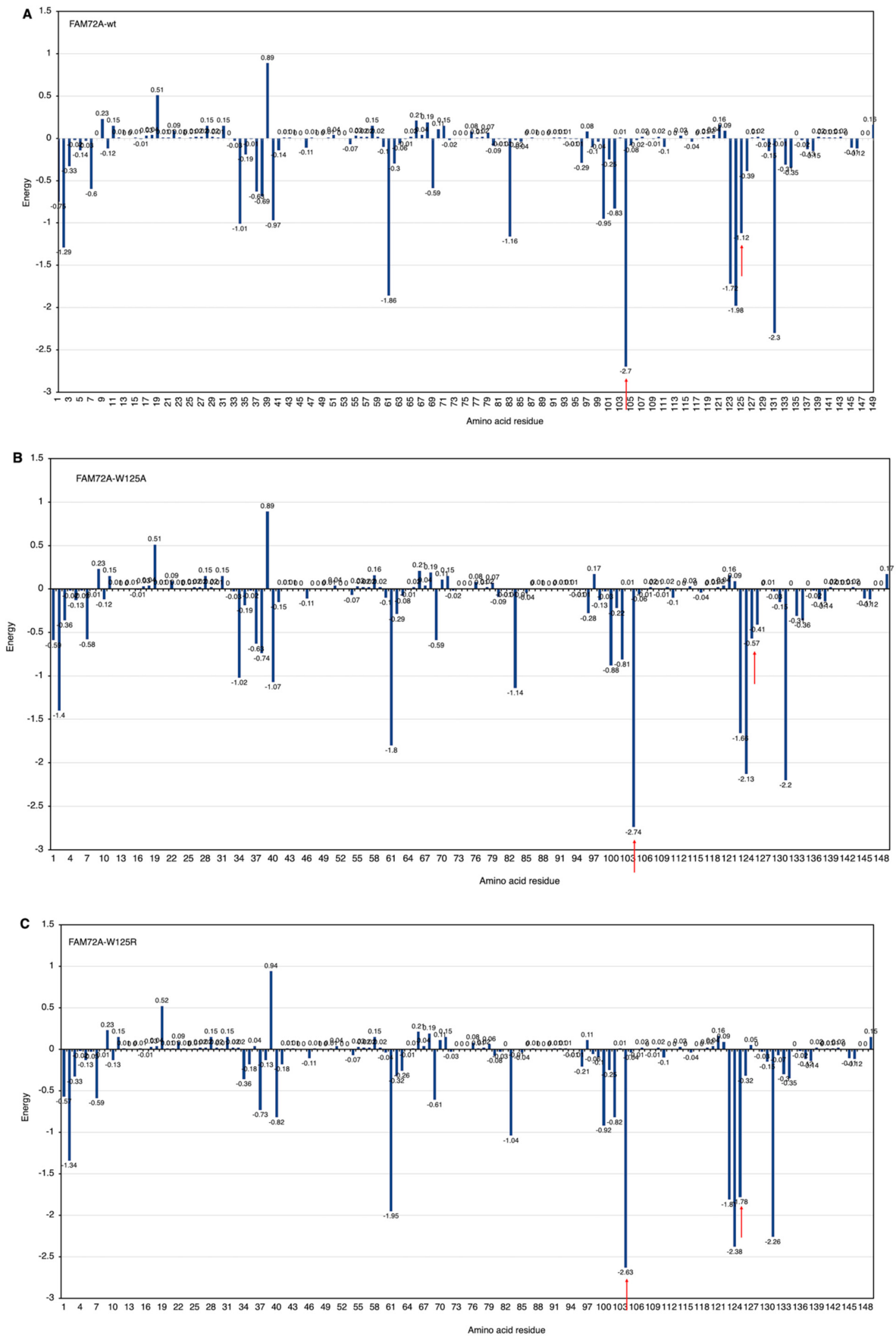


Figure 6. Cont.

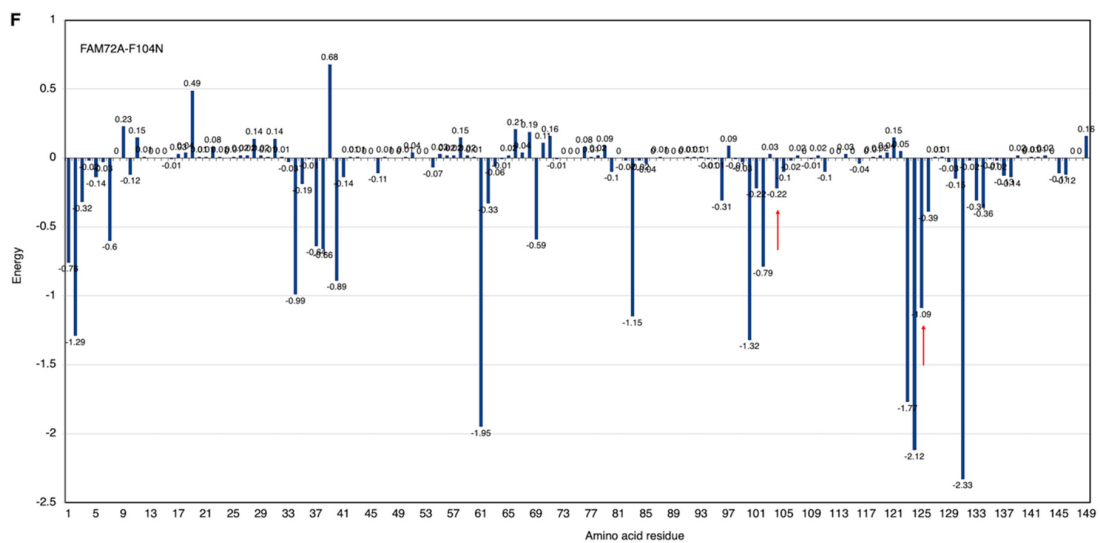
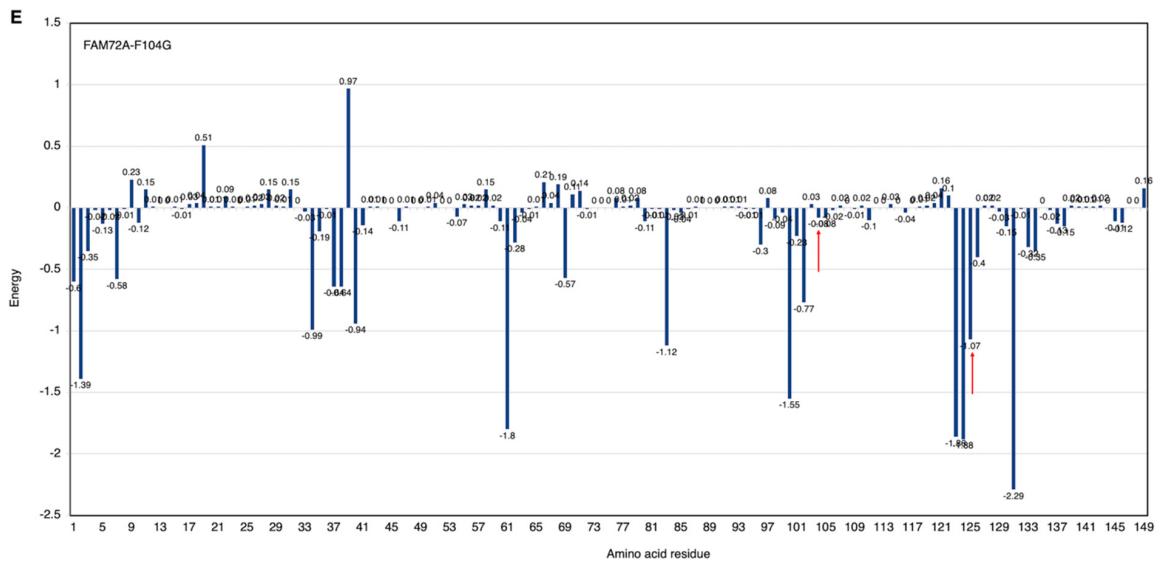
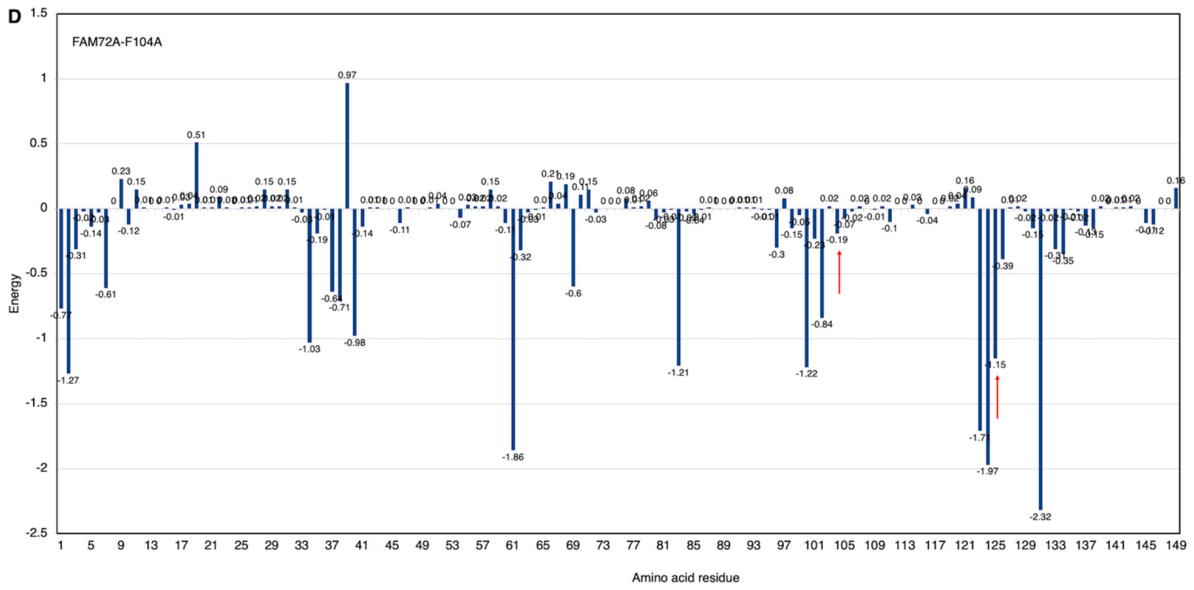
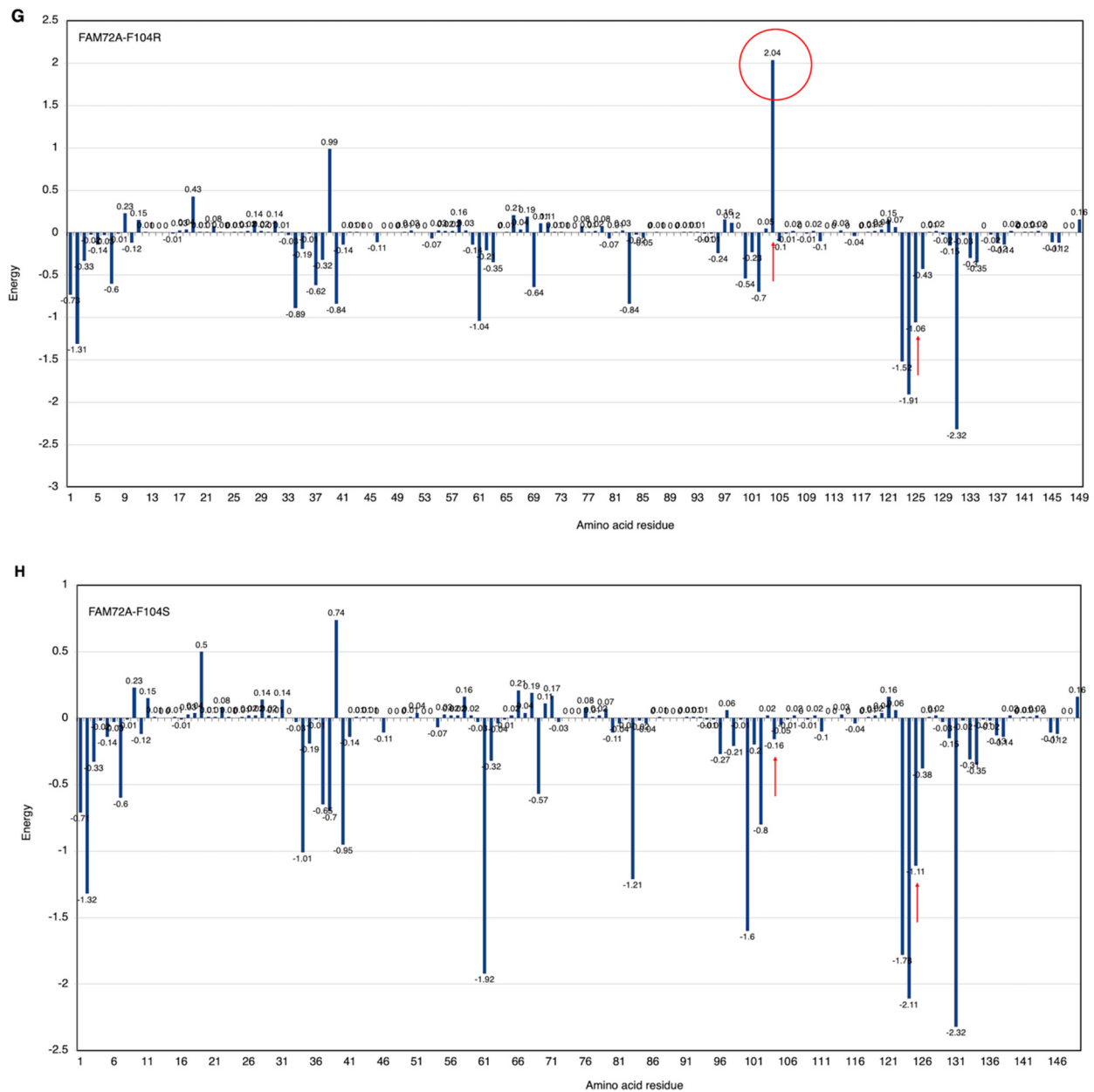
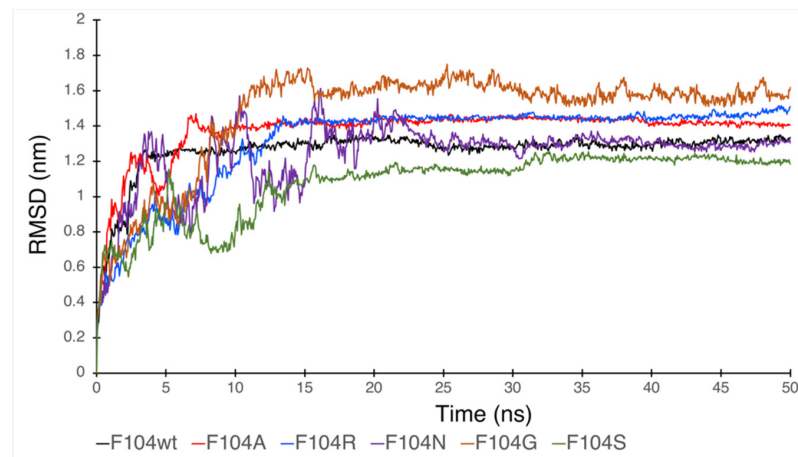


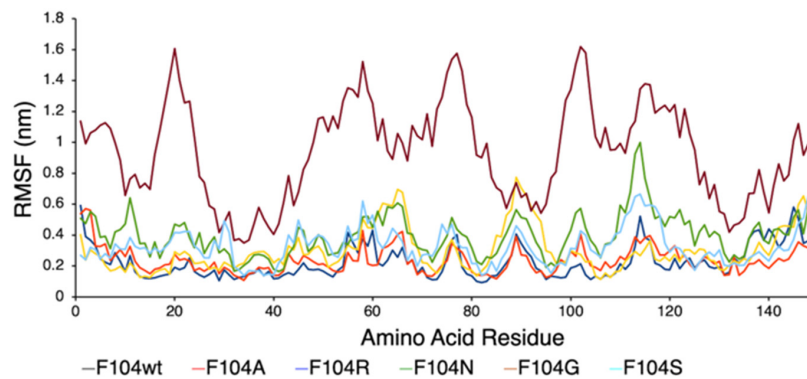
Figure 6. Cont.



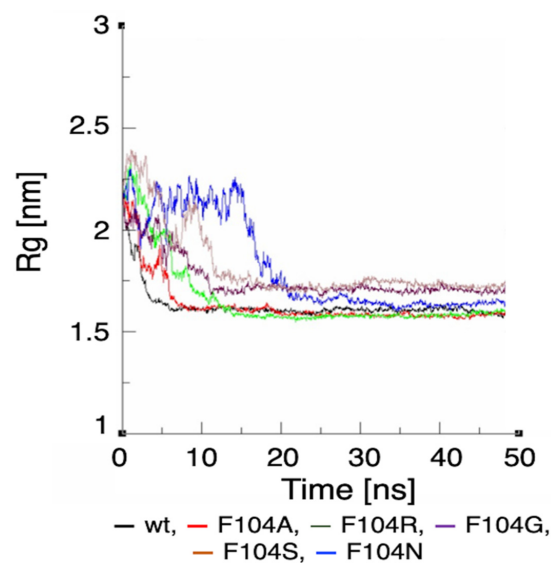
**Figure 6.** (A–H) FWMF motif (AA 101–104)-specific mutation-dependent FAM72A protein and UNG2 (AA 1–45) peptide heterodimer free binding energy. FAM72A mutations at AA F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S). The essential AA residues in FAM72A-wt and UNG2 interaction are the 2, 37–40, 61, 83, 104, 123–126, and 131 AAs of FAM72A. The highest binding energy (kcal/mol) goes to FAM72A’s AA F104 within the FWMF motif (AA 101–104). Positive values indicate low binding energies and weak interactions, whereas more negative values indicate stronger interactions. Values around 0 indicate moderate interactions between FAM72A and UNG2. A change in the FWMF motif from wt F104 to F104R shows the largest effect, turning the binding energy from negative (−2.7; strong binding/hydrophobic core) to positive (+2.04; strong binding/hydrogen bonding).



**Figure 7.** RMSD plot showing the effect of FAM72A mutations at wt AA F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S) within the FWMF motif (AA 101–104) on the dynamic nature of the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer binding.



**Figure 8.** RMSF plot showing the effect of FAM72A mutations at wt AA F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S) within the FWMF motif (AA 101–104) on the dynamic nature of the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer binding.

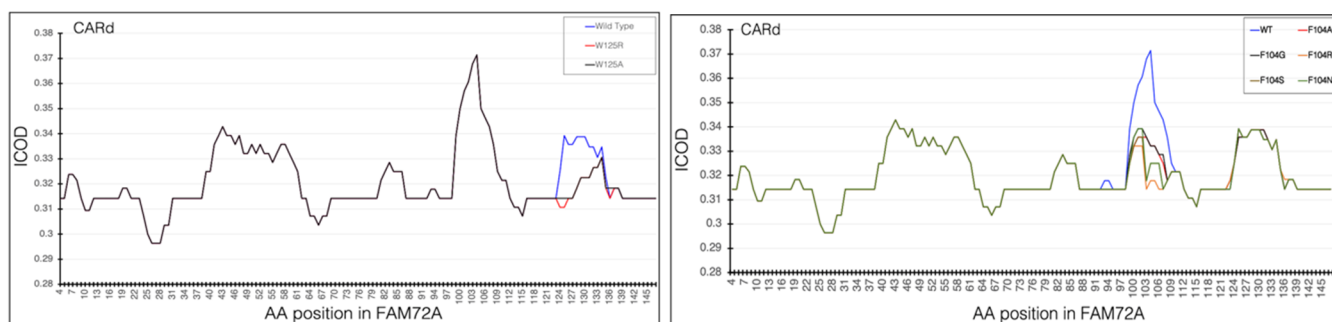


**Figure 9.** Rg plot showing the effect of FAM72A mutations at AA F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S) within the FWMF motif (AA 101–104) on the dynamic nature of the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer binding.

We assessed the effect of F104 mutations in FAM72A on the dynamic confirmation changes, stability, and rigidity of core and buried regions. Trajectories recorded up to 50 ns were plotted as RMSD, RMSF, and Rg, respectively. Figures 7–9 show the effect of these mutations on protein backbone changes in mutated FAM72A and signifies the pivotal role of this FWMF motif (AA 101–104) for the FAM72A-UNG2 interaction. These data confirm the FWMF motif as a suitable target to interfere with FAM72A-UNG2 signaling pathways.

### 3.7. CARd Analysis

Knowing that AAs Trp/W125 and F104 in FAM72A are important for the FAM72A-UNG2 interaction, we changed wt FAM72A-W125 (FAM72A-wt) to FAM72A-W125A or W125R, and wt FAM72A-F104 to F104A, F104R, F104N, F104G, and F104S, respectively, and performed CARd analysis (Figure 10).



**Figure 10.** CARd analysis showing the effect of FAM72A mutations at the AA tryptophan Trp W125 (W125 → W125A and W125R) and F104 phenylalanine (F104 → F104A, F104R, F104N, F104G, and F104S; within the FWMF motif (AA 101–104)) on the dynamic nature of the FAM72A protein and UNG2 (AA 1–45) peptide heterodimer binding. CARd analysis confirmed the FAM72A-UNG2 interaction and the effect of the various FAM72A mutations on it, as carbon governs cohesiveness at the active site in proteins. Accordingly, internal carbon optimized domains (ICODs) of FAM72Awt and UNG2 (AA 1–45) interaction validate a strong overlap of FAM72A-UNG2 binding. In particular, mutations in FAM72A turn the ICODs of FAM72A at AAs 100–104 and 125–135 into non-ICODs, encouraging a less favorable and weaker FAM72A<sub>mut</sub>-UNG2 induced-fit-binding than that of FAM72A<sub>wt</sub>-UNG2. F104A, F104G: change from hydrophobic (W) and carbon-rich (W) to hydrophobic and carbon-low; F104R: change from hydrophobic (F) and carbon-rich (F) to hydrophilic, carbon-rich, and basic AA; F104S, F104N: change from hydrophobic (F) and carbon-rich (F) to hydrophilic, carbon-low, and polar AA. CARd reveals that F104 tends to switch to F104R > F104S > F104N > F104G > F104A. Code information: F104: uuu, A104: gcu, G104: ggu, S104: ucu, N104: aau, R104: cgu.

Internal carbon optimized domains (ICODs) of the FAM72A<sub>wt</sub>-UNG2 (AA 1–45) heterodimer confirmed the FAM72A<sub>wt</sub>-UNG2 (AA 1–45) interaction and the effect of the various FAM72A mutations on it, as carbon governs cohesiveness at the active site in proteins. Accordingly, ICODs of the FAM72A<sub>wt</sub>-UNG2 (AA 1–45) interaction validate a strong overlap of FAM72A-UNG2 binding. Particularly, mutations in FAM72A turn the ICODs of FAM72A at AAs 100–104 and 125–135 into non-ICODs, encouraging a less favorable and weaker FAM72A<sub>mut</sub>-UNG2 induced-fit-binding than that of FAM72A<sub>wt</sub>-UNG2. Based on CARd analysis, F104 tends to switch to F104R > F104S > F104N > F104G > F104A. The likelihood of the F104 (uuu) mutation to S104: ucu appears to be highest, as only one nucleotide would be changed compared with the others (F104R: cgu, F104N: aau, F104G: ggu, and F104A: gcu).

### 3.8. Lead Discovery and Chemical Docking—Interference with FAM72A-UNG2 Interaction and Activity

We performed a virtual high-throughput screening to detect a potential lead interference with the FAM72A-UNG2 interaction [71–73]. Binding scores, along with drug-likeness and pharmacophore properties, were considered [35,74–78]. The MTiOpen Screen required Autodock vina to exercise the docking of chemical libraries, scoring, and ensemble

analysis [54,55]. The virtual screening suggested 100 compounds, and predicted binding energies (kcal/mol) were considered to filter the hits for the selection of the “best” hit for optimization in order to identify the promising lead compound. Based on predicted binding energies, we identified withaferin B (PubChem CID: 11113907) as the “best” hit (binding energy  $-0.5$  kcal/mol) molecule that could potentially interfere with the FAM72A-UNG2 interaction (Figure 11). In the lead generation, the Glide XP docking analysis showed a strong binding affinity of  $-1.868$  kcal/mol at the FAM72A-UNG2 interference site. Active AA residues contributing to the interaction were visualized by LIGPLOT (Figure 11).

Interestingly, withaferin B has structural similarities with withaferin A. Both compounds are withanolide analogues (derived from *Withania somnifera* (Indian ginseng)) and contain an oxapentacyclo moiety. Besides, in their central moieties, withaferin B contains octadecan-5-yl, whereas withaferin A contains octadec-4-en-3-one. Of note, withaferin A has been reported to be a potential anti-cancer molecule that can inhibit cell proliferation, cell migration, and cell invasion [79–86]. Similarly, withaferin B, bound to the FAM72A-UNG2 heterodimer, could possibly block FAM72A-UNG2 signaling pathways in cancer cells [26,29]; however, thus far, the biological and therapeutic properties of withaferin B remain unknown.

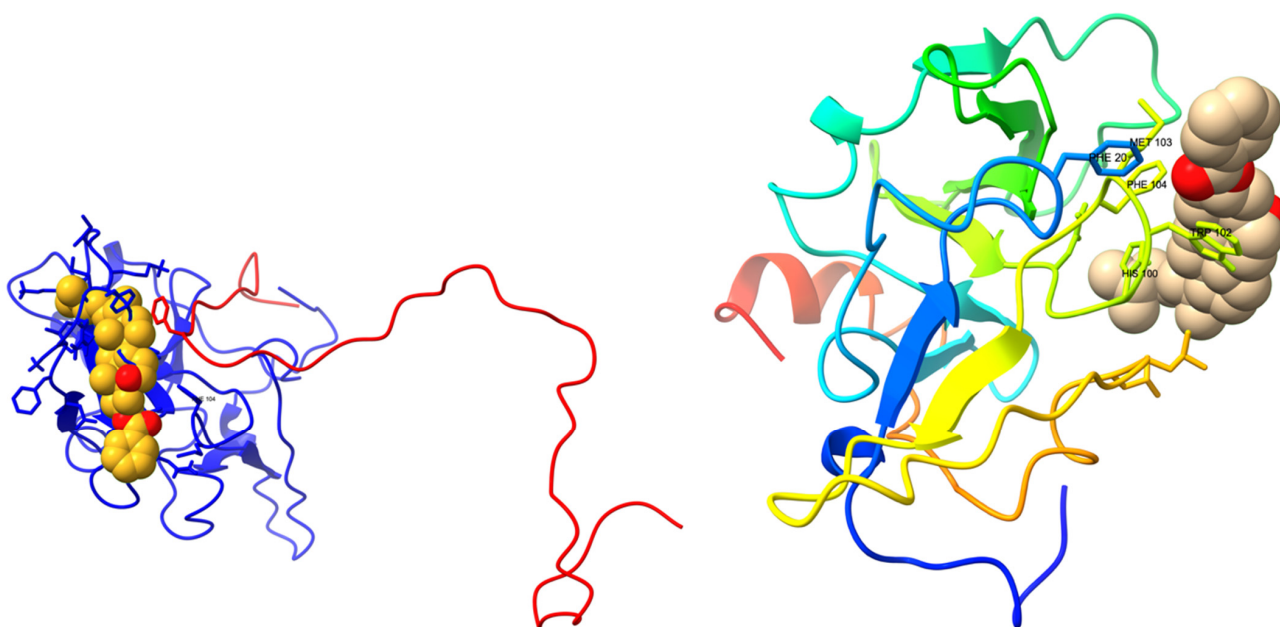
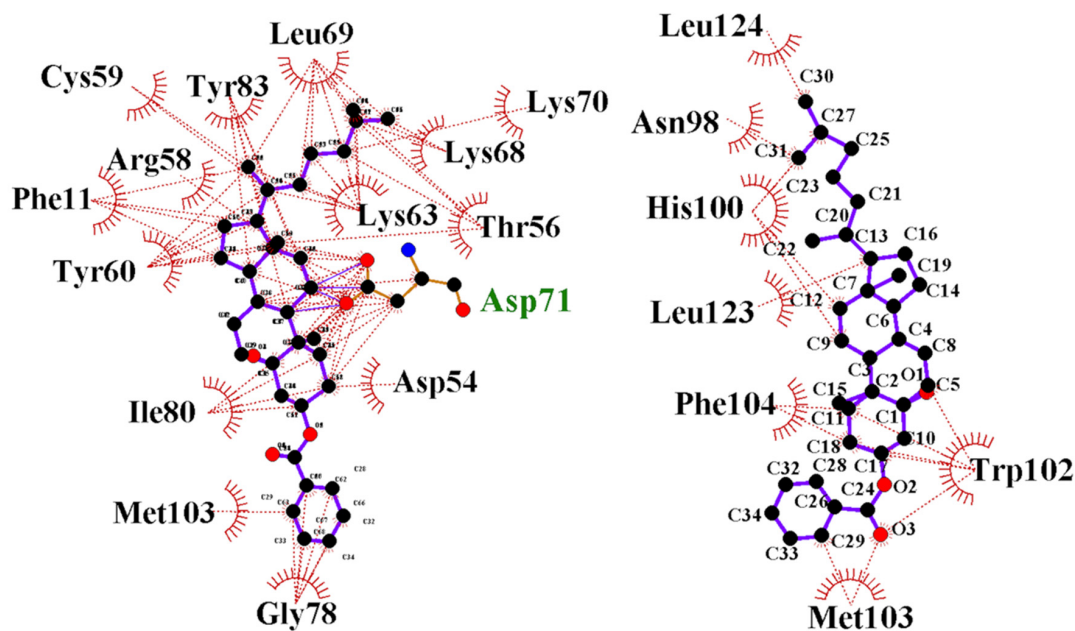


Figure 11. Cont.



**Figure 11.** Interaction of the FAM72Awt-UNG2 (AA 1–45) heterodimer with withaferin B, another withanolide analog. (**Top left-hand side**): Schematic 3D interaction of FAM72A protein and UNG2 (AA 1–45) peptide heterodimer (cartoon model) with withaferin B (space-fill model) is shown. (**Top right-hand side**): FAM72A protein monomer (cartoon model) with withaferin B (space-fill model) is shown. (**Bottom left-hand side**): LIGPLOT output showing the interaction between the FAM72Awt-UNG2 (AA 1–45) heterodimer and withaferin B. Green Asp71 indicates hydrogen bonding. (**Bottom right-hand side**): LIGPLOT output showing the interaction of FAM72Awt monomer with withaferin B. The binding affinity was calculated as  $-1.868$  kCal/Mol. Withaferin B, PubChem compound ID (CID): 11113907, molecular formula:  $C_{34}H_{50}O_3$ , molecular weight: 506.8 g/mol. IUPAC Name: [(1S,2R,5S,7S,9R,11S,12S,15R,16R)-2,16-dimethyl-15-[(2R)-6-methylheptan-2-yl]-8-oxapentacyclo[9.7.0.0<sup>2,7</sup>.0<sup>7,9</sup>.0<sup>12,16</sup>] octadecan-5-yl] benzoate. UNG2' AAs Tyr(Y)8 and Phe(F)11 are the most important AAs for the interaction with the FAM72A FWMF (AAs 101–104) motif. Withaferin B binds to the FAM72A-UNG2 heterodimer's interface at the FWMF motif and interacts strongly with both FAM72A and UNG2. Our data show that withaferin B could probably bind to the FAM72A-UNG2 heterodimer using electrostatic interactions and hydrophobic contacts via FAM72A' AAs Y60, T56, C59, and M103 and hydrogen bonding with FAM72A' D71. Moreover, withaferin B could probably bind to the FAM72A-UNG2 heterodimer using hydrophobic contacts via UNG2' AA F11 to disrupt the stability of FAM72A-UNG2 chain attachment and, thus, to inhibit the formation of active FAM72A-UNG2 protein complexes. As a result, FAM72A-UNG2 cell signaling could be turned off.

#### 4. Conclusions

Accumulating evidence indicates the involvement of FAM72A in tumorigenesis [24–26,29,87,88]. Elevated FAM72A causes reduced UNG2 levels, eventually leading to new mutations [24,25,27–29]. Our data pave the way for new investigative experimental approaches to validate the prevention of cancer by interfering with the FAM72A-UNG2 signaling pathways using withaferin B. Withaferin B is a potential candidate for future investigations in the interference with genome stability, centromere formation, and genome editing, and on potential therapeutic strategies for the treatment of cancer.

**Author Contributions:** Conceptualization, K.H.; methodology, S.R., S.P., R.E. and K.H.; validation, S.R., S.P., R.E., K.H., P.-S.K. and A.K.; formal analysis, S.R., S.P., R.E. and K.H.; investigation, S.R., S.P. and K.H.; resources, S.R. and K.H.; data curation, S.R., S.P. and K.H.; writing—original draft preparation, S.R., S.P. and K.H.; writing—review and editing, A.K., P.-S.K., S.R. and K.H.; visualization, S.R., P.-S.K. and A.K.; supervision, K.H.; project administration, K.H.; funding acquisition, S.R., A.K., P.-S.K. and K.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), which was funded by the Ministry of Education (2019R1F1A1056445).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding authors.

**Conflicts of Interest:** The authors have no competing financial interest.

## References

1. Chastain, P.D., II; Nakamura, J.; Rao, S.; Chu, H.; Ibrahim, J.G.; Swenberg, J.A.; Kaufman, D.G. Abasic sites preferentially form at regions undergoing DNA replication. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2010**, *24*, 3674–3680. [[CrossRef](#)]
2. Schormann, N.; Ricciardi, R.; Chattopadhyay, D. Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **2014**, *23*, 1667–1685. [[CrossRef](#)] [[PubMed](#)]
3. Beard, W.A.; Horton, J.K.; Prasad, R.; Wilson, S.H. Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annu. Rev. Biochem.* **2019**, *88*, 137–162. [[CrossRef](#)] [[PubMed](#)]
4. Akbari, M.; Otterlei, M.; Peña-Diaz, J.; Krokan, H.E. Different organization of base excision repair of uracil in DNA in nuclei and mitochondria and selective upregulation of mitochondrial uracil-DNA glycosylase after oxidative stress. *Neuroscience* **2007**, *145*, 1201–1212. [[CrossRef](#)] [[PubMed](#)]
5. Krokan, H.E.; Otterlei, M.; Nilsen, H.; Kavli, B.; Skorpen, F.; Andersen, S.; Skjelbred, C.; Akbari, M.; Aas, P.A.; Slupphaug, G. Properties and functions of human uracil-DNA glycosylase from the UNG gene. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *68*, 365–386. [[CrossRef](#)]
6. Akbari, M.; Otterlei, M.; Peña-Diaz, J.; Aas, P.A.; Kavli, B.; Liabakk, N.B.; Hagen, L.; Imai, K.; Durandy, A.; Slupphaug, G.; et al. Repair of U/G and U/A in DNA by UNG2-associated repair complexes takes place predominantly by short-patch repair both in proliferating and growth-arrested cells. *Nucleic Acids Res.* **2004**, *32*, 5486–5498. [[CrossRef](#)] [[PubMed](#)]
7. Parker, J.B.; Bianchet, M.A.; Krosky, D.J.; Friedman, J.I.; Amzel, L.M.; Stivers, J.T. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **2007**, *449*, 433–437. [[CrossRef](#)]
8. Wyatt, M.D. Advances in understanding the coupling of DNA base modifying enzymes to processes involving base excision repair. *Adv. Cancer Res.* **2013**, *119*, 63–106. [[CrossRef](#)]
9. Lindahl, T. DNA repair enzymes. *Annu. Rev. Biochem.* **1982**, *51*, 61–87. [[CrossRef](#)] [[PubMed](#)]
10. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709–715. [[CrossRef](#)]
11. Loeb, L.A. Apurinic sites as mutagenic intermediates. *Cell* **1985**, *40*, 483–484. [[CrossRef](#)]
12. Slupphaug, G.; Markussen, F.H.; Olsen, L.C.; Aasland, R.; Aarsaether, N.; Bakke, O.; Krokan, H.E.; Helland, D.E. Nuclear and mitochondrial forms of human uracil-DNA glycosylase are encoded by the same gene. *Nucleic Acids Res.* **1993**, *21*, 2579–2584. [[CrossRef](#)]
13. Haug, T.; Skorpen, F.; Kvaløy, K.; Eftedal, I.; Lund, H.; Krokan, H.E. Human uracil-DNA glycosylase gene: Sequence organization, methylation pattern, and mapping to chromosome 12q23-q24.1. *Genomics* **1996**, *36*, 408–416. [[CrossRef](#)]
14. Nilsen, H.; Otterlei, M.; Haug, T.; Solum, K.; Nagelhus, T.A.; Skorpen, F.; Krokan, H.E. Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the UNG gene. *Nucleic Acids Res.* **1997**, *25*, 750–755. [[CrossRef](#)]
15. Otterlei, M.; Haug, T.; Nagelhus, T.A.; Slupphaug, G.; Lindmo, T.; Krokan, H.E. Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. *Nucleic Acids Res.* **1998**, *26*, 4611–4617. [[CrossRef](#)]
16. Rodriguez, G.; Orris, B.; Majumdar, A.; Bhat, S.; Stivers, J.T. Macromolecular crowding induces compaction and DNA binding in the disordered N-terminal domain of hUNG2. *DNA Repair* **2020**, *86*, 102764. [[CrossRef](#)] [[PubMed](#)]
17. Perkins, J.L.; Zhao, L. The N-terminal domain of uracil-DNA glycosylase: Roles for disordered regions. *DNA Repair* **2021**, *101*, 103077. [[CrossRef](#)] [[PubMed](#)]
18. Weiser, B.P.; Rodriguez, G.; Cole, P.A.; Stivers, J.T. N-terminal domain of human uracil DNA glycosylase (hUNG2) promotes targeting to uracil sites adjacent to ssDNA-dsDNA junctions. *Nucleic Acids Res.* **2018**, *46*, 7169–7178. [[CrossRef](#)] [[PubMed](#)]
19. Zeitlin, S.G.; Chapados, B.R.; Baker, N.M.; Tai, C.; Slupphaug, G.; Wang, J.Y. Uracil DNA N-glycosylase promotes assembly of human centromere protein A. *PLoS ONE* **2011**, *6*, e17151. [[CrossRef](#)] [[PubMed](#)]
20. Nehar, S.; Mishra, M.; Heese, K. Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *FEBS Lett.* **2009**, *583*, 3247–3253. [[CrossRef](#)]
21. Ho, N.T.T.; Kutzner, A.; Heese, K. Brain plasticity, cognitive functions and neural stem cells: A pivotal role for the brain-specific neural master gene |SRGAP2-FAM72-|. *Biol. Chem.* **2017**, *399*, 55–61. [[CrossRef](#)]
22. Ho, N.T.; Kim, P.S.; Kutzner, A.; Heese, K. Cognitive Functions: Human vs. Animal-4:1 Advantage |-FAM72-SRGAP2-|. *J. Mol. Neurosci.* **2017**, *61*, 603–606. [[CrossRef](#)]



23. Kutzner, A.; Pramanik, S.; Kim, P.S.; Heese, K. All-or-(N)One—an epistemological characterization of the human tumorigenic neuronal paralogous FAM72 gene loci. *Genomics* **2015**, *106*, 278–285. [[CrossRef](#)]
24. Rahane, C.S.; Kutzner, A.; Heese, K. Establishing a human adrenocortical carcinoma (ACC)-specific gene mutation signature. *Cancer Genet* **2019**, *230*, 1–12. [[CrossRef](#)]
25. Rahane, C.S.; Kutzner, A.; Heese, K. A cancer tissue-specific FAM72 expression profile defines a novel glioblastoma multiform (GBM) gene-mutation signature. *J. Neurooncol.* **2019**, *141*, 57–70. [[CrossRef](#)] [[PubMed](#)]
26. Guo, C.; Zhang, X.; Fink, S.P.; Platzer, P.; Wilson, K.; Willson, J.K.; Wang, Z.; Markowitz, S.D. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res.* **2008**, *68*, 6118–6126. [[CrossRef](#)]
27. Feng, Y.; Li, C.; Stewart, J.; Barbulescu, P.; Desivo, N.S.; Álvarez-Quilón, A.; Pezo, R.C.; Perera, M.L.W.; Chan, K.; Tong, A.H.Y.; et al. FAM72A antagonizes UNG2 to promote mutagenic uracil repair during antibody maturation. *bioRxiv* **2020**. [[CrossRef](#)]
28. Rogier, M.; Moritz, J.; Robert, I.; Lescale, C.; Heyer, V.; Thomas-Claudepierre, A.-S.; Abello, A.; Deriano, L.; Reina-San-Martin, B. Fam72a controls the balance between error-prone and error-free DNA repair during antibody diversification. *bioRxiv* **2020**. [[CrossRef](#)]
29. Ho, N.T.T.; Rahane, C.S.; Pramanik, S.; Kim, P.S.; Kutzner, A.; Heese, K. FAM72, Glioblastoma Multiforme (GBM) and Beyond. *Cancers* **2021**, *13*, 1025. [[CrossRef](#)] [[PubMed](#)]
30. Pramanik, S.; Kutzner, A.; Heese, K. Lead discovery and in silico 3D structure modeling of tumorigenic FAM72A (p17). *Tumour Biol.* **2015**, *36*, 239–249. [[CrossRef](#)] [[PubMed](#)]
31. Parikh, S.S.; Mol, C.D.; Slupphaug, G.; Bharati, S.; Krokan, H.E.; Tainer, J.A. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **1998**, *17*, 5214–5226. [[CrossRef](#)] [[PubMed](#)]
32. Berman, H.M.; Battistuz, T.; Bhat, T.N.; Bluhm, W.F.; Bourne, P.E.; Burkhardt, K.; Feng, Z.; Gilliland, G.L.; Iype, L.; Jain, S.; et al. The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899–907. [[CrossRef](#)] [[PubMed](#)]
33. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [[CrossRef](#)]
34. Pramanik, S.; Kutzner, A.; Heese, K. 3D Structure, Dimerization Modeling, and Lead Discovery by Ligand-protein Interaction Analysis of p60 Transcription Regulator Protein (p60TRP). *Mol. Inform.* **2016**, *35*, 99–108. [[CrossRef](#)]
35. Pramanik, S.; Thaker, M.; Perumal, A.G.; Ekambaram, R.; Poondla, N.; Schmidt, M.; Kim, P.S.; Kutzner, A.; Heese, K. Proteomic Atomic Reveals a Distinctive Uracil-5-Methyltransferase. *Mol. Inform.* **2020**, *39*, e1900135. [[CrossRef](#)]
36. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40. [[CrossRef](#)] [[PubMed](#)]
37. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [[CrossRef](#)]
38. Webb, B.; Sali, A. Protein structure modeling with MODELLER. *Methods Mol. Biol.* **2014**, *1137*, 1–15. [[CrossRef](#)]
39. Webb, B.; Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* **2016**, *86*, 2.9.1–2.9.37. [[CrossRef](#)]
40. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]
41. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596. [[CrossRef](#)] [[PubMed](#)]
42. Buchinger, E.; Wiik, S.A.; Kusnierczyk, A.; Rabe, R.; Aas, P.A.; Kavli, B.; Slupphaug, G.; Aachmann, F.L. Backbone (1)H, (13)C and (15)N chemical shift assignment of full-length human uracil DNA glycosylase UNG2. *Biomol. NMR Assign* **2018**, *12*, 15–22. [[CrossRef](#)] [[PubMed](#)]
43. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. [[CrossRef](#)]
44. Zhou, P.; Jin, B.; Li, H.; Huang, S.Y. HPEPDOCK: A web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46*, W443–W450. [[CrossRef](#)]
45. Chen, F.; Liu, H.; Sun, H.; Pan, P.; Li, Y.; Li, D.; Hou, T. 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. [[CrossRef](#)] [[PubMed](#)]
46. Weng, G.; Wang, E.; Wang, Z.; Liu, H.; Zhu, F.; Li, D.; Hou, T. HawkDock: A web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* **2019**, *47*, W322–W330. [[CrossRef](#)] [[PubMed](#)]
47. Rajasekaran, E. CARd: Carbon distribution analysis program for protein sequences. *Bioinformatics* **2012**, *8*, 508–512. [[CrossRef](#)]
48. Kemmish, H.; Fasnacht, M.; Yan, L. Fully automated antibody structure prediction using BIOVIA tools: Validation study. *PLoS ONE* **2017**, *12*, e0177923. [[CrossRef](#)]
49. Kalathiya, U.; Padariya, M.; Baginski, M. Structural, functional, and stability change predictions in human telomerase upon specific point mutations. *Sci. Rep.* **2019**, *9*, 8707. [[CrossRef](#)]
50. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [[CrossRef](#)]
51. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 133–144. [[CrossRef](#)]

52. Lomb, L.; Barends, T.R.; Kassemeyer, S.; Aquila, A.; Epp, S.W.; Erk, B.; Foucar, L.; Hartmann, R.; Rudek, B.; Rolles, D.; et al. Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Review B Condens. Matter Mater. Phys.* **2011**, *84*, 214111. [[CrossRef](#)] [[PubMed](#)]
53. Betancourt, M.R.; Skolnick, J. Universal similarity measure for comparing protein structures. *Biopolymers* **2001**, *59*, 305–309. [[CrossRef](#)]
54. Labbé, C.M.; Rey, J.; Lagorce, D.; Vavruša, M.; Becot, J.; Sperandio, O.; Villoutreix, B.O.; Tufféry, P.; Miteva, M.A. MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Res.* **2015**, *43*, W448–W454. [[CrossRef](#)]
55. 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*, 455–461. [[CrossRef](#)]
56. Tsai, T.-Y.; Chang, K.-W.; Chen, C.Y.-C. iScreen: World's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 525–531. [[CrossRef](#)] [[PubMed](#)]
57. Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H. PharmMapper 2017 update: A web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res.* **2017**, *45*, W356–W360. [[CrossRef](#)]
58. Xu, D.; Zhang, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **2011**, *101*, 2525–2534. [[CrossRef](#)]
59. Yang, J.; Roy, A.; Zhang, Y. BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.* **2013**, *41*, D1096–D1103. [[CrossRef](#)]
60. Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **2009**, *77* (Suppl. S9), 114–122. [[CrossRef](#)]
61. 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*, 2785–2791. [[CrossRef](#)]
62. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [[CrossRef](#)]
63. Senthil, R.; Sakthivel, M.; Usha, S. Structure-based drug design of peroxisome proliferator-activated receptor gamma inhibitors: Ferulic acid and derivatives. *J. Biomol. Struct. Dyn.* **2020**, *39*, 1–17. [[CrossRef](#)]
64. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [[CrossRef](#)]
65. Uversky, V.N. Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* **2015**, *282*, 1182–1189. [[CrossRef](#)] [[PubMed](#)]
66. Bondos, S.E.; Dunker, A.K.; Uversky, V.N. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun. Signal* **2021**, *19*, 88. [[CrossRef](#)] [[PubMed](#)]
67. Uversky, V.N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins* **2013**, *1*, e24684. [[CrossRef](#)] [[PubMed](#)]
68. Berná, L.; Alvarez-Valin, F. Evolutionary volatile Cysteines and protein disorder in the fast evolving tunicate *Oikopleura dioica*. *Mar. Genom.* **2015**, *24 Pt 1*, 47–54. [[CrossRef](#)]
69. Mejia, J.S.; Arthun, E.N.; Titus, R.G. Cysteine-free proteins in the immunobiology of arthropod-borne diseases. *J. Biomed. Biotechnol.* **2010**, *2010*, 171537. [[CrossRef](#)]
70. Uversky, V.N. Protein intrinsic disorder and structure-function continuum. *Prog. Mol. Biol. Transl. Sci.* **2019**, *166*, 1–17. [[CrossRef](#)]
71. Schröder, J.; Klinger, A.; Oellien, F.; Marhöfer, R.J.; Duszenko, M.; Selzer, P.M. Docking-based virtual screening of covalently binding ligands: An orthogonal lead discovery approach. *J. Med. Chem.* **2013**, *56*, 1478–1490. [[CrossRef](#)]
72. Cavasotto, C.N.; Orry, A.J. Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top Med. Chem.* **2007**, *7*, 1006–1014. [[CrossRef](#)] [[PubMed](#)]
73. Kontoyianni, M. Docking and Virtual Screening in Drug Discovery. *Methods Mol. Biol.* **2017**, *1647*, 255–266. [[CrossRef](#)]
74. Ji, D.; Xu, M.; Udenigwe, C.C.; Agyei, D. Physicochemical characterisation, molecular docking, and drug-likeness evaluation of hypotensive peptides encrypted in flaxseed proteome. *Curr. Res. Food Sci.* **2020**, *3*, 41–50. [[CrossRef](#)] [[PubMed](#)]
75. Benet, L.Z.; Hosey, C.M.; Ursu, O.; Oprea, T.I. BDDCS, the Rule of 5 and drugability. *Adv. Drug Deliv. Rev.* **2016**, *101*, 89–98. [[CrossRef](#)] [[PubMed](#)]
76. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [[CrossRef](#)]
77. Opo, F.; Rahman, M.M.; Ahammad, F.; Ahmed, I.; Bhuiyan, M.A.; Asiri, A.M. Author Correction: Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci. Rep.* **2021**, *11*, 19106. [[CrossRef](#)]
78. Renganathan, S.; Manokaran, S.; Vasanthakumar, P.; Singaravelu, U.; Kim, P.S.; Kutzner, A.; Heese, K. Phytochemical Profiling in Conjunction with In Vitro and In Silico Studies to Identify Human  $\alpha$ -Amylase Inhibitors in *Leucaena leucocephala* (Lam.) De Wit for the Treatment of Diabetes Mellitus. *ACS Omega* **2021**, *6*, 19045–19057. [[CrossRef](#)]

79. Kaileh, M.; Vanden Berghe, W.; Heyerick, A.; Horion, J.; Piette, J.; Libert, C.; De Keukeleire, D.; Essawi, T.; Haegeman, G. Withaferin A strongly elicits I $\kappa$ B kinase beta hyperphosphorylation concomitant with potent inhibition of its kinase activity. *J. Biol. Chem.* **2007**, *282*, 4253–4264. [[CrossRef](#)]
80. McKenna, M.K.; Gachuki, B.W.; Alhakeem, S.S.; Oben, K.N.; Rangnekar, V.M.; Gupta, R.C.; Bondada, S. Anti-cancer activity of withaferin A in B-cell lymphoma. *Cancer Biol. Ther.* **2015**, *16*, 1088–1098. [[CrossRef](#)]
81. Yadav, D.K.; Kumar, S.; Saloni, H.S.; Kim, M.H.; Sharma, P.; Misra, S.; Khan, F. Molecular docking, QSAR and ADMET studies of withanolide analogs against breast cancer. *Drug Des. Devel Ther.* **2017**, *11*, 1859–1870. [[CrossRef](#)] [[PubMed](#)]
82. Dom, M.; Vanden Berghe, W.; Van Ostade, X. Broad-spectrum antitumor properties of Withaferin A: A proteomic perspective. *RSC Med. Chem.* **2020**, *11*, 30–50. [[CrossRef](#)]
83. Chirumamilla, C.S.; Pérez-Novo, C.; Van Ostade, X.; Vanden Berghe, W. Molecular insights into cancer therapeutic effects of the dietary medicinal phytochemical withaferin A. *Proc. Nutr. Soc.* **2017**, *76*, 96–105. [[CrossRef](#)] [[PubMed](#)]
84. Bonandi, E.; Mori, M.; Infante, P.; Basili, I.; Di Marcotullio, L.; Calcaterra, A.; Catti, F.; Botta, B.; Passarella, D. Design and Synthesis of New Withaferin A Inspired Hedgehog Pathway Inhibitors. *Chemistry* **2021**, *27*, 8350–8357. [[CrossRef](#)]
85. Odongo, R.; Demiroglu-Zergeroglu, A.; Çakır, T. A systems pharmacology approach based on oncogenic signalling pathways to determine the mechanisms of action of natural products in breast cancer from transcriptome data. *BMC Complement Med. Ther.* **2021**, *21*, 181. [[CrossRef](#)] [[PubMed](#)]
86. Xu, K.; Shi, H.; Du, Y.; Ou, J. Withaferin A inhibits proliferation of human endometrial cancer cells via transforming growth factor- $\beta$  (TGF- $\beta$ ) signalling. *3 Biotech* **2021**, *11*, 323. [[CrossRef](#)] [[PubMed](#)]
87. Yu, Y.; Wang, Z.; Zheng, Q.; Li, J. FAM72 serves as a biomarker of poor prognosis in human lung adenocarcinoma. *Aging* **2021**, *13*, 8155–8176. [[CrossRef](#)]
88. Zhang, T.; Nie, Y.; Gu, J.; Cai, K.; Chen, X.; Li, H.; Wang, J. Identification of Mitochondrial-Related Prognostic Biomarkers Associated With Primary Bile Acid Biosynthesis and Tumor Microenvironment of Hepatocellular Carcinoma. *Front. Oncol.* **2021**, *11*, 587479. [[CrossRef](#)]