



OPEN

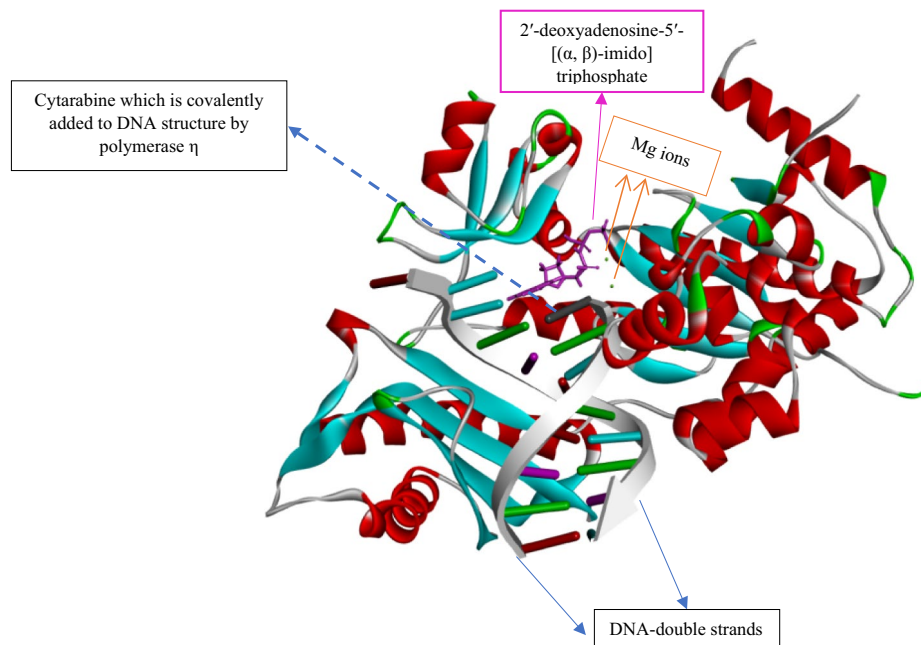
## Design of amino acid- and carbohydrate-based anticancer drugs to inhibit polymerase $\eta$

Sepideh Kalhor & Alireza Fattahi✉

DNA polymerase  $\eta$  (pol $\eta$ ) is of significant value for designing new families of anticancer drugs. This protein takes a role in many stages of the cell cycle, including DNA replication, translesion DNA synthesis, and the repairing process of DNA. According to many studies, a high level of expression of pol $\eta$  in most cases has been associated with low rates of patients' survival, regardless of considering the stage of tumor cells. Thus, the design of new drugs with fewer side effects to inhibit pol $\eta$  in cancerous cells has attracted attention in recent years. This project aims to design and explore the alternative inhibitors for pol $\eta$ , which are based on carbohydrates and amino acids. In terms of physicochemical properties, they are similar to the traditional anticancer drugs such as Cytarabine (cytosine arabinose). These alternative inhibitors are supposed to disrupt the DNA replication process in cancerous cells and prevent the tumor cells from mitosis. These newly designed structures, which are based on natural products, are expected to be non-toxic and to have the same chemotherapeutic impact as the traditional agents. The combinatorial use of quantum mechanics studies and molecular dynamic simulation has enabled us to precisely predict the inhibition mechanism of the newly designed structure, which is based on carbohydrates and amino acids, and compare it with that of the traditional chemotherapeutic drugs such as Cytarabine. Our results suggest that the inhibitors containing the natural building blocks of amino acid and carbohydrate could be considered alternative drugs for Cytarabine to block pol $\eta$ .

As it is widely known, the correct transmission of chromosomes is valuable for inheritance, which is facilitated by the DNA replication process. The life cycle of cells is not flawless. A variety of environmental agents alongside reactive metabolites generated during the life cycle of cells can cause lesions in DNA structures<sup>1</sup>. On the other hand, neoplastic transformation and genome instability, which are the replication stress results, can be another reason for DNA damage<sup>2</sup>. At first glance, we expect cells to employ pathways to repair DNA damage. Under replication stress, the cells are supposed to replicate the damaged DNA before the mitosis phase of cells' life cycle. Undoubtedly, the copies of this damaged DNA will pass down to both daughter cells, and this mechanism, which is called DNA damage tolerance, will be repeated for daughter cells and result in cancerous cells<sup>3</sup>. Many proteins involved in producing these cancerous cells, including DNA polymerase  $\eta$ , can target the newly designed anticancer drugs. DNA polymerases are participating in crucial steps of DNA replication in that they involve in functioning in the replication of the genomic material, playing a role in DNA repair, and taking part in a process called translesion DNA synthesis (TLS)<sup>4</sup>. Fifteen mammalian DNA polymerases have been identified, but our focus in this study is DNA polymerase  $\eta$  (Pol $\eta$ ). Polymerase  $\eta$  among the other family members is a unique enzyme. For instance, it bypasses the barriers caused by dominant UV lesion<sup>5</sup>. According to many studies, a high level of expression of polymerase  $\eta$  has been associated with low rates of patients' survivals<sup>6</sup>. Even in cases such as acute myeloid leukaemia, the polymerase  $\eta$  can be a good target for the design of new anticancer drugs. Many attempts have been made to discover the pathway in which the polymerase  $\eta$  is involved. For example, this enzyme can perform translesion DNA synthesis, even in the presence of cisplatin adducts<sup>7,8</sup>. As the therapeutic drug, Cytarabine is used to treat acute myeloid leukemia and expected to inhibit DNA polymerase  $\eta$ . However, recently, it was found that this drug could form covalent bonds with DNA by the aim of DNA polymerase  $\eta$ <sup>9</sup>. Thus, DNA polymerase  $\eta$  (Pol $\eta$ ) can continue DNA synthesis from Cytarabine terminated primers, which leads to drug resistance and DNA mutation. Also, high-resolution structures of the complex of DNA, DNA polymerase  $\eta$  (Pol $\eta$ ), and Cytarabine (all three structures together) have been provided (PDB code: 6D0Z)<sup>9</sup>. The above-described issues of Cytarabine attracted our attention to design alternative drugs to inhibit DNA polymerase  $\eta$  (Pol $\eta$ ) predominantly. Cytarabine, as the synthetic analog of deoxycytidine, enters the cell in different ways. It can be transported across the cell membrane by a nucleoside carrier system<sup>10</sup>. Diffusion at a high concentration of this

Department of Chemistry, Sharif University of Technology, Tehran, Iran. ✉email: fattahi@sharif.edu

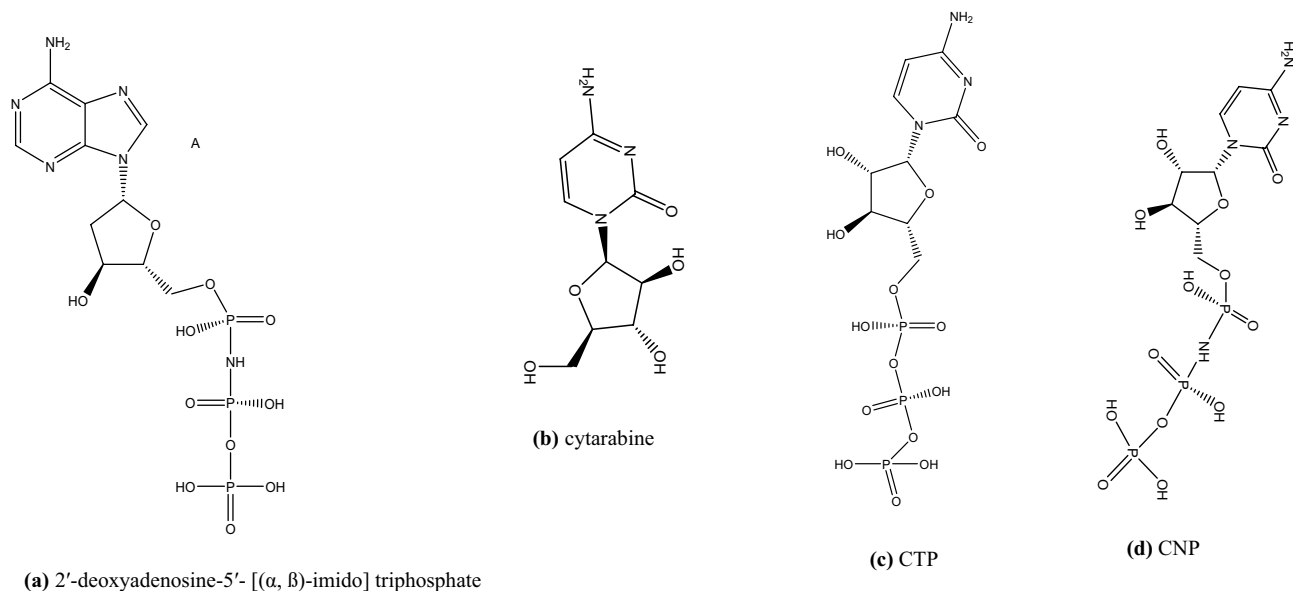


**Figure 1.** Structure of polymerase  $\eta$  complexed with DNA to which cytarabine is added covalently. Also 2'-deoxyadenosine-5'-[( $\alpha$ ,  $\beta$ )-imido] triphosphate (pink molecule) is incoming nucleotide.

drug has been reported, too<sup>11</sup>. Many adverse effects of Cytarabine including myelosuppression<sup>12</sup>, Anaphylactic reactions<sup>13</sup>, Drug Fever<sup>14</sup>, a very painful erythematous palmar-plantar bulla formation<sup>15</sup>, Subacute pulmonary failure<sup>16</sup>, and gastrointestinal toxicity<sup>17</sup> stimulated our research group to find the natural-based drug that can be used instead of Cytarabine. For 40 years, progress in protein crystallography and NMR has enabled the chemists to have three-dimensional structures of Protein–ligand complexes to visualize the selective interaction of drugs better and to optimize the structures of drugs to achieve more efficient patients' treatment<sup>18</sup>. Simultaneously, the achievements in computational chemistry have allowed scientists to predict molecular interactions and in-silico characterization of protein–ligand complexes<sup>19</sup>. The diversity of carbohydrates in nature has proved to be one of the best tools in drug discovery<sup>20</sup>. Moreover, well-defined stereochemical centers with flexible pyran and furan rings make them functional building blocks to present pharmacophoric groups<sup>21–24</sup>. Also, it has been found that the combination of sugars and peptides or sugars and small molecules can have a specific influence on structural properties such as metabolic stability<sup>25</sup>, solubility, membrane permeability<sup>26</sup>, biodistribution<sup>27</sup>, and ligand–target interactions<sup>28,29</sup>. As a result, our recent interest is to design structures based on sugars and amino acids via computational methods and molecular dynamic (MD) simulations. These designed amino acid-sugar-based structures could be potentially good candidates to replace Cytarabine. We used the structure with the PDB ID of 6D0Z as our initial structure (Fig. 1)<sup>9</sup>. In Fig. 1, the post-insertion step is observed, in which DNA polymerase  $\eta$  has covalently inserted cytarabine residue to one of the DNA strands. Also, in Fig. 1, the small molecule shown by pink color is the incoming nucleotide called 2'-deoxyadenosine-5'-[( $\alpha$ ,  $\beta$ )-imido] triphosphate in its inactive form. Green dots in Fig. 1 are  $Mg^{2+}$  (to fix the pH of crystallization at 8.0)<sup>9</sup>. Figure 1 also shows the crystallized form of the human Pol $\eta$  catalytic core (residues 1–432) complexed with a DNA including 12-nucleotide template (5'-CATGACAGTGCT-3')/8-nucleotide primer (5'-AGCACTGT-3') in which Cytarabine is covalently bonded to the 8-nucleotide strand. Also, as it can be seen in Fig. 1, the human Pol $\eta$  inserts 2'-deoxyadenosine-5'-[( $\alpha$ ,  $\beta$ )-imido] (containing triphosphate) into the DNA structure from the part of DNA terminated with Cytarabine. Thus, we can infer that triphosphate should have been added to Cytarabine before forming the covalent bond with the primer strand of DNA (Fig. 2c). One of the essential steps in the cell cycle is the simultaneous interaction of ((2R,3S,4S,5R)-5-(4-amino-2-oxypyrimidin-1(2H)-yl)-3,4-dihydroxytetrahydrofuran-2-yl) methyl tetrahydrogen (S, S)-triphosphate (CTP) with DNA strands and Pol $\eta$ . We considered this fact to design our new amino acid-sugar-based structure to kill cancerous cells undergoing DNA synthesis by Pol $\eta$  in the S-phase of the cell cycle.

As Cytarabine prodrug (Fig. 2b) is entering the cell, deoxycytidine and pyrimidine kinases will phosphorylate this prodrug and convert it to CTP (Fig. 2c)<sup>30</sup>. As seen in Fig. 2, the structure known as cytarabine-5'-[( $\alpha$ ,  $\beta$ )-imido] triphosphate (CNP) (Fig. 2d) is very similar to CTP (Fig. 2c), but CNP is non-hydrolysable compared to CTP. As a result, CNP is expected to be more stable than CTP in the presence of ions such as  $Mg^{2+}$  and water molecules. In this study, we have compared our designed structures with CNP. The synthesis of CNP can be achieved by methods described by Hirokazu Nakayama et al.<sup>31</sup>.

Our designed structures are based on amino acids and D-fructose, and their synthesis can be achieved by Amadori and Heyns methods<sup>32</sup>. We assume that our suggested structures can compete with cytarabine-5'-[( $\alpha$ ,  $\beta$ )-imido] triphosphate (CNP) in the interactions with Pol $\eta$  and DNA, as mentioned above. In this study, the summary of our approach toward drug design can be explained as follows. First, the suggested structures were optimized at the B3LYP/6-311++G (d, p) level<sup>33</sup>. Then, the docking procedures were performed to predict the



**Figure 2.** (a) 2'-deoxyadenosine-5'-[(α, β)-imido] triphosphate, (b) cytarabine, (c) CTP, and (d) CNP.

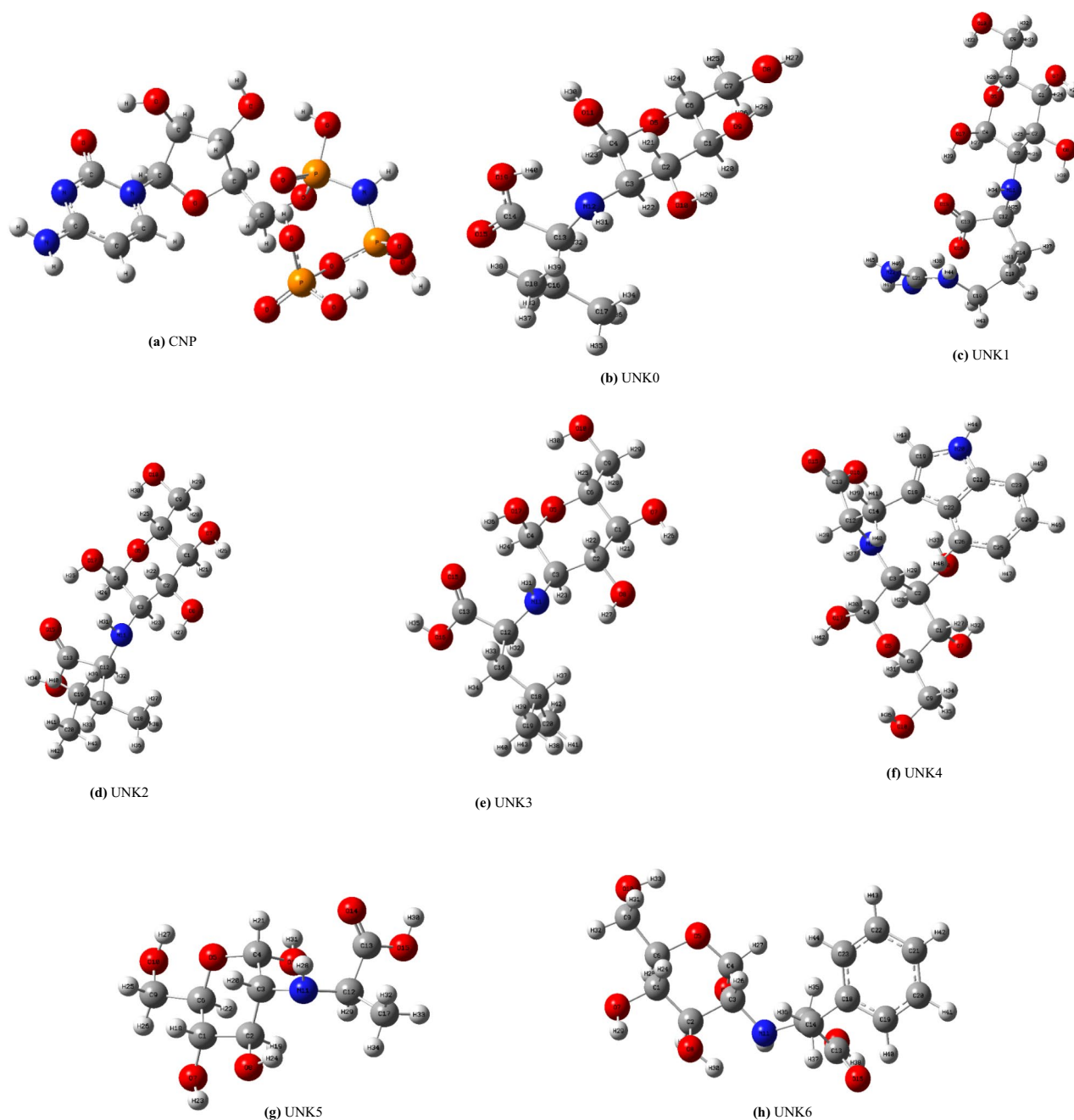
conformations and binding affinities of the drugs (complexed with DNA and DNA polymerase  $\eta$  (Pol $\eta$ )) studied herein including CNP and our suggested structures using the Auto Dock/Vina plugin with scripts from the Auto Dock Tools Package<sup>34,35</sup>. The intermolecular interactions, including hydrogen bonds, electrostatic interactions, and van der Waals for these drugs complexed with DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) were analyzed by discovery studio modeling environment (Discovery studio 4.5)<sup>36</sup>. Finally, the extracted data from the previous steps were used as inputs for MD simulation (molecular dynamic simulation) (The GROMACS 5.2)<sup>37</sup>. Whereas the experimental techniques are valuable tools to predict the behavior of biomolecular dynamics, some details such as pressure distributions inside membranes cannot be accessible through experiments. To address this issue, in this study, we applied molecular dynamic simulation (parameterized using the experimental data) to explore the interaction of the drugs with DNA and DNA polymerase  $\eta$  (Pol $\eta$ )<sup>38</sup>.

## Method and theoretical calculations

**Molecular docking and Geometry optimizations of the suggested structures.** The knowledge about the gas phase structural properties (including electronic properties, bond lengths, angles, and proper dihedrals) of CNP and our amino acid-carbohydrate-based suggested structures is necessary before the docking process and molecular dynamic simulations. The initial conformational searches at the relative energy of 10 kcal/mol were carried out with the MMFF force field (molecular mechanics) using Spartan software<sup>33</sup>. The geometry optimizations and frequency calculations were carried out using the 6-311++G\*\* basis set. The absence of imaginary frequencies confirmed that the optimized structures corresponded to the real minima<sup>33</sup>. The optimized structures of CNP and suggested structures are illustrated in Fig. 3. See Table 1 for nomenclatures and codes of molecules. For example, UNK4 refers to ((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-tryptophan structure (Table 1, entry 6). Before the docking procedure, the initial PDB was edited by VMD package<sup>39</sup>. All water molecules and other solvents, as well as the co-crystallized ligands including nucleotide (2'-deoxyadenosine-5'-[(α, β)-imido] triphosphate) and Cytarabine, incorporated in DNA, were removed using Discovery studio 4.5. To perform a docking procedure by Auto Dock/Vina, we considered the complex of protein and DNA as our receptors and CNP and designed structures as ligands. Then, the PDB files of the receptors and the ligands were saved as pdbqt files by auto dock tools<sup>40</sup>. In pdbqt files, information such as atomic charges, atom-type definitions, and rotatable bonds (like single bonds) is stored<sup>34</sup>. We used grid box with the dimensions  $20 \times 24 \times 28 \text{ \AA}^3$  with its center located at  $x = 52.861$ ,  $y = 6.193$ ,  $z = -12.977$  (Table 1).

To find the best conformer for CNP (which will be used for MD calculations), we used Auto Dock Vina, which suggested nine conformers. Figure 4 illustrates the structural information (obtained by Auto Dock Vina) for the first conformer of CNP complexed with DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) with the least RMSD = 0.0323 Å. This RMSD was calculated by overlaying this conformer over the Cytarabine structure obtained from the main pdb file (6D0Z). The structural information for all nine conformers of CNP is given in Fig. S1 in Supporting Information.

As seen in Table 1, our approach to design the new structures as anticancer drugs are based on computing some physical properties such as volume, area, and electrostatic potential map (EPS) of the designed structures, and comparison of these properties with those of the main drug, which is CNP in this study. The volume, area, and electrostatic potential map (EPS) of CNP and designed structures (listed in Table 1) were obtained by single-point energy calculations on the structures optimized at the B3LYP/6-311++G (d, p) level (in a vacuum) by Spartan software<sup>32</sup>. The last column of Table 1 represents the EPS. The red color depicts the regions with a higher electron density (nucleophilic centers), and blue color depicts the regions with a lower electron density



**Figure 3.** Optimized geometries of the lowest energy conformers of the suggested drugs based on carbohydrate and amino acids at B3LYP/6-311++G (d, p) level (see Table 1 for nomenclatures and codes of molecules).

(electrophilic centers). The EPS patterns allow us to select the candidate structures (designed structures) with the charge distribution most similar to the charge distribution of the main drug (i.e., CNP). Comparing the volume, area, EPS, and binding affinity of CNP with those of the designed compounds indicates that UNK4 (Entry 6 in Table 1) should be the most suitable candidate to replace CNP. For instance, in terms of binding affinity obtained by the Auto Dock/Vina, the binding affinity values of our designed structure UNK4 and CNP were calculated to be  $-9.2$  kcal/mol and  $-8.8$  kcal/mol, respectively, suggesting that UNK4 could replace CNP.

Moreover, we used Auto Dock Vina to study the interaction of cytarabine (Fig. 2b) and the complex of DNA and DNA polymerase  $\eta$  (Pol $\eta$ ). The results from this study can be seen in entry 1 of Table S1. Accordingly, the designed compound UNK4 has a better affinity towards the receptors (including DNA and DNA polymerase  $\eta$  (Pol $\eta$ )) than cytarabine. The binding affinity values of cytarabine (entry 1 of Table S1) and UNK4 (Entry 6 in Table 1) were calculated to be  $-6.6$  kcal/mol and  $-9.2$  kcal/mol, respectively, representing that UNK4 can interact more effectively with the complex of DNA and DNA polymerase  $\eta$  than cytarabine.

Also, to explore the probable off-target interactions of the designed compound UNK4, we chose receptors with which cytarabine structure and its derivatives interact. For example, one of the targets was DNA Polymerase

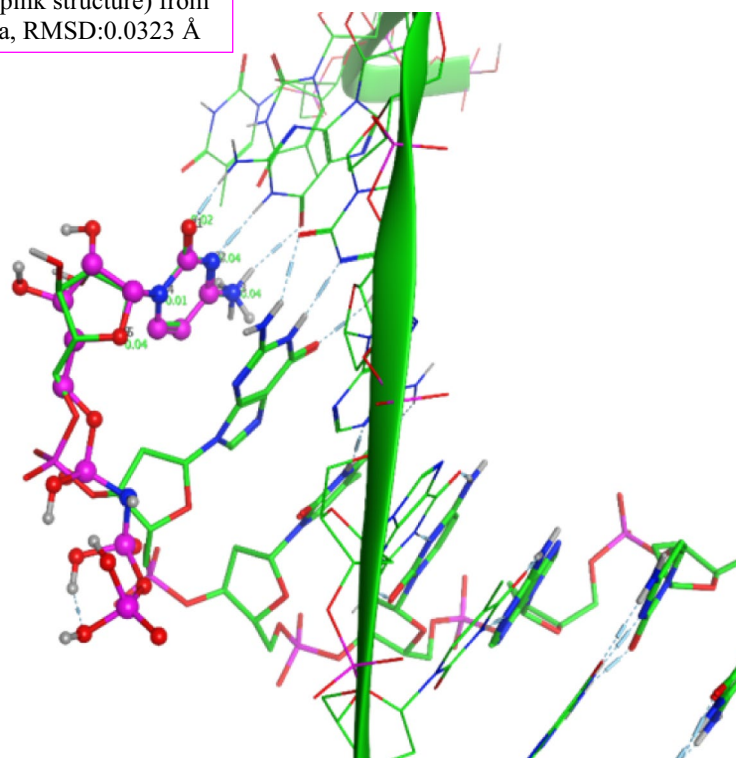
Entry	Name of structures	Structure	Structures* codes	E <sub>scf</sub> (Hartree/Particle)	Receptor	AFFINTY (kcal/mol)	Area (Å <sup>2</sup> )	Volume (Å <sup>3</sup> )	Electrostatic potential map
1	cytarabine-5'-[(α, β)-imido] triphosphate (CNP)		CNP	The suggested structure for CNP, before insertion into the DNA structure. -2594.4253	complex of protein and DNA	-8.8	389.77	349.33	
2	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-valine		UNK0	-1013.1124	complex of protein and DNA	-6.9	291.53	263.44	
3	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-arginine		UNK1	-1217.2906	complex of protein and DNA	-7.8	336.01	308.53	
4	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-isoleucine		UNK2	-1052.4080	complex of protein and DNA	-6.9	312.44	282.33	
5	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-leucine		UNK3	-1052.4104	complex of protein and DNA	-7.0	314.40	282.42	
6	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-tryptophan		UNK4	-1297.1131	complex of protein and DNA	-9.2	366.93	344.14	
7	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-D-alanine		UNK5	-934.5210	complex of protein and DNA	-6.5	256.98	227.56	
8	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-phenylalanine		UNK6	-1165.5427	complex of protein and DNA	-8.3	341.95	312.44	

**Table 1.** The structural information for cytarabine-5'-triphosphate and our designed amino acid-carbohydrate based structures.

$\lambda$ <sup>41</sup>, having a significant role in meiotic recombination and DNA repair<sup>42</sup>; the interactions of both structures, including cytarabine and UNK4, with Polymerase  $\lambda$ , were evaluated (entries 2 and 3, Table S1). Our designed compound UNK4 and cytarabine binding affinity values towards Polymerase  $\lambda$  were calculated to be  $-8.5$  kcal/mol and  $-7.9$  kcal/mol. These results show that UNK4 has more affinities towards receptors, including DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) than Polymerase  $\lambda$ . In contrast, cytarabine has more affinity towards Polymerase  $\lambda$  than towards receptors in this study, suggesting that UNK4 can interact better with receptors including DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) in comparison with cytarabine.

The other target was Human deoxyribonucleoside kinase playing a valuable role in the phosphorylation of the natural deoxyribonucleosides<sup>43</sup>. Also, this enzyme phosphorylates a significant number of nucleoside-based

(a) Conformer 1 (pink structure) from Auto dock vina, RMSD:0.0323 Å



**Figure 4.** The structural information (obtained by Auto Dock Vina) for the first conformer of CNP (brown) complexed with DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) with the least RMSD = 0.0323 Å (structure of Pol $\eta$  not shown here).

drugs<sup>43</sup>, which are vastly used as anticancer drugs such as cytarabine. As a result, we performed a docking process for cytarabine and dCK. Simultaneously, the binding affinity of UNK4 and dCK was obtained by the Auto Dock/Vina. The results can be observed in entries 4 and 5, Table S1. According to entries 4 and 5, Table S1, UNK4 and cytarabine bear binding affinity towards dCK as  $-7.2$  kcal/mol and  $-7.7$  kcal/mol, showing that UNK4 can possibly interact more effectively with receptors, including DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) compared to dCK. The studies about the possible mentioned off-targets interactions of UNK4 alongside with the results from Table 1 show that UNK4 can be a good replacement for cytarabine.


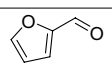
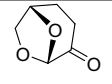
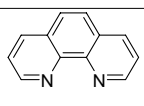
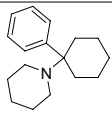
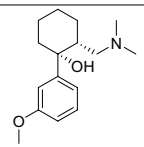
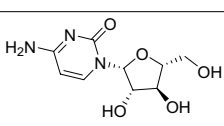
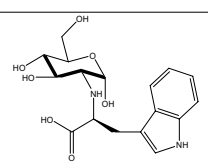
**Prediction of log P for the designed anticancer structure.** P is the partition coefficient defined as the ratio of the concentration of neutral drug molecule in the lipophilic solvent to its concentration in water. The lipophilicity and hydrophilicity of drug molecules are given by the logarithm of P (log P). 1-octanol is usually considered the lipophilic solvent because it represents properties similar to the membrane of cells<sup>44</sup>. To illustrate that our designed structure (UNK4) has good solubility in both water and lipophilic solvents, we calculated log P of UNK4. The free energy changes of transferring process of UNK4 from water to 1-octanol can be related to log P of UNK4 using the Eqs. (1) and (2).

$$\Delta G_{\text{transferring process}} = \Delta G_1 - \Delta G_2 \quad (1)$$

$$\Delta G_{\text{transferring process}} = - (RT) \ln P = - 2.303(RT) \log P \quad (2)$$

The free energy changes of transferring UNK4 in both water and 1-octanol solvents were calculated ( $\Delta G_1$  and  $\Delta G_2$ , respectively). From the difference between these two free energy changes ( $\Delta G_{\text{transferring process}}$ ), it is possible to obtain log P<sup>45</sup>. In the equation 2, R is the ideal gas constant, T is temperature (298 K), and P is the partition coefficient.

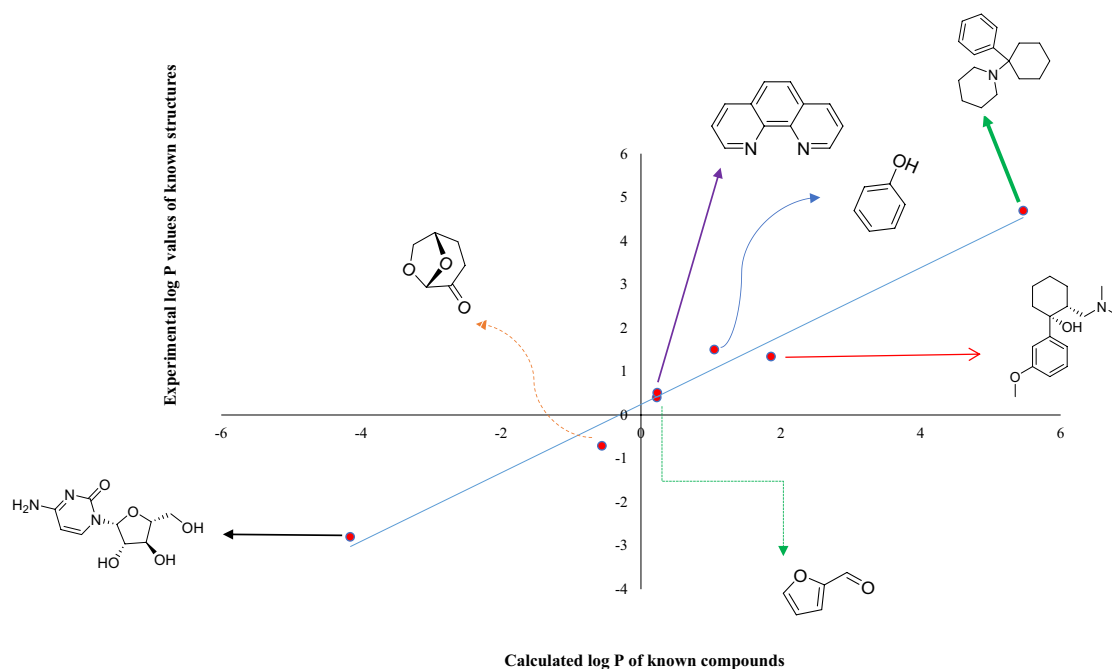
First, we computed log P of the structures with known experimentally-measured log P values, including phenol, furfural, dihydrolevoglucosenone<sup>46</sup>, and some known drugs such as Phenanthroline<sup>47,48</sup>, phencyclidine<sup>49</sup>, tramadol<sup>50</sup> and cytarabine<sup>51</sup> by employing density functional theory (DFT) calculations in combination with a solvation model based on density (SMD) (refer to Table 2)<sup>52</sup>. Then, we plotted the experimentally reported log P of known structures versus their computed log P (Fig. 5). We used this plot to predict log P of UNK4, which will be discussed below. The geometries of all molecules used in this part of the study were optimized at the M06-2X/6-311+G (d, p) level<sup>53</sup> in both water and 1-Octanol solvents<sup>32</sup>. Figures S2 and S3 show the optimized molecules and their charge distributions in water and 1-octanol, respectively. These figures illustrate that the

Entry	Name of structures	Structures of compounds	Predicted log P calculated at M06-2X/6-311+G (d, p)	Experimentally reported log P
1	Phenol		1.05	1.5 <sup>a</sup>
2	Furfural		0.23	0.4 <sup>a</sup>
3	Dihydrolevoglucosenone		-0.56	-0.7 <sup>a</sup>
4	Phenanthroline		0.23	0.51 <sup>b</sup>
5	Phencyclidine		5.47	4.69 <sup>c</sup>
6	Tramadol		1.86	1.34 <sup>d</sup>
7	Cytarabine		-4.16	-2.8 <sup>e</sup>
8	((2S,3R,4R,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-3-yl)-L-tryptophan		-2.39	-1.64 <sup>g</sup>

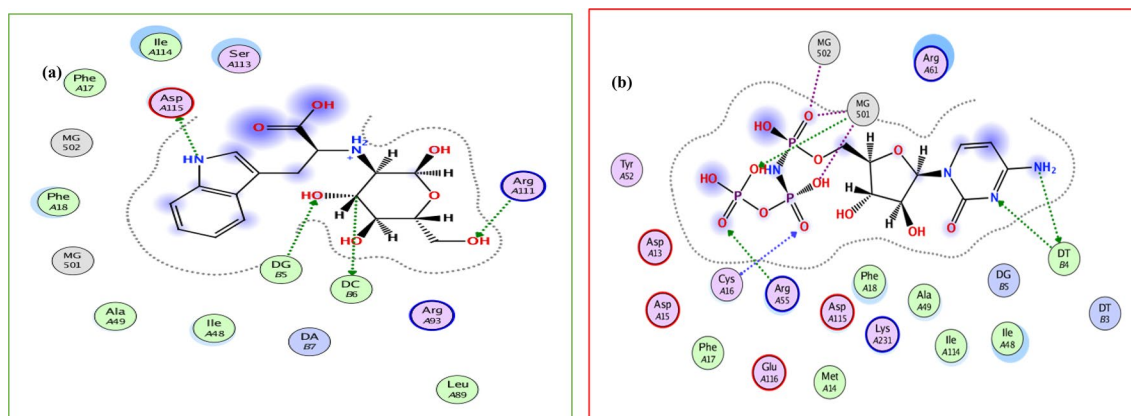
**Table 2.** The calculated (at the M06-2X/6-311+G (d, p) level) and experimental log P values. References for a-g: a<sup>46</sup>; b<sup>47,48</sup>; c<sup>49</sup>; d<sup>50</sup>; e<sup>51</sup>; g; predicted based on the plot in Fig. 5.

overall charge distribution of atoms in molecules can be different in various solvents. According to the Hessian analysis, no imaginary frequencies were found for the structures, indicating that all of the optimized structures were real minima. The properties of the solvents, such as the dielectric constant, refractive index, bulk surface tension, acidity, and basicity are required to calculate the free energy of each solute in both water and 1-Octanol solvents<sup>46</sup>. Figure 5 represents the plot of the experimentally measured log P values versus clog P (calculated log P) for the compounds listed in the entries 1–7 in Table 2 with  $y = 0.785 - 0.241x$  and  $R^2 = 0.98$ . Where  $y$  and  $x$  indicate the experimental and calculated log P values, respectively. This plot predicts the experimental log P value of UNK4 to be  $-1.64$  and comparing it with the log P value of Cytarabine ( $-2.8$ )<sup>51</sup> suggests that UNK4 tend to have more distribution in the octanol phase than water, which can have positive effects on the penetration of UNK4 through the cell membrane. In contrast, the log P value of Cytarabine indicates that it is a more hydrophilic molecule.

**Grid box searching.** According to the results extracted from Tables 1 and 2, and Fig. 3, among our suggested structures, the structure most similar to CNP is UNK4 (see Table 1). We assumed the cyclic structure for D-fructose incorporated into UNK4, according to the results reported by Barclay et al.<sup>54</sup>. They reported that 99.5% of D-fructose in water is in cyclic form, and just 0.5% of its structure is in acyclic form. We also used Auto Dock 4.2.2 software<sup>55</sup> to ensure that our grid boxes in auto dock/vina studies for UNK4 have been chosen correctly. The spacing between grid points was considered to be  $0.375 \text{ \AA}$ , the grid box size was set at  $126 \times 126 \times 126 \text{ \AA}^3$ , UNK4 as a ligand was considered to be flexible, and the grid searching was carried out at the local search genetic algorithm (LGA)<sup>56,57</sup>. 519 docking runs were performed to docking the ligand (UNK4) on the receptor (including the complex of DNA and pol $\eta$ ). The results obtained in this part of our project were in good agreement with the Auto dock/vina results in that the target of UNK4 is the same as the target of CNP, as confirmed by interaction sites shown in Fig. 6. Figure 6a indicates that UNK4 is surrounded by the residues



**Figure 5.** Plot of the experimental log P against the calculated log P (clog P).

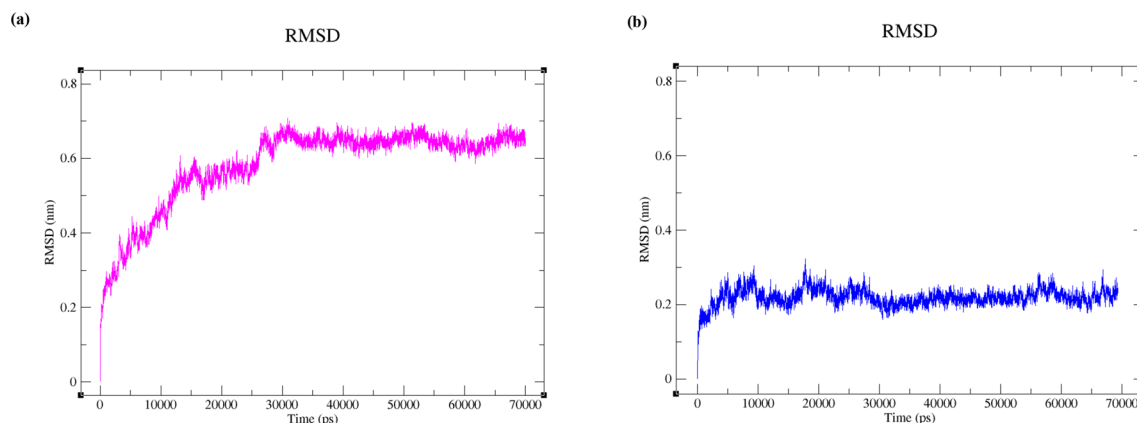


**Figure 6.** (a) The best pose of UNK4 in pocket site of receptors after 519 docking runs of Auto dock, (b) The best pose of CNP.

of DNA polymerase  $\eta$  (Pol $\eta$ ), including phenylalanine 17, phenylalanine 18, Isoleucine 48, alanine 49, leucine 89, Arginine 93, arginine 111, serine 113, Isoleucine 114, and aspartic acid 115. Figure 6b shows that CNP is also surrounded by the same residues of DNA polymerase  $\eta$  (Pol $\eta$ ), including aspartic acid 13, methionine 14, aspartic acid 15, Cysteine 16, phenylalanine 17, phenylalanine 18, alanine 49, Tyrosine 52, Arginine 55, Arginine 61, Isoleucine 114, aspartic acid 115, glutamic acid 116, and lysine 231, indicating that the active sites of UNK4 and CNP are very similar.

**Molecular dynamic simulation (MD).** MD simulations were carried out in four stages for two systems, including CNP and our designed structure (UNK4, see Table 1) complexed with their receptors (including DNA and DNA polymerase  $\eta$ ). The topology files for UNK4, CNP, and the nucleic acids of DNA were constructed based on the Gromos 43 a1 force field using the PRODRG 2.x online server. The complexes were solvated by a layer of SPC water in all directions<sup>58</sup>. Our simulation box was considered a triclinic box with faces at least 1.0 nm apart from the closest atom of our system. The simulation box for each system involved at least 18,607 water molecules, 8 chloride ions, and 2 magnesium ions. MD simulation for minimization of each system was performed in three steps. The first step included the minimization of the whole system by applying the steepest descent. The second step included the conjugate gradient methods to minimize the system. The third minimization step was again the steepest descent method with emtol value of 100.0 kJ/mol.nm. Following minimization steps, the overall system was equilibrated in the NVT ensemble (where N, V, and T stand for the number of particles,





**Figure 7.** (a) Polymerase  $\eta$  backbone RMSD values in the presence of CNP (series 1, pink plot). (b) polymerase  $\eta$  backbone RMSD values in the presence of UNK4 (series 2, blue plot).

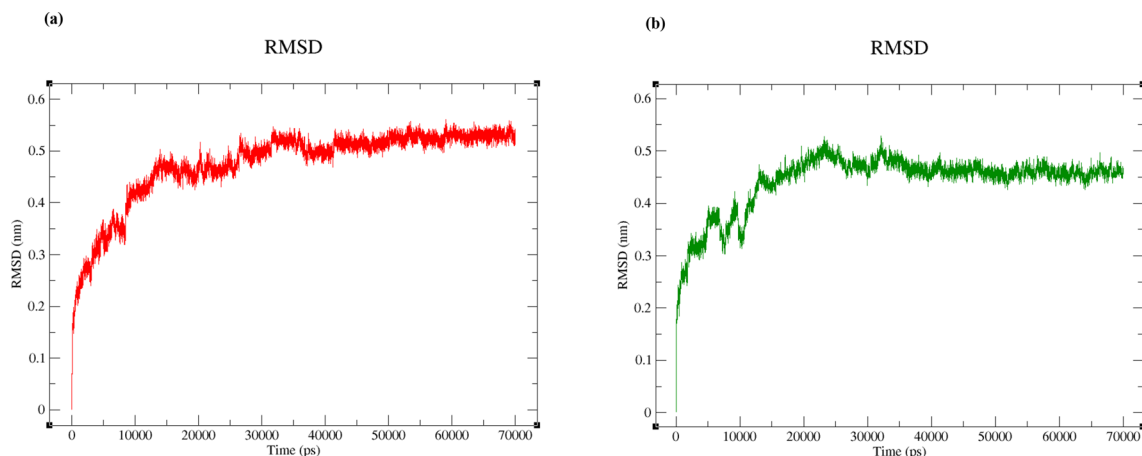
volume, and temperature, respectively) for 500 ps ( $1 \text{ ps} = 10^{-12} \text{ s}$ ) at 100 K. In this NVT ensemble, temperature-coupling times were 0.1 ps. The NPT ensemble followed this step (where N, P, and T stand for the number of particles, pressure, and temperature). The NPT equilibration was performed using md integrator for 1000 ps at 300 K and 1 bar. Berendsen algorithm was selected for both barostat and thermostat coupling algorithms in this step<sup>58</sup>. In the NPT ensemble, to increase the system's temperature from 100 to 300 K, the velocities were raised based on the Maxwell–Boltzmann distribution<sup>59</sup>. Finally, the MD production steps for each system was initiated for 70 ns ( $1 \text{ ns} = 10^{-9} \text{ s}$ ) at 300 K with a time step of 1 fs ( $1 \text{ fs} = 10^{-15} \text{ s}$ ) with 70 million steps in total. In this part, as we assumed that the overall 70 ns of MD simulation is enough to consider the various motions of protein including its folding as polymerase  $\eta$  is in complex with DNA and thus it is less flexible than free polymerase  $\eta$ . Parrinello–Rahman barostat<sup>60</sup> and Nose–Hoover thermostat algorithms<sup>60</sup> were used in MD simulation stages for pressure and temperature couplings, respectively. Furthermore, the periodic boundary condition was used for NPT simulation stage. To calculate van der Waals interactions, ‘Cut-off’ method was applied. In MD simulations, the van der Waals (rvdw) interactions were accompanied by neighbor list cut-off (rlist), where  $\text{rvdw} \geq \text{rlist}$ . The frequency of updating the list of neighbors was considered 10 ps. In addition, the Particle Mesh Ewald (PME) model was exploited to estimate the electrostatic interactions<sup>61,62</sup>. For constraints on hydrogen bonds in systems, the LINCS algorithm was employed in all MD simulations<sup>63</sup>.

## Results and discussions

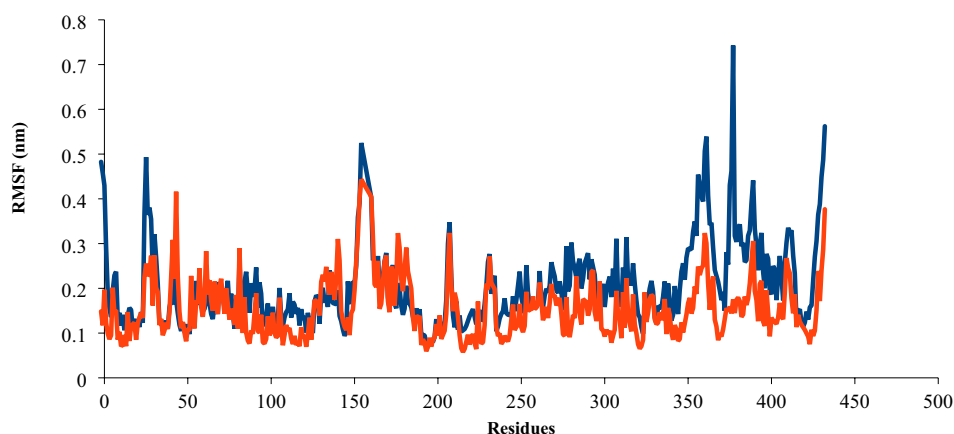
**Comparison of inhibitory effects of UNK4 and CNP.** To explore the similarity between UNK4 and CNP in inhibition of their receptors (the complex of DNA and DNA polymerase  $\eta$  (Pol $\eta$ )), the patterns of the change in the root-mean-square deviation (RMSD) value of polymerase  $\eta$  during 70 ns of MD simulations were compared for these two designed (UNK4) and main (CNP) drugs. All the calculations of RMSD values of DNA and protein were performed with the aid of the GROMACS 5.2 package. Figure 7a,b indicate the patterns of RMSD values for the backbone of polymerase  $\eta$  in the presence of CNP and UNK4, respectively, during 70 ns of simulation. According to Fig. 7a, the significant changes in RMSD values of polymerase  $\eta$  can be observed in the first 15 ns of MD simulation. As can be seen in the Fig. 7a, the RMSD values for the backbone of polymerase  $\eta$  in the presence of CNP have been achieved at the 0.59 nm with respect to the reference structures (PDB ID: 6D0Z); however, after 20 ns of MD simulation, the backbone of polymerase  $\eta$  has fluctuated in the range of 0.6–0.7 nm. While in the system, including the complex of DNA and DNA polymerase  $\eta$  (Pol $\eta$ ) and UNK4 (Fig. 7b), in the primary 2.5 ns of MD simulations, the sharp changes in RMSD values with respect to the reference structures (PDB ID: 6D0Z) are observed. After the first 2.5 ns, the perseverance of RMSD values in MD simulation in Fig. 7b can be observed. RMSD plot in Fig. 7b indicates that the backbone of polymerase  $\eta$  has fluctuated in the range of 0.2–0.3 nm, indicating that the system, including the complex of DNA and Pol $\eta$  along with UNK4, was in an equilibrated and converged stage and kept stable between 2.5 and 70 ns of MD simulation.

As DNA molecules were also considered to be part of the receptors, the same comparison was made between the changes in the root-mean-square deviation (RMSD) values of DNA in the presence of CNP (Fig. 8a) and UNK4 (Fig. 8b). As seen in Fig. 8a,b, the fluctuations in RMSD values of DNA with respect to the reference structures (PDB ID: 6D0Z) in both systems were sharp for the first 15 ns of MD simulations. The dynamic steadiness of DNA structure (for both systems) is observed between 15 and 70 ns. After the first 15 ns, both systems' perseverance of DNA RMSD values (Fig. 8a,b) are observed. RMSD plots for DNA also obviously indicate that the backbone of DNA has fluctuated in the range of 0.5–0.55 nm in the complex of DNA and Pol $\eta$  along with CNP (Fig. 8a). The backbone of DNA has fluctuated in the range of 0.45–0.5 nm in the complex of DNA and Pol $\eta$  along with UNK4 (Fig. 8b), suggesting that both systems were in an equilibrated and converged stage and kept stable between 15 and 70 ns of MD simulations.

RMSF (root mean square fluctuation) of protein versus its residue was plotted for both systems to characterize further the dynamic behaviour or structural flexibility of the above-described systems (Fig. 9). The blue and red plots in Fig. 9 correspond to the Polymerase  $\eta$  complexed with CNP and UNK4, respectively. Polymerase  $\eta$  (Pol $\eta$ ) has the template-primer, including palm (including the residues 1–13 and 90–238), fingers (including the



**Figure 8.** (a) RMSD values of DNA in the presence of CNP. (b) RMSD values of DNA in the presence of UNK4.



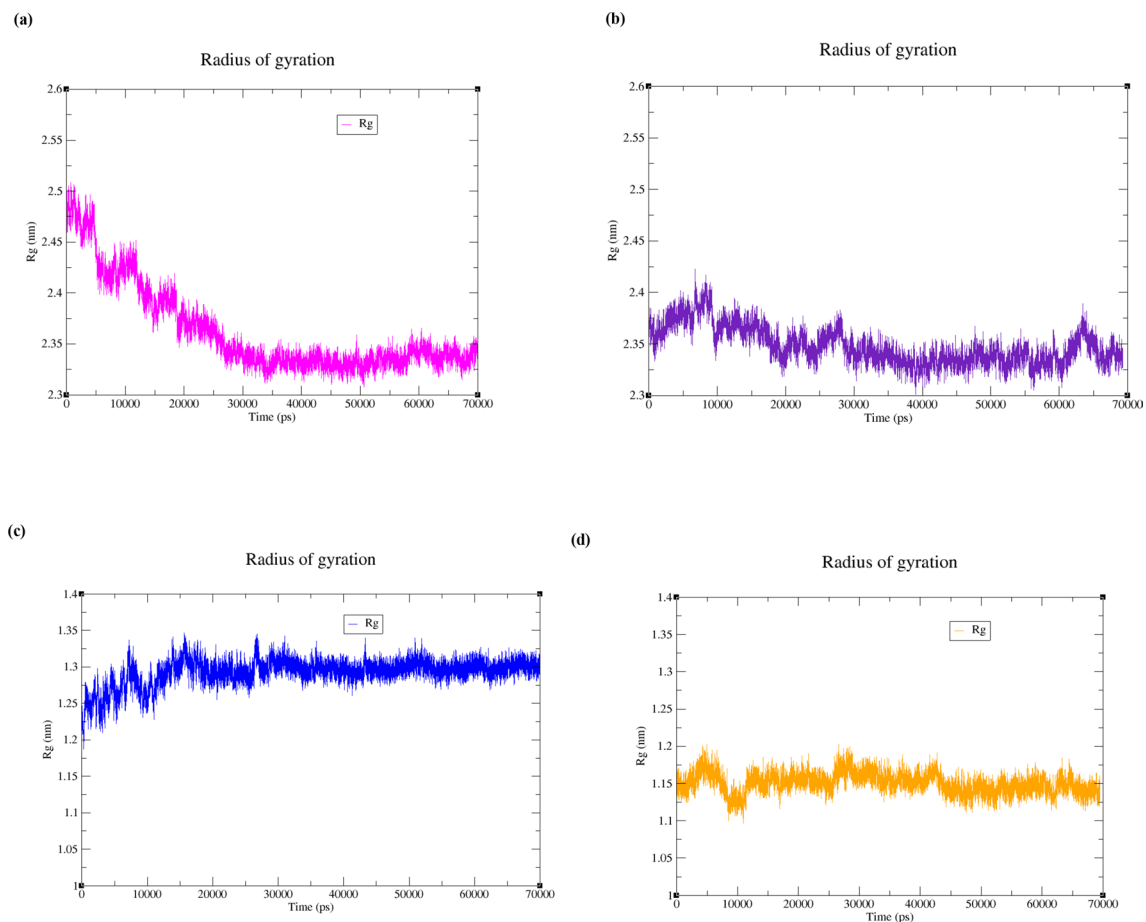
**Figure 9.** RMSF plots of the protein backbone Pol $\eta$  complexed with CNP (series 1, blue plot) and UNK4 (series 2, red plot) during 70 ns MD simulation.

residues 17–87), and thumb (including the residues 289–378) as well as the little finger domain of polymerase  $\eta$  or polymerase associated domain (PAD: the residues 319–432), which especially belongs to Y-family polymerases<sup>9</sup>. As discussed above, in Fig. 6a,b, the main residues of Pol $\eta$  interacting with UNK4 and CNP are phenylalanine18, Isoleucine 48, and alanine 49, Isoleucine114, and aspartic acid 115. These residues mainly belong to the palm and finger domains of Pol $\eta$ . As seen in Fig. 9, in series 1, RMSF values of the residues 1–50 in the presence of CNP vary in the range of 0.116–0.49 nm. While in series 2, RMSF values of the residues 1–50 in the presence of UNK4 vary in the range of 0.085–0.41 nm. This analysis shows that the fingers and palm domains of polymerase  $\eta$  in CNP and UNK4 have similar fluctuations over 70 ns of MD simulation.

To evaluate the compactness of systems in MD simulations, we considered the changes in the radius of gyration (Rg) for the protein Pol $\eta$  complexed with CNP (Fig. 10a) and UNK4 (Fig. 10b). According to Fig. 10a, the radius of gyrations of polymerase  $\eta$  in the presence of CNP is in the range of 2.35–2.50 nm. Figure 10b depicts that the radius of gyrations of polymerase  $\eta$  in the presence of UNK4 is in the range of 2.3–2.40 nm. Figure 10c,d represent the change of radius of gyration of DNA in the presence of CNP and UNK4, which is in the range of 1.25–1.325 nm for the system CNP (Fig. 10c) and is in the range of 1.1–1.2 in the system UNK4, (Fig. 10d).

As seen in Fig. 11a,b, the minimum distances between polymerase  $\eta$  and CNP (Fig. 11a) and UNK4 (Fig. 11b) during 70 ns of MD simulations are acceptable in both systems, as they are < 0.25 nm in most of the frames. Besides, under 6 Å, the number of contacts between polymerase  $\eta$  and UNK4 (Fig. 11d) is more than that between polymerase  $\eta$  and CNP (Fig. 11c). The intermolecular hydrogen bond analyses (Fig. 11e,f) display a possibility of forming a maximum of 7 and 8 hydrogen bonds between polymerase  $\eta$  and UNK4 (Fig. 11f). In contrast, the maximum number of hydrogen bonds between polymerase  $\eta$  and CNP is estimated to be 5 and 4 (Fig. 11e).

As seen in Fig. 12a,b, the minimum distances between DNA and CNP (Fig. 12a) as well as between DNA and UNK4 (Fig. 12b) during 70 ns of MD simulations are < 0.25 nm in most of the frames. Figure 12c,d show that the number of contacts between DNA and UNK4 is more than that between DNA and CNP. The intermolecular



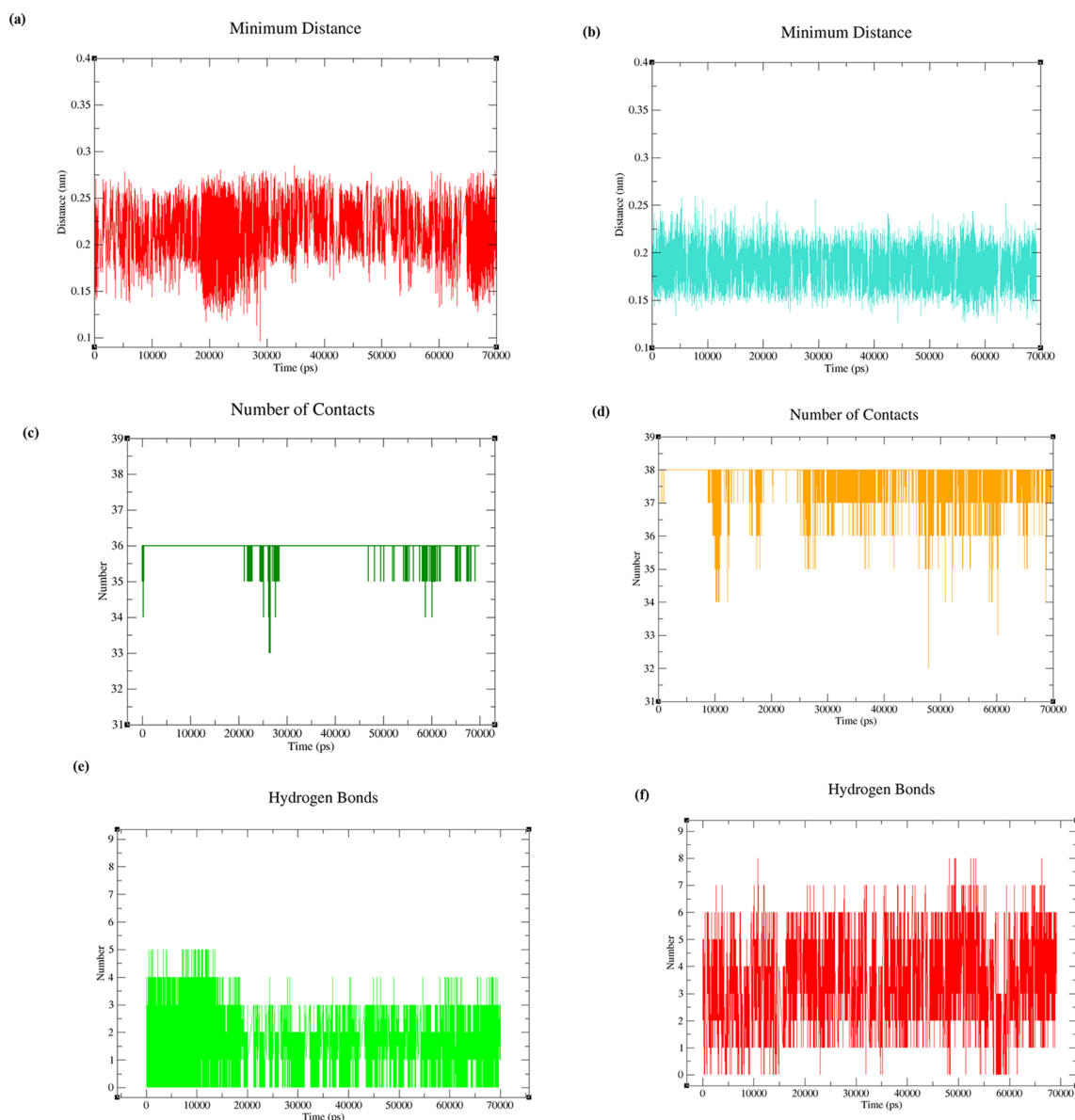
**Figure 10.** (a) Rg plot of protein in the presence of CNP. (b) Rg plot of protein in the presence of UNK4. (c) Rg plot of DNA strands in the presence of CNP. (d) Rg plot of DNA strands in the presence of UNK4.

hydrogen bond analyses (Fig. 12e,f) show a maximum of 5 hydrogen bonds between DNA and CNP (Fig. 12e) and a maximum number of 6 hydrogen bonds between DNA and UNK4 (Fig. 12f).

To ensure that the structure of UNK4 was reliable during 70 ns of molecular dynamics simulation, the RMSD of UNK4 with respect to the optimized geometry of the lowest energy conformer of UNK4 at B3LYP/6-311++G (d, p) level was plotted against time (ps). As seen in Fig. S4, the RMSD plots for UNK4 indicates that the backbone of this drug has fluctuated in the range of 0–0.1 nm, which illustrates that this ligand has been correctly parametrized for MD simulation. In addition, eight snapshots during the first 30 ns of MD simulation were taken for both CNP (Fig. S5) and UNK4 (Fig. S6) complexed with Pol $\eta$  and DNA. According to Figs. S5 and S6, UNK4, compared to CNP, can have a more prominent effect on the complex of Pol $\eta$  and DNA. In these Figs. S5 and S6, Mg ions are shown in pink color, and chloride ions are yellow. Two strands are shown in cyan and blue colors. H-bond model in red displays solvent (water molecules), Pol $\eta$  is shown using the new cartoon style, and both UNK4 and CNP are shown by the ball and stick model.

**MM/PBSA calculation of UNK4 and CNP on receptors (DNA and polymerase  $\eta$ ).** The binding free energies of UNK4 and CNP on the receptor (DNA and polymerase  $\eta$ ) were obtained by considering the Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PBSA) method during a 70 ns period of MD trajectory (for these calculations, 179 snapshots in both mentioned systems were chosen). The energy terms calculated by the *g\_mmpbsa* package are listed in Table 3<sup>64</sup>. MM/PBSA results illustrate that the binding energy of UNK4 to DNA (–70.14 kJ/mol) is more negative than that of CNP to DNA (–55.81 kJ/mol). Additionally, according to our calculations, UNK4 can bind to the polymerase  $\eta$  with the binding energy of –144.99 kJ/mol, indicating the stronger binding of UNK4 with polymerase  $\eta$ , as compared with the binding energy of CNP with polymerase  $\eta$  (–125.16 kJ/mol). According to our results, the binding energy values of both CNP and UNK4 are negative, and their interactions with the receptor (DNA and polymerase  $\eta$ ) are thermodynamically very favourable. As a result, UNK4, with the volume, area, and electron density similar to CNP, could be considered an alternative anticancer drug for Cytarabine with much fewer side effects because its building blocks are amino acid and cyclic carbohydrate.

**Evaluations of the validity of our theoretical calculations with respects to the experimental data.** To validate our theoretical studies on UNK4 as an anticancer drug, we studied the interaction of UNK4

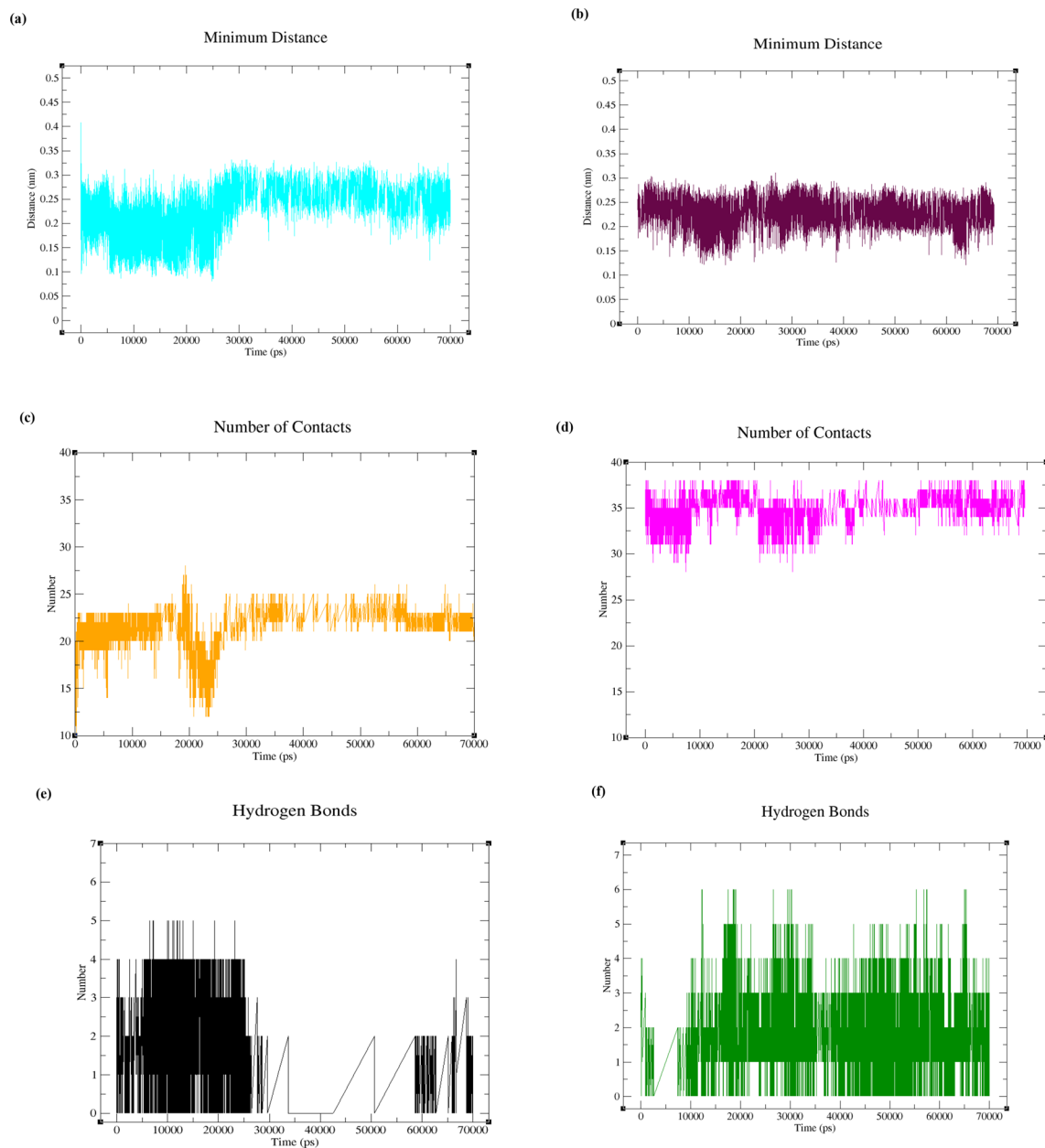


**Figure 11.** Plots of the protein backbone (pol $\eta$ ) complexed with CNP and UNK4 during 70 ns MD simulation: (a) plot of average of the minimum distance between pol $\eta$  and CNP, (b) plot of average of the minimum distance between pol $\eta$  and UNK4, (c) plot of number of contacts between pol $\eta$  and CNP, (d) plot of number of contacts between pol $\eta$  and UNK4. The intermolecular hydrogen-bonding distribution plot versus time for (e) pol $\eta$ -CNP and (f) pol $\eta$ -UNK4.

with SI/II-pocket of KRAS protein using Auto Dock/Vina. The experimental data of the complexes of this protein with different drugs protein are reported with the PDB ID codes, including 6GJ5, 6GJ6, 6GJ7, and 6GJ8. In Table 4, entries 1–4 indicate the structures of inhibitors (drugs) of the SI/II-pocket of KRAS protein extracted from their PDB files. Dirk Kessler et al. first introduced these inhibitors<sup>65</sup>. The SI/II-pocket of KRAS protein was selected for three reasons. First, KRAS protein plays an essential role in activating various signaling molecules that permit the transduction of signals from the cell surface to the nucleus. In turn, these processes can control different stages of the cell cycle, such as chemotaxis<sup>66</sup>, growth, and apoptosis<sup>67,68</sup>. Secondly, KRAS with the mutation is known as a prominent driver of human cancers<sup>69</sup>. Thirdly, the inhibitors indicated in entries 1–4 in Table 4 contain indole substructures, and they have been reported to have nanomolar binding affinity to the SI/II-pocket of KRAS protein<sup>65</sup>. As UNK4 has indole fragment in its structure, we examined UNK4 inhibitory activities against KRAS to support our theoretical results.

The experimental binding affinities of KRAS toward these known inhibitors were estimated from measured dissociation constants  $K_D$ <sup>70</sup>, and by employing Eqs. (3) and (4).

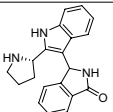
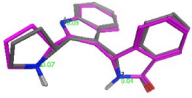
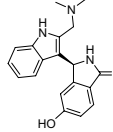
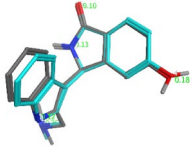
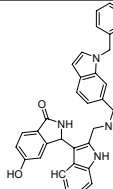
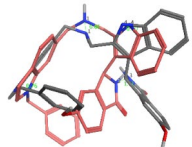
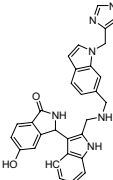
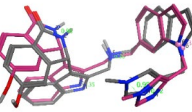
$$\Delta G_{\text{Dissociation}} = -RT \ln K_D \quad (3)$$



**Figure 12.** Plots of the DNA backbone complexed with CNP and UNK4 complexes during 70 ns MD simulation: (a) plot of average of the minimum distance between DNA and CNP, (b) plot of average of the minimum distance between DNA and UNK4, (c) plot of number of contacts between DNA and CNP, (d) plot of number of contacts between DNA and UNK4. The intermolecular hydrogen-bonding distribution plot versus time for (e) DNA–CNP and (f) DNA–UNK4.

MM/PBSA	DNA and CNP	polymerase $\eta$ and CNP	DNA and UNK4	polymerase $\eta$ and UNK4
van der Waals energy (kJ/mol)	−67.02	−217.02	−80.75	−158.17
Electrostatic energy(kJ/mol)	−2.15	−11.32	−6.57	−146.62
Polar solvation energy(kJ/mol)	19.25	117.84	24.57	175.83
Non-polar solvation energy (kJ/mol)	−5.89	−14.66	−7.39	−16.03
Binding energy (kJ/mol)	−55.81 ± 17.58	−125.16 ± 35.68	−70.14 ± 10.74	−144.99 ± 18.94

**Table 3.** Various energy terms obtained through MM/PBSA calculations.

Entry	Ligand	The structure of ligand	PDB ID	Receptor	$K_D$ ( $\mu$ M) (experimental)	$\Delta G_{\text{Binding}}$ (kcal/mol) (experimental)	$\Delta G_{\text{Binding}}$ (kcal/mol) Auto dock vina (theoretical)	RMSD (Calculated with MOE) ( $\text{\AA}$ )	RMSD (best pose)
1	3-(2-((S)-pyrrolidin-2-yl)-1H-indol-3-yl)isoindolin-1-one		6GI5	KRAS <sup>G12D</sup>	440	-4.58	-6.6	0.070	
2	(S)-3-(2-((dimethylamino)methyl)-1H-indol-3-yl)-5-hydroxyisoindolin-1-one		6GI6	KRAS <sup>G12D</sup>	17	-6.5	-8.3	0.168	
3	3-(2-(((1-benzyl-1H-indol-6-yl) methyl) amino) methyl)-1H-indol-3-yl)-5-hydroxyisoindolin-1-one		6GI7	KRAS <sup>G12D</sup>	0.87	-8.2	-9.2	0.500	
4	5-hydroxy-3-(2-(((1-(1-methyl-1H-imidazol-4-yl) methyl)-1H-indol-6-yl) methyl) amino) methyl)-1H-indol-3-yl)isoindolin-1-one		6GI8	KRAS <sup>G12D</sup>	0.74	-8.35	-8.4	0.600	

**Table 4.** The calculated and experimental  $\Delta G_{\text{Binding}}$  (kcal/mol) values of some known compounds toward KRAS protein.

$$\Delta G_{\text{Dissociation}} = -\Delta G_{\text{Binding}} \quad (4)$$

In this part, we calculated the binding affinity of the SI/II-pocket of KRAS protein toward these inhibitors (entries 1–4 in Table 4) and UNK4 using the Auto Dock/Vina plugin with scripts from the Auto Dock Tools Package. We then obtained the linear graph shown in Fig. 13 by plotting these calculated binding affinities versus their experimentally binding affinities (see Table 4 for calculated and experimental binding affinity values). Using the linear equation obtained from this linear graph ( $y = 0.564x - 4.23$ ,  $R^2 = 0.828$ ) and the binding affinity of KRAS for UNK4 obtained by Auto Dock/Vina, we predicted the experimental binding affinity of KRAS for UNK4 to be  $-8.1$  kcal/mol, which is in good agreement with the binding affinity of KRAS for UNK4 calculated by Auto Dock/Vina ( $-6.9$  kcal/mol).

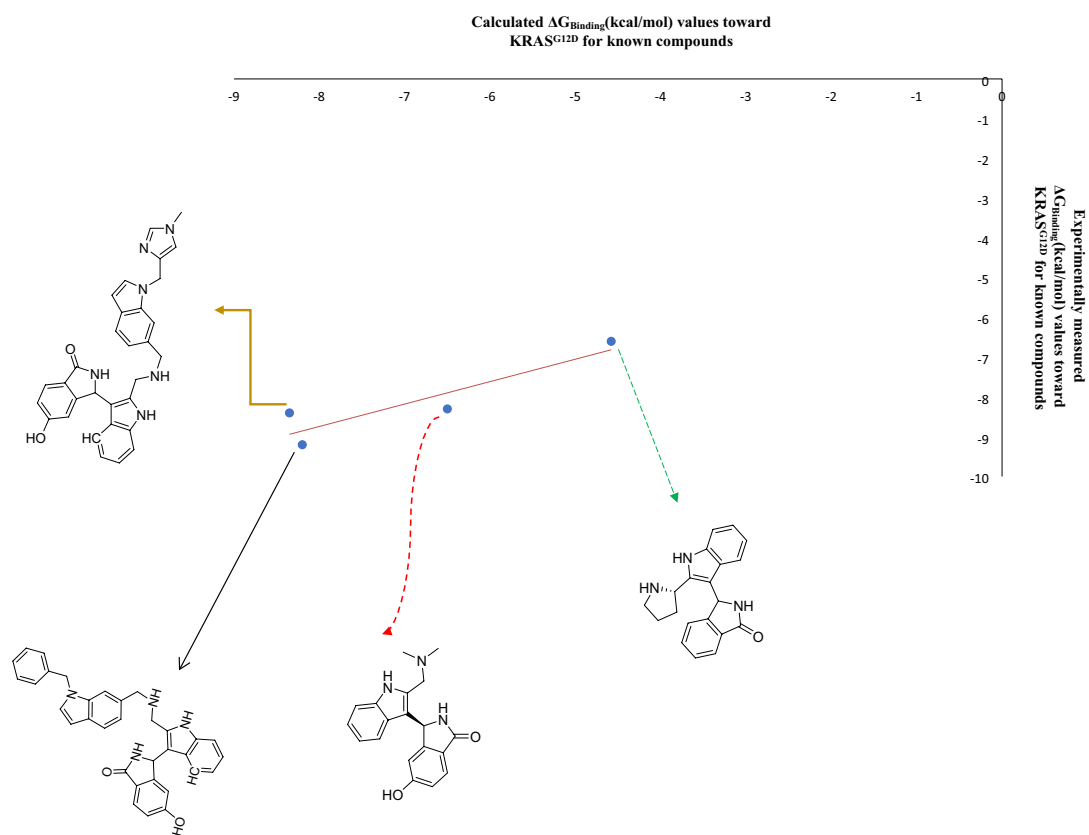
### Synthesis of UNK4

In this section, we used the approach reported by Chanda et al.<sup>32</sup> for the synthesis of UNK4 (Table 1, entry 6). In a round bottom flask containing 7 mL MeOH and 0.5 mL of acetic acid, tryptophan (1 mmol) was added, and D-fructose (1 mmol) was loaded into this flask. Proceeding this stage, Zn (OAc)<sub>2</sub> (1 mg) was added to the reaction environment followed by the reflux of the reaction mixture between 6 and 12 h. During the progress of the mentioned reaction, the mixture turned reddish-brown colour. In the last steps the received solution was filtered through Sintered glass filter disc under the influence of a partial vacuum. After being concentrated with the aid of a rotary evaporator, the filtrate was washed with acetone four times (20 mL in total), and the obtained crystal was again decanted with water and distilled ethyl acetate. Then, the organic phase was separated from the aqueous phase and subjected to vacuum. The recrystallized product was dried under vacuum. The FT-IR spectrum and H-NMR data of the product UNK4 are given in Fig. S7 and Table S2. In the next section, we report the measured log P of the synthesized UNK4.

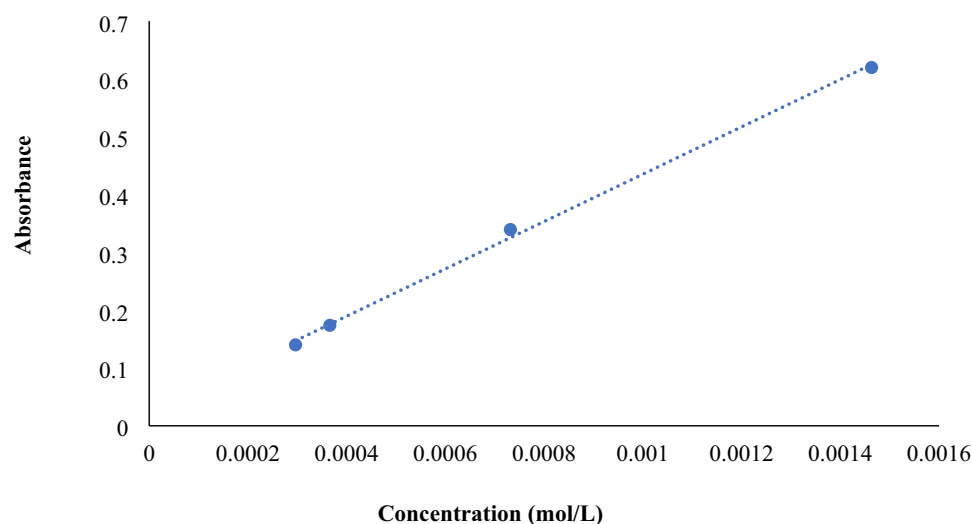
### Experimental measurement of log P of UNK4

To obtain log P of UNK4, this compound was dissolved in water to make the solutions with known concentrations, including 0.0002964 M, 0.0003658 M, 0.0007316 M and 0.0014632 M. Subsequently, the absorptions of the mentioned solutions were scanned between 200 and 800 nm. Then, the calibration curve for UNK4 was plotted at the  $\lambda_{\text{max}}$  280 nm. Figure 14 illustrates the UV calibration curve with  $y = 409.4x + 0.0258$ ,  $R^2 = 0.9981$ . The y-axis in Fig. 14 displays the absorbance of each solution, and the x-axis depicts their concentrations. The UV spectra of the solutions mentioned above can be found in Fig. S8.

With a calibration curve in hand, 10 mL of the UNK4 solutions with the concentrations of 0.0014632 M, 0.0002964 M and 0.0007316 were added to 3 separatory funnels. Then, 10 mL of 1-Octanol was loaded into each of the funnels and the funnels were shaken. Then, mixture in each funnel was allowed to achieve equilibrium in 2



**Figure 13.** Plot of the experimental  $\Delta G_{\text{Binding}}$  (kcal/mol) values against the calculated  $\Delta G_{\text{Binding}}$  (kcal/mol) values for the known KRAS inhibitors.



**Figure 14.** The UV calibration curve of UNK4.

phases. Then absorbance of each aqueous solution was again obtained at the  $\lambda_{\text{max}}$  280 nm to measure concentration of UNK4 in water using the calibration curve described above (Fig. 14 and Table S3). Using the Eqs. (5)–(8)<sup>71</sup>, we measured the experimental value of log P for UNK4. In Eq. (5),  $C_{\text{aq, Init}}$  represents the initial concentration of the aqueous solution of UNK4, while  $C_{\text{aq, eq}}$  and  $C_{\text{Oct, eq}}$  show the concentrations of UNK4 in the aqueous and octanol solutions after equilibration, respectively;  $V_{\text{Oct}}$  and  $V_{\text{aq}}$  are the epithets of the volumes of aqueous solution and 1-Octanol in the separatory funnels. According to the mass balance, the relationship between  $C_{\text{aq, Init}}$ ,  $C_{\text{aq, eq}}$  and  $C_{\text{Oct, eq}}$  can be explained by Eq. (5). The Eq. (6) can result from the Eq. (5) and  $K_{\text{ow}}$ <sup>72</sup> represents the ratio of

the concentrations of a compound between 1-Octanol and water, which can be expressed as Eq. (7). Using the Eqs. (7) and (8), log P of the UNK4 solutions were measured to be  $-0.78$ ,  $-1.61$  and  $-1.52$  with the average value of  $-1.3$ . A good correlation exists between this experimentally measured log P ( $-1.3$ ) and theoretically calculated log P of UNK4 ( $-1.64$ ) obtained in Section “[Prediction of log P for the designed anticancer structure](#)”.

$$C_{aq,init} V_{aq} = C_{aq,eq} V_{aq} + C_{Oct,eq} V_{Oct} \quad (5)$$

$$V_{aq} = V_{Oct}$$

$$C_{aq,init} - C_{aq,eq} = C_{Oct,eq} \quad (6)$$

$$K_{OW} = \frac{C_{Oct,eq}}{C_{aq,eq}} = \frac{C_{aq,init} - C_{aq,eq}}{C_{aq,eq}} \quad (7)$$

$$\log K_{OW} = \log P \quad (8)$$

The comparison of our experimentally measured log P of UNK4 ( $-1.3$ ) with the experimentally reported log P of Cytarabine ( $-2.8$ )<sup>51</sup> indicate that the suggested structure of UNK4 is comparable to the main drug (Cytarabine) in terms of cell membrane permeability.

## Conclusion

According to many studies, carbohydrates are involved in many cells' recognition processes<sup>73–75</sup>. These carbohydrate structures bear high polarities that can affect their pharmacokinetic properties as a drug. In this study, we designed the new structures containing both carbohydrates and amino acids and explored their interactions with the combined receptor containing the DNA and polymerase  $\eta$  ( $pol\eta$ ). We chose small molecules composed of D-fructose and amino acids such as valine, arginine, isoleucine, leucine, tryptophan, alanine, and phenylalanine. These small designed molecules should have acceptable solvation in both lipophilic and hydrophilic media. Therefore, they might achieve their targets in the human cells quickly. These newly designed amino acid-carbohydrate-based structures should have fewer toxic effects in high dosage as new anticancer drugs. We used in silico studies to find the best candidates for these newly designed drugs. Our results indicated that designed UNK4 structure (Table 1) could be a suitable replacement for Cytarabine or other traditional drugs used to inhibit  $pol\eta$ , which plays an essential role in the replication of damaged DNA in the cancerous cell and leads to the formation of daughter cells having the genome with some lesions. This process will be repeated over and over and result in the accumulation of mutations and tumorigenesis. To stop this replication process in the tumor cells, in this study, both DNA and  $pol\eta$  were considered the combined receptor. As Cytarabine is one of the well-known drugs to stop the replication process of DNA in tumor cells, the first goal was to design a structure most similar to Cytarabine, but with fewer side effects. Thus, the physical properties such as volumes, areas, and the electrostatic potential maps of the newly designed amino acid-carbohydrate-based structures were obtained and compared with those of Cytarabine. We selected the candidate structures which were very similar to Cytarabine in terms of these physical properties. Then, the docking procedures were performed to predict the bound conformations and the binding affinities of Cytarabine and UNK4 with the combined receptor of DNA and  $pol\eta$  using the Auto Dock/Vina plugin with scripts from the Auto Dock Tools Package. It was found that UNK4 has an excellent binding affinity towards the receptor of DNA and  $pol\eta$  ( $-9.2$  kcal/mol), in comparison with the binding affinity of CNP toward this receptor ( $-8.8$  kcal/mol). DFT calculations in combination with a solvation model based on density (SMD) (refer to Table 2)<sup>49</sup> were employed to calculate log P values of Cytarabine and UNK4. The log P for UNK4 was found to be  $-1.64$  comparable to the reported experimental log P value of Cytarabine ( $-2.8$ ). Finally, MD simulations were carried out for the CNP and UNK4 complexed with their receptor (DNA and  $Pol\eta$ ) to comprehend their inhibition capability better to replicate DNA. MD simulations and its energy analysis performed by MM/PBSA calculations indicated that UNK4, containing the natural building blocks of amino acid and carbohydrate, could be a promising inhibitor for the combined receptor of DNA and  $pol\eta$ , and thus a suitable replacement for Cytarabine.

## Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 12 August 2022; Accepted: 19 October 2022

Published online: 02 November 2022

## References

1. Waters, L. S. *et al.* Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.* **73**, 134–154. <https://doi.org/10.1128/MMBR.00034-08> (2009).
2. Barnes, R. P., Tsao, W. C., Moldovan, G. L. & Eckert, K. A. DNA polymerase  $\eta$  prevents tumor cell-cycle arrest and cell death during recovery from replication stress. *Cancer Res.* **78**, 6549–6560. <https://doi.org/10.1158/0008-5472.CAN-17-3931> (2018).
3. Sale, J. E., Lehmann, A. R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **13**, 141–152. <https://doi.org/10.1038/nrm3289> (2012).
4. Lange, S. S., Takata, K. I. & Wood, R. D. DNA polymerases and cancer. *Nat. Rev. Cancer* **11**, 96–110. <https://doi.org/10.1038/nrc998> (2011).



5. Masutani, C., Kusumoto, R., Iwai, S. & Hanaoka, F. Mechanisms of accurate translesion synthesis by human DNA polymerase  $\eta$ . *EMBO J.* **19**, 3100–3109. <https://doi.org/10.1093/emboj/19.12.3100> (2000).
6. Ceppi, P. *et al.* Polymerase  $\eta$  mRNA expression predicts survival of non-small cell lung cancer patients treated with platinum-based chemotherapy. *Clin. Cancer Res.* **15**, 1039–1045. <https://doi.org/10.1158/1078-0432.CCR-08-1227> (2009).
7. Ummat, A. *et al.* Structural basis for cisplatin DNA damage tolerance by human polymerase  $\eta$  during cancer chemotherapy. *Nat. Struct. Mol. Biol.* **19**, 628. <https://doi.org/10.1038/nsmb.2295> (2012).
8. Zhao, Y. *et al.* Structural basis of human DNA polymerase  $\eta$ -mediated chemoresistance to cisplatin. *PNAS USA* **109**, 7269–7274. <https://doi.org/10.1073/pnas.1202681109> (2012).
9. Rechtkoblit, O. *et al.* Structural basis for polymerase  $\eta$ -promoted resistance to the anticancer nucleoside analog cytarabine. *Sci. Rep.* **8**, 1–9. <https://doi.org/10.1038/s41598-018-30796-w> (2018).
10. Wiley, J. S., Jones, S. P., Sawyer, W. H. & Paterson, A. R. P. Cytosine arabinoside influx and nucleoside transport sites in acute leukemia. *J. Clin. Investig.* **69**, 479–489. <https://doi.org/10.1172/JCI110472> (1982).
11. Paterson, A. R. P. & Oliver, J. M. Nucleoside transport. II. Inhibition by p-nitrobenzylthioguanosine and related compounds. *Can. J. Biochem. Physiol.* **49**, 271–274. <https://doi.org/10.1139/o71-039> (1971).
12. Goodell, B., Leventhal, B. & Henderson, E. Cytosine arabinoside in acute granulocytic leukemia. *Clin. Pharmacol. Ther.* **12**, 599–606. <https://doi.org/10.1002/cpt.1971124599> (1971).
13. Rassiga, A. L., Schwartz, H. J., Forman, W. B. & Crum, E. D. Cytarabine-induced anaphylaxis: demonstration of antibody and successful desensitization. *Arch. Intern. Med.* **140**, 425–426. <https://doi.org/10.1001/archinte.1980.0033015013903> (1980).
14. Pizzo, P. A., Robichaud, K. J., Gill, F. A. & Witebsky, F. G. Empiric antibiotic and antifungal therapy for cancer patients with prolonged fever and granulocytopenia. *Am. J. Med.* **72**, 101–111. [https://doi.org/10.1016/0002-9343\(82\)90594-0](https://doi.org/10.1016/0002-9343(82)90594-0) (1982).
15. Burgdorf, W. H., Gilmore, W. A. & Ganick, R. G. Peculiar acral erythema secondary to high-dose chemotherapy for acute myelogenous leukemia. *Ann. Intern. Med.* **97**, 61–62. <https://doi.org/10.7326/0003-4819-97-1-61> (1982).
16. Jehn, U., Gödel, N., Rienmüller, R. & Wilmanns, W. Non-cardiogenic pulmonary edema complicating intermediate and high-dose Ara C treatment for relapsed acute leukemia. *Med. Oncol. Tumor Pharmacother.* **5**, 41–47. <https://doi.org/10.1007/BF03003180> (1988).
17. Donehower, R. C., Karp, J. E. & Burke, P. J. Pharmacology and toxicity of high-dose cytarabine by 72-hour continuous infusion 1, 2. *Cancer Treat. Rep.* **70**, 1059 (1986) (PMID: 3461882).
18. Congreve, M., Murray, C. W. & Blundell, T. L. Keynote review: Structural biology and drug discovery. *Drug Discov. Today* **10**, 895–907. [https://doi.org/10.1016/S1359-6446\(05\)03484-7](https://doi.org/10.1016/S1359-6446(05)03484-7) (2005).
19. Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S. & Wootton, R. Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.* **57**, 201–209. <https://doi.org/10.1111/j.1476-5381> (1976).
20. Gruner, S. A., Locardi, E., Lohof, E. & Kessler, H. Carbohydrate-based mimetics in drug design: Sugar amino acids and carbohydrate scaffolds. *Chem. Rev.* **102**, 491–514. <https://doi.org/10.1021/cr0004409> (2002).
21. Hirschmann, R. *et al.* Modulation of receptor and receptor subtype affinities using diastereomeric and enantiomeric monosaccharide scaffolds as a means to structural and biological diversity. *J. Med. Chem.* **41**, 1382–1391. <https://doi.org/10.1021/jm980346> (1998).
22. Blocks, C. B. & Bols, B. M. *Carbohydrate Building Blocks* (Wiley, 1996).
23. Wunberg, T. *et al.* Carbohydrates as multifunctional chiral scaffolds in combinatorial synthesis. *Angew. Chem.* **37**, 2503–2505. [https://doi.org/10.1002/\(SICI\)1521-3773\(19981002\)37:18%3C2503::AID-ANIE2503%3E3.0.CO;2-R\(1998\)](https://doi.org/10.1002/(SICI)1521-3773(19981002)37:18%3C2503::AID-ANIE2503%3E3.0.CO;2-R(1998)) (1998).
24. Schweizer, F. & Hindsgaul, O. Combinatorial synthesis of carbohydrates. *Curr. Opin. Chem. Biol.* **3**, 291–298. [https://doi.org/10.1016/s1367-5931\(99\)80045-3](https://doi.org/10.1016/s1367-5931(99)80045-3) (1999).
25. Bednarska, N. G., Wren, B. W. & Willcocks, S. J. The importance of the glycosylation of antimicrobial peptides: natural and synthetic approaches. *Drug Discov. Today* **22**, 919–926. <https://doi.org/10.1016/j.drudis.2017.02.001> (2017).
26. Varamini, P. *et al.* Synthesis and biological evaluation of an orally active glycosylated endomorphin-1. *J. Med. Chem.* **55**, 5859–5867. <https://doi.org/10.1021/jm300418d> (2012).
27. Polt, R., Dhanasekaran, M. & Keyari, C. M. Glycosylated neuropeptides: A new vista for neuropsychopharmacology?. *Med. Res. Rev.* **25**, 557–585. <https://doi.org/10.1002/med.20039> (2005).
28. Ho, H. H., Gilbert, M. T., Nussenzveig, D. R. & Gershengorn, M. C. Glycosylation is important for binding to human calcitonin receptors. *Biochemistry* **38**, 1866–1872. <https://doi.org/10.1021/bi981195e> (1999).
29. Herzner, H., Reipen, T., Schultz, M. & Kunz, H. Synthesis of glycopeptides containing carbohydrate and peptide recognition motifs. *Chem. Rev.* **100**, 4495–4538. <https://doi.org/10.1021/cr990308c> (2000).
30. Grant, S. Ara-C: Cellular and molecular pharmacology. *Adv. Cancer Res.* **72**, 197–233. <https://doi.org/10.1016/S0065-230X> (1997).
31. Maeda, H., Kusuhara, T., Tsubako, M. & Nakayama, H. Phosphorylation of 5'-deoxy-5-fluorouridine with inorganic phosphorylating agents. *Chem. Pharm. Bull.* **56**, 1698–1703. <https://doi.org/10.1248/cpb.59.1447> (2008).
32. Chanda, D. & Harohally, N. V. Revisiting Amadori and Heyns synthesis: Critical percentage of acyclic form play the trick in addition to catalyst. *Tetrahedron Lett.* **59**, 2983–2988. <https://doi.org/10.1016/j.tetlet.2018.06.050> (2018).
33. Shao, Y., Fusti-Molnar, L., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S. T., Gilbert, A. T. B., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P. & DiStasio Jr, R. A. Wavefunct. Inc. (Irvine CA, 2011).
34. 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.* **31**, 455–461. <https://doi.org/10.1002/jcc.21334> (2010).
35. Seeliger, D. & de Groot, B. L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* **24**, 417–422. <https://doi.org/10.1007/s10822-010-9352-6> (2010).
36. Biovia, D. S. *BIOVIA Discovery Studio Client*, (v16. 1.0. 15350.) (Dassault Systems, San Diego, 2017).
37. Páll, S., Abraham, M. J., Kutzner, C., Hess, B. & Lindahl, E. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *EASCS* [https://doi.org/10.1007/978-3-319-15976-8\\_1](https://doi.org/10.1007/978-3-319-15976-8_1) (2014).
38. Kukol, A. (ed.) *Molecular Modeling of Proteins* Vol. 443 (Humana Press, 2008).
39. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Gr.* **14**, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) (1996).
40. Huey, R. & Morris, G. M. *Using AutoDock 4 with AutoDocktools: A Tutorial* 54–56 (The Scripps Research Institute, USA, 2008).
41. Garcia-Diaz, M., Murray, M. S., Kunkel, T. A. & Chou, K. M. Interaction between DNA polymerase  $\lambda$  and anticancer nucleoside analogs. *J. Biol. Chem.* **285**, 16874–16879. <https://doi.org/10.1074/jbc.M109.094391> (2010).
42. Garcia-Diaz, M. *et al.* DNA polymerase  $\lambda$ , a novel DNA repair enzyme in human cells. *J. Biol. Chem.* **277**, 13184–13191. <https://doi.org/10.1074/jbc.M111601200> (2002).
43. Sabini, E. *et al.* Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat. Struct. Mol. Biol.* **10**, 513–519. <https://doi.org/10.1038/nsb942> (2003).
44. Mannhold, R. *et al.* (eds) *Molecular Drug Properties: Measurement and Prediction* Vol. 37 (Wiley-VCH, 2008).
45. Lemkul, J. A. & Bevan, D. R. Destabilizing Alzheimer's A $\beta$ 42 protofibrils with morin: mechanistic insights from molecular dynamics simulations. *Biochemistry* **49**, 3935–3946. <https://doi.org/10.1021/bi1000855> (2010).
46. Nedyalkova, M. A., Madurga, S., Tobiszewski, M. & Simeonov, V. Calculating the partition coefficients of organic solvents in octanol/water and octanol/air. *J. Chem. Inf. Model.* **59**, 2257–2263. <https://doi.org/10.1021/acs.jcim.9b00212> (2019).

47. Hoekman, D. Exploring QSAR fundamentals and applications in chemistry and biology, volume 1. Hydrophobic, electronic and steric constants. *J. Am. Chem. Soc.* **118**, 10678–10678. <https://doi.org/10.1021/ja965433%2B> (1996).
48. Labnetwork.com. *LabNetwork*. <https://www.labnetwork.com/frontend-app/p/#!/moleculerdetails/LN00008170> (2021).
49. Sangster, J. A databank of evaluated octanol-water partition coefficients (LogP) on microcomputer diskette. Bulletin for Sangster Research Laboratories. Canadian National Committee for CODATA: Montreal, Quebec, Canada (1994).
50. Thompson, M. Material safety data sheets. *Am. J. Nurs.* **110**, 12–14. <https://doi.org/10.1097/01.NAJ.0000390501.84780.e6> (2010).
51. Schaduagrat, N. *et al.* Towards reproducible computational drug discovery. *J. Cheminform.* **12**, 9. <https://doi.org/10.1186/s13321-020-0408-x> (2020).
52. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B.* **113**, 6378–6396. <https://doi.org/10.1021/jp810292n> (2009).
53. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241. <https://doi.org/10.1007/s00214-007-0310-x> (2008).
54. Barclay, T., Ginic-Markovic, M., Johnston, M. R., Cooper, P. & Petrovsky, N. Observation of the keto tautomer of D-fructose in D2O using 1H NMR spectroscopy. *Carbohydr. Res.* **347**, 136–141. <https://doi.org/10.1016/j.carres.2011.11.003> (2012).
55. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791. <https://doi.org/10.1002/jcc.21256> (2009).
56. Mansourian, M., Mahnam, K., Madadkar-Sobhani, A., Fassihi, A. & Saghaie, L. Insights into the human A1 adenosine receptor from molecular dynamics simulation: Structural study in the presence of lipid membrane. *Med. Chem. Res.* **24**, 3645–3659. <https://doi.org/10.1007/s00044-015-1409-6> (2015).
57. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14%3c1639::AID-JCC10%3e3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14%3c1639::AID-JCC10%3e3.0.CO;2-B) (1998).
58. Davydov, A. S. *Solitons in Molecular Systems* 113 (Reidel, Dordrecht, 1985).
59. Yadava, U., Shukla, B. K., Roychoudhury, M. & Kumar, D. Pyrazolo [3, 4-d] pyrimidines as novel inhibitors of O-acetyl-L-serine sulphydrylase of *Entamoeba histolytica*: An in-silico study. *J. Mol. Model.* **21**, 96. <https://doi.org/10.1007/s00894-015-2631-3> (2015).
60. Berendsen, H. J., Postma, J. V., van Gunsteren, W. F., DiNola, A. R. H. J. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690. <https://doi.org/10.1063/1.448118> (1984).
61. Yadav, R. K. & Yadava, U. Molecular dynamics simulation of DNA duplex, analog of PPT (polypurine tract), its conformation and hydration: A theoretical study. *Med. Chem. Res.* **23**, 280–286. <https://doi.org/10.1007/s00044-013-0631-3> (2014).
62. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089. <https://doi.org/10.1063/1.464397> (1993).
63. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
64. Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, & Lynn, A. g\_mmpbsa: A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**, 1951–1962. <https://doi.org/10.1021/ci500020m> (2014).
65. Kessler, D. *et al.* Drugging an undruggable pocket on KRAS. *PNAS USA* **116**, 15823–15829. <https://doi.org/10.1073/pnas.1904529116> (2019).
66. Karnoub, A. E. *et al.* Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* **449**, 557–563. <https://doi.org/10.1038/nature06188> (2007).
67. Esser, D. *et al.* Structure determination of the Ras-binding domain of the Ral-specific guanine nucleotide exchange factor Rlf. *Biochemistry* **37**, 13453–13462. <https://doi.org/10.1021/bi9811664> (1998).
68. Zuber, J. *et al.* A genome-wide survey of RAS transformation targets. *Nat. Genet.* **24**, 144–152. <https://doi.org/10.1038/72799> (2000).
69. Biankin, A. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405. <https://doi.org/10.1038/nature11547> (2012).
70. Varki, A., Kannagi, R. & Toole, B. P. *Glycosylation Changes in Cancer. Essentials of Glycobiology* 2nd edn. (Cold Spring Harbor Laboratory Press, 2009).
71. Brittain, H. G. *Profiles of Drug Substances, Excipients, and Related Methodology* (Academic press, 2020).
72. Moldoveanu, S. C. & David, V. *Modern Sample Preparation for Chromatography* (Elsevier, 2021).
73. Van Kooyk, Y. & Rabinovich, G. A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **9**, 593–601. <https://doi.org/10.1038/ni.f.203> (2008).
74. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104. <https://doi.org/10.1039/B907931A> (2009).
75. Ernst, B. & Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nat. Rev. Drug Discov.* **8**, 661–677. <https://doi.org/10.1038/nrd2852> (2009).

## Acknowledgements

We would like to thank the Sharif High-Performance Computing (HPC) centre for providing the computational resources of this research.

## Author contributions

A.F. is the advisor of S.K. The S.K. and A.F. authors have the same contributions in terms of defining the project. S.K. prepared the first draft of manuscript. A.F. revised it. Both authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22810-z>.

**Correspondence** and requests for materials should be addressed to A.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022