

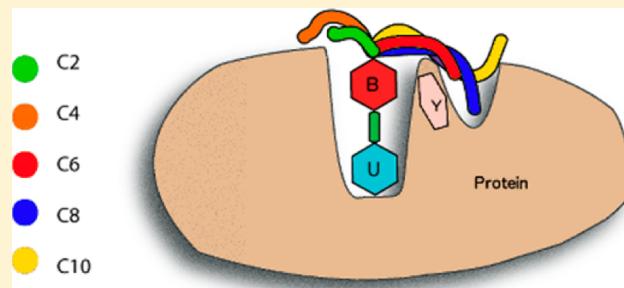
Computational Rationale for the Selective Inhibition of the Herpes Simplex Virus Type 1 Uracil-DNA Glycosylase Enzyme

Umraan Hendricks, Werner Crous, and Kevin J. Naidoo*

Scientific Computing Research Unit and Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

 Supporting Information

ABSTRACT: The herpes simplex virus uracil-DNA glycosylase (HSV-UNG) enzyme is responsible for the reactivation of the virus from latency and efficient viral replication in nerve tissue. The lack of uracil-DNA glycosylase enzyme in human neurons and the continuous deamination of cytosine create an environment where the presence of viral uracil-DNA glycosylase is a necessity for the proliferation of the virus. A series of 6-(4-alkylanilino)-uracil inhibitors has been developed that selectively and strongly binds to the HSV-UNG enzyme while weakly binding to human uracil-DNA glycosylase (hUNG). Here, by using a combination of sequence and structural comparisons between the two enzymes along with free energy of binding computations and principal component analysis of the ligands, we investigate and rationalize the inhibitory effect of the 6-(4-alkylanilino)-uracil series as a function of alkyl chain length on the HSV-UNG. The results of these computations corroborate the experimental finding that the inhibitor with an octyl aliphatic chain selectively binds HSV-UNG best. More importantly we find that 6-(4-octylanilino)-uracil's selective inhibition of HSV-UNG over hUNG is due to the combination of the solution preconfigured bent conformation of that specific chain length and the position of HIS92 (absent in hUNG) just outside HSV-UNG's hydrophobic gorge lying adjacent to its uracil binding pocket. The similarities between the uracil binding pockets in HSV-UNG and hUNG obfuscate an understanding of the preferential inhibition of the virus enzyme. However, the differences in the enzymes' shallow hydrophobic grooves adjacent to the binding pockets, such as the gorge we identify here, rationalizes 6-(4-alkylanilino)-uracil with an octyl chain length as an excellent pharmacophore template for HSV-UNG inhibitor design.



INTRODUCTION

During the latency stage of the herpes simplex virus type 1 (HSV-1), its DNA undergoes various mutations. The presence of the RNA base uracil in the HSV-1 DNA is a result of one of two mechanistic actions. Either deoxyuridine triphosphate (dUTP) is misincorporated during cellular replication or cytosine (C) is hydrolytically deaminated to produce a guanine-uracil (G:U) base pair mismatch.¹ If uracil remains in the HSV-1 DNA then a C → T transition mutation results in that DNA strand, and a G → A transition mutation occurs in the opposite strand of the DNA after the next round of replication. This leads to adenine-thymine (A:T) base pair mutations.

The process of uracil mutation correction in human DNA and the herpes simplex virus both occur via the base-excision repair pathway, which is initiated by a uracil-DNA glycosylase (UNG) enzyme (Figure 1A). Four families of the UNG enzyme have been identified. Family-1 enzymes are thought to adopt a "base sampling" mechanism of identifying the uracil base.^{2–4} Base sampling requires the uracil nucleotide to assume the extrahelical conformation (Figure 1B) allowing it to enter the UNG binding site and so initiate selective favorable interactions.^{5,6} A hypothesis is that the herpes simplex virus uracil-DNA glycosylase (HSV-UNG) is critical to the reactivation of HSV-1 following a latency period.⁷ A second functional role proposed

for this enzyme is the efficient replication in nerve tissue.⁸ The lack of cellular UNG in neurons combined with the continual deamination of cytosine creates an environment where the need for viral UNG is a necessity for proliferation of HSV-1.⁹ To understand the functional role that HSV-UNG plays in the proliferation of HSV-1 selective inhibition of the virus is necessary. However, the highly conserved secondary structure and the amino acid composition and location of catalytic residues in enzymes of family-1 UNG makes selective inhibition of HSV-UNG over human uracil-DNA glycosylase (hUNG) difficult. There are however behavioral differences between these enzymes that can be exploited. Here the structures of the HSV-UNG and hUNG enzymes are compared. This is done with a view to understanding the selective binding pattern of a series 6-(4-alkylanilino)-uracil HSV-UNG inhibitors compared with their lack of inhibition of hUNG, as well as to rationalize the increased inhibitory effect on HSV-UNG of 6-(4-octylanilino)-uracil compared with the rest of the compounds in the series of different aliphatic chain lengths.

Received: June 26, 2014



ACS Publications

© XXXX American Chemical Society

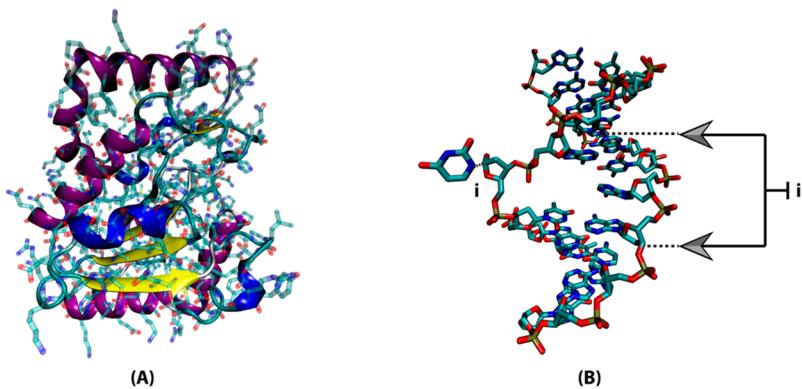


Figure 1. (A) UNG structure (PDB code 1SSP) and (B) segment of DNA double helix taken from Brookhaven Protein Data Bank structure 1SSP. In B, i marks the cleaved glycosidic bond, while ii marks the strand of DNA used in simulations presented in this work.

The 6-(4-alkylanilino)-uracil compounds (Figure 2) are synthetic analogues of the uracil substrate that have been shown to be competitive with DNA and so inhibit hsvUNG.

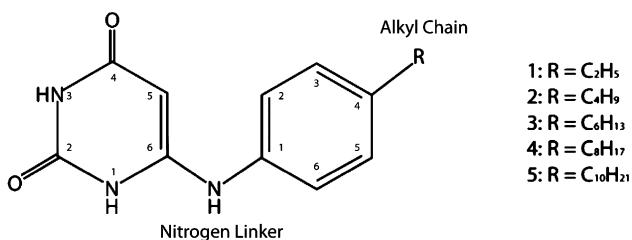


Figure 2. Molecular structure of the series of 6-(4-alkylanilino)-uracil inhibitors showing variable alkyl chain length (R).

COMPUTATIONAL METHODS

All simulations were carried out using the CHARMM program.¹⁰ Incomplete amino acid side chains were built in and the protonation states of all the residues in the crystal structure were corrected to a pH of 7 as based on the estimation of the pK_a of each residue in the protein environment.¹¹ In Figure 1A the unprotonated protein from the structure (PDB code 1SSP) obtained from the Protein Data Bank is shown. The sugar glycosidic bond and uracil base depicted in Figure 1B by i were missing and required computational insertion and repair. The truncated DNA used in the simulation, that include only 5 base pairs, is marked by ii in the same figure. CHARMM parameters for the protein¹² and nucleic acids¹³ were used while the water was modeled using the TIP3P water model^{14,15} as implemented in CHARMM.¹⁶ Ab initio calculations were carried out with the GAUSSIAN 03 program.¹⁷ The sequence analysis work was performed with Geneious v5.1.¹⁸

Parameter Consistency. To ensure that the 6-(4-alkylanilino)-uracil inhibitor series models energetic and conformational performance consistent with that of the natural substrate, we parametrized the charges for NN2U and HN2 and the dihedral parameters Φ and Ψ (given in Figure S1 of the Supporting Information) as these were not in the protein and nucleic acid parameter set. We present the parameters and methods used in the Supporting Information.

Molecular Dynamics. Inhibitors in hUNG and hsvUNG Water Solutions. Simulations were carried out on both hUNG (PDB ID 1SSP) and hsvUNG (PDB ID 1UDG) with the natural substrate as well as the five inhibitors as shown in Figure 2. The

newly parametrized charge and dihedral function coefficients were used for the inhibitors. The systems were solvated in a truncated octahedron cut from a 90 Å water box and water molecules 3 Å away from the heavy atoms were deleted. This resulted in systems with 10536 TIP3P water molecules in each of the hUNG and hsvUNG solution simulation setups. The systems were first heated from 145 to 300 K using periodic boundary conditions in a truncated octahedron of unit cell with an axis length of 77.94 Å. The heating took place over a period of 100 ps followed by an equilibration of 8 ns at a pressure of 1 bar and a temperature of 300 K using the isobaric–isothermal ensemble (NPT) to allow the system's volume to change in order to obtain a water density of 1.0 g·cm⁻³. These equilibrated solution configurations were run for 10 ns in a canonical ensemble (NVT) of which the trajectories were used for analysis.

In the case of each simulation with spherical boundary conditions the protein was covered by a 35 Å TIP3P water sphere with a reaction region with a radius of 31 Å and a 4 Å buffer region. Langevin dynamics was applied in the buffer region to control the temperature of the entire system and molecular dynamics utilizing the Leapfrog integrator was used in the reaction region with no explicit temperature control for this region. A switching function was used to gradually dampen the potential energy to reach zero at the periphery of the nonbonded interaction energy calculation. The switching function was initiated at a cutoff distance of 10 Å and truncated at 12 Å. Distances of hydrogen to heavy atoms were constrained using the SHAKE¹⁹ algorithm. The nonbonded interaction list was updated every 10 fs and a group-by-group selection criterium was used for the inclusion lists. The water sphere was centered on the protein binding pocket's center of geometry. The standard deviation in temperature fluctuation during the production period was ± 1.4 K.

Inhibitors in Water. Simulations were run in explicit TIP3P water for each of the inhibitors. A truncated octahedron with initial crystallographic unit cell axes of length 77.94 Å was used to perform simulations in the isobaric isothermal ensemble. A piston mass of 500 amu and a gamma value of 20 ps⁻¹ at a temperature of 298.15 K were used. Particle mesh Ewald (PME)²⁰ was applied with a kappa value of 0.28 Å, and the reciprocal sum was interpolated with cardinal B-splines of order 6. A grid of 80 increments was used along each of the axes. The pressure was controlled by a Nose Hoover barostat while the temperature was maintained using a Nose Hoover thermostat with a fictitious mass of 1000 amu. Each simulation was 1 ns in length.^{21,22} A force switching function was used for the



Figure 3. Sequence alignment for hsvUNG and hUNG including analysis of hydrophobic regions as illustrative of the large similarity in the protein sequences of the two enzymes.

electrostatics and a potential shifting function for the van der Waals interaction. The switching function was initiated at a cutoff distance of 10 Å and truncated at 12 Å. A pressure of 1 atm was applied.

Principal Component Analysis (PCA). A PCA²³ of the solution trajectories for each of the inhibitors was performed using an internal coordinate distance, instead of the Cartesian coordinates, as the former naturally provide a correct separation of internal and overall dynamics. Specifically the internal coordinate chosen was the distance between the para carbon atom of the benzene aromatic ring (C4 in Figure 2) to the last aliphatic carbon of the chain. Reducing the dimensionality of the data set in this way allowed us to determine the probability and extent of bending of the aliphatic carbon chain. Since the uracil and benzene rings are the same across all the inhibitors their exclusion from the clustering criteria gives an improved signal-to-noise ratio. A cluster radius of 1.0 Å was used for inhibitors 1–2 while a cluster radius of 1.5 Å was used for inhibitor 3. A cluster radius of 2.0 and 3.0 Å was used for inhibitor 4 and 5 respectively. The choice of cluster radii was informed by the need to minimize the number of clusters while preserving a separation of the alkyl chain conformations. The maximum number of clusters was set to 100.

Free Energy of Binding. The starting coordinates for each of the simulations were obtained from the individual molecular dynamics simulations for each inhibitor. In each of the free energy perturbation simulations (Supporting Information Figure S4), a hydrogen atom in the reactant is replaced by a ethyl group in the product. The dual-topology method was used, and the intermediate points between the physical end points ($\lambda_A = 0$ and $\lambda_B = 1$) were defined at coupling parameters (λ) in intervals of 0.05 except near the end points where a spacing of 0.025 was used between windows. For each λ window, 700 ps of equilibration was performed followed by 1.5 ns of data collection. Double-wide sampling was used over the full range of λ in order to double the sampling at the λ values and increase the rate of convergence. A complete description of the free energy protocol is provided as Supporting Information.

RESULTS AND DISCUSSION

Comparative Structural and Sequence Analysis of hsvUNG and hUNG.

The common DNA repair function of

UNG enzymes has resulted in convergent evolutionary similarities in their amino acid sequence and structure leading to a large number of identical amino acid residues in the protein sequence.² Critical amino acids forming close contacts with uracil are identical between hsvUNG and hUNG (Figures 3 and 4); further, a comparison of the overall secondary structure of the proteins reveal that they are strikingly similar. Consequently, very few inhibitors have been able to selectively inhibit hsvUNG to any significant extent.²⁴

Computational analyses using alignment algorithms^{18,25} were used to illustrate the similarities in the amino acid sequence (Figure 3). There is a 40.1% match of the sequence identity in the amino acid sequence between the two proteins. Despite this low sequence identity match between the hsvUNG and hUNG proteins, the secondary structures of the two proteins show large structural similarities with the α helix appearing to be the dominant structural motif that is conserved.

The majority of the identical amino acids are located in the regions including and surrounding the binding pockets (Figure 4). This illustrates the evolution of a conserved binding pocket for the common purpose of recognition of the uracil base and the cleavage of the glycosidic bond between the deoxyribose sugar and the uracil base.

Binding Pocket Analyses. Initial coordinates for both systems were obtained from the Brookhaven Protein Data Bank. The crystal structures used were resolved at 1.9 and 1.75 Å for 1SSP (hUNG)²⁶ and 1UDG (hsvUNG),²⁷ respectively. The protein structure was corrected²⁸ and protonation²⁹ states determined. After evaluating the protonation states of all titratable groups the overall charge of the protein structures were estimated to be +8. We used the Ewald summation method to evaluate electrostatic nonbonded interactions therefore the overall charge of the system had to be adjusted to zero to ensure efficient convergence of the Fourier space Ewald sum.³⁰ We achieved this by including the 5 base pairs that carries a negative charge totalling -8. The base sequence in the double helix being STRAND-1:5'-D(*GP*TP*UP*AP*T)-3' and STRAND-2:5'-D(*AP*TP*AP*AP*C)-3'. The combination of the protein and the DNA that forms a complex results in the overall charge of zero for the system.

The energetically significant amino acids lining the uracil binding pocket are illustrated in Figure 4 for both of the enzymes.

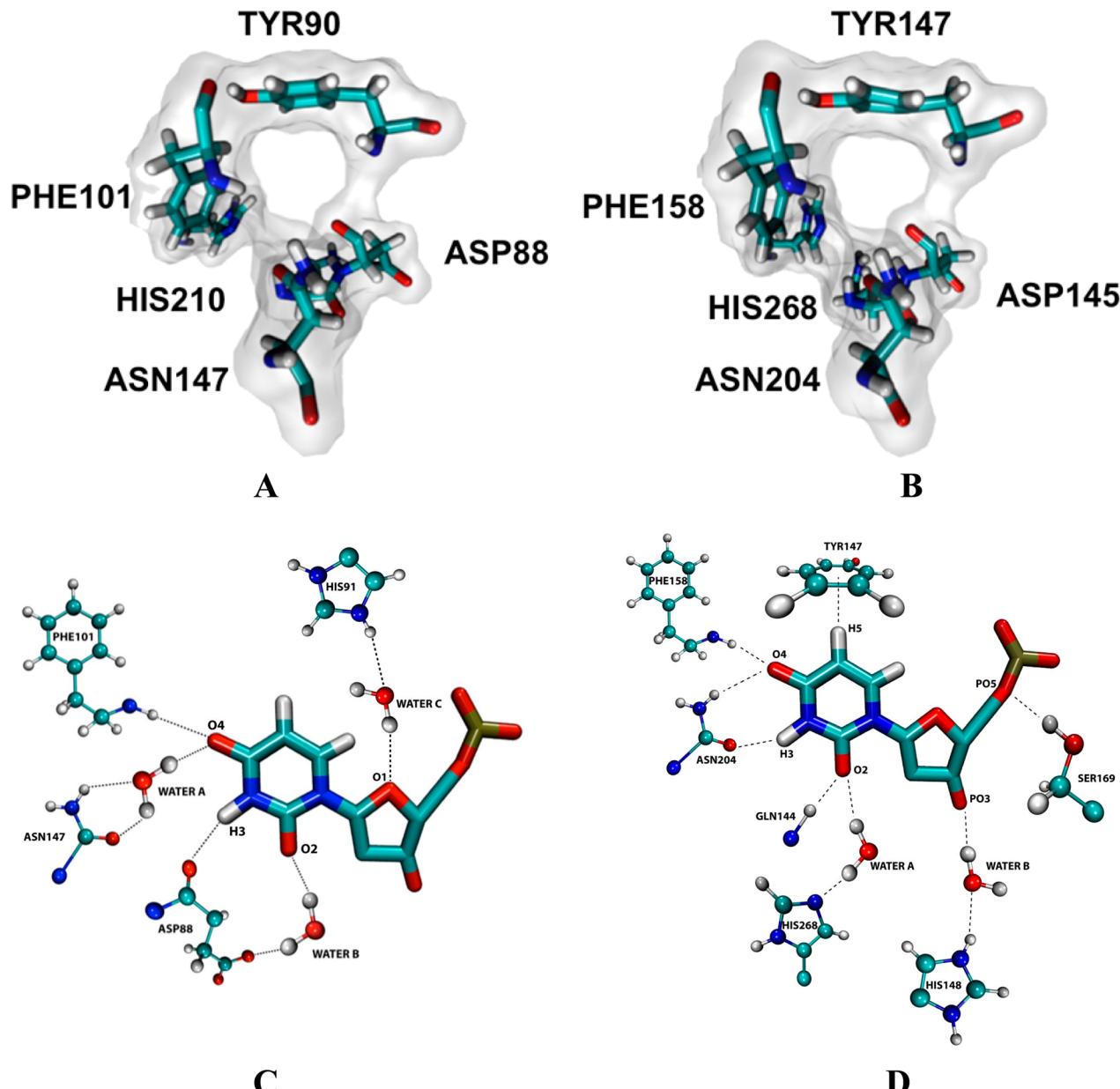


Figure 4. Identical amino acids in the binding pocket of (A) hsvUNG and (B) hUNG. Energetically significant amino acids interactions between (C) hsvUNG and the uracil nucleotide and (D) hUNG and the uracil nucleotide.

Table 1. Inhibitor Binding Regions around the Uracil Binding Pocket Identified for the hsvUNG and hUNG Enzymes

inhibitor binding areas	hsvUNG	hUNG
hydrophobic residues 5 Å away from DNA substrate	PRO89, HIS91, PRO110, PRO111, LEU113, ALA157, ALA158, HIS191, PRO211, PRO213, LEU214, VAL217	HIS148, PRO150, LEU156, PRO168, LEU170, HIS212, ALA214, PRO271, LEU272
gorge region	GLN95, VAL103, VAL107, PRO108, PRO110	GLN152, VAL160, VAL164, PRO165, PRO167
uracil binding pocket	GLY86, ASN87, TYR90, ALA100, PHE101, SER112, GLN147, HIS210	ASP145, GLY143, GLN144, PRO146, TYR147, CYS157, PHE158, ASN204, HIS268

Differences in DNA Binding and Identification of Important Regions for Binding. We performed a STAMP structural alignment in VMD³¹ on the human and the virus proteins with specific attention being paid to the binding regions. The most significant structural difference between the two proteins is that HIS92 in the virus enzyme occurs in the same position as GLY149 in the human enzyme. The conformational effect is observed through the positional similarity of the α carbon of HIS92 in the virus enzyme and the α carbon of

GLY149 in the human enzyme. There are hydrophobic residues 5 Å away from the pentameric DNA model system in the case of both enzymes that are in contact with most of the DNA backbone (see Table 1). We will call this the “DNA binding region” from now on. Hydrophobic analyses of the enzymes revealed a second hydrophobic region just outside of the uracil binding pocket which we will call the gorge region. This gorge region does not associate with the pentameric DNA model system but, as we will show later, is important for binding of the

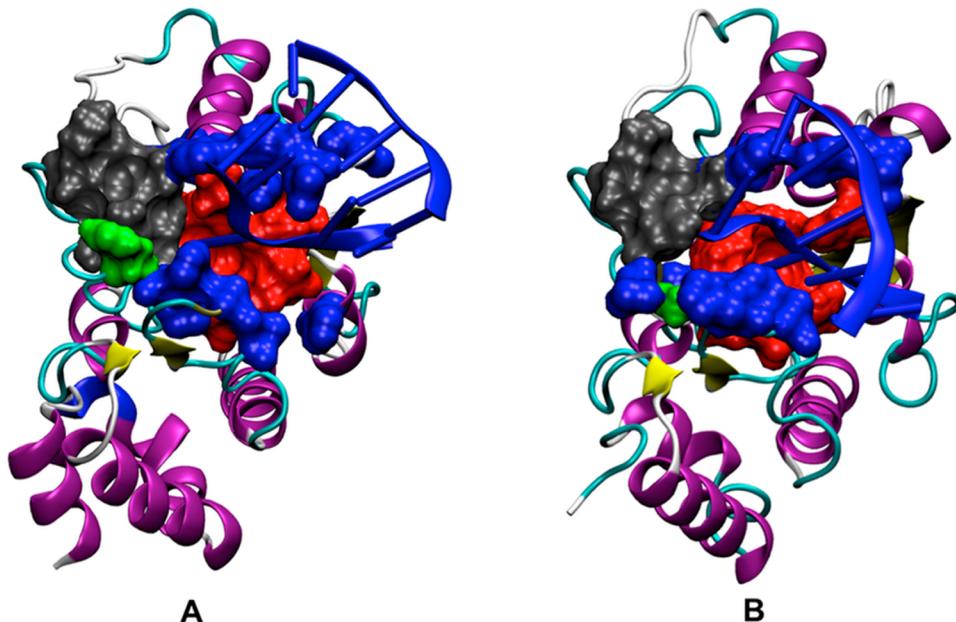


Figure 5. DNA substrate (shown in blue) binding modes of (A) the hsvUNG enzyme and (B) the hUNG enzyme. The uracil binding pockets are shown in red, the “gorge” regions in dark gray, and the hydrophobic residues 5 Å away from the DNA substrate are shown in blue. The HIS92 residue (hsvUNG) and the corresponding GLY149 residue (hUNG) are shown in green.

Table 2. (a) IC₅₀ Values for the 6-(4-Alkylanilino)-uracil Inhibitors for hsvUNG and (b) Relative Binding Free Energies of the 6-(4-Alkylanilino)-uracil Inhibitors to hsvUNG and hUNG^a

(a)		(b)		$\Delta\Delta A$ (kcal·mol ⁻¹)	
Inhibitor	IC ₅₀ (μ M)	perturbation	hsvUNG	hUNG	
1	500	1 → 2	-0.9 (-0.71)	-4.0	
2	150	2 → 3	-1.6 (-0.94)	-1.3	
3	30	3 → 4	-1.8 (-0.78)	-0.9	
4	8	4 → 5	1.9 (0.87)	0.5	
5	35				

^a $\Delta\Delta A$ values as estimated from the experimental IC₅₀ values are shown in parentheses.

Table 3. Average Interaction Energies (kcal·mol⁻¹) between Inhibitors and Key Amino Acid Residues in hsvUNG for Alkyl Chain Lengths for Inhibitors 1–5^a

amino acid	hsvUNG Interaction Energy with Inhibitors				
	1	2	3	4	5
PHE101	-7.4	-7.5	-7.0	-7.7	-7.6
HIS210	-15.4	-16.7	-16.2	-16.1	-16.2
SER112	-5.3	-3.2	-1.4	-2.8	-2.7
GLN87	-3.0	-3.8	-2.2	-2.2	-2.1
ASN147	-9.9	-10.0	-9.8	-9.9	-9.6
GLY86	-1.6	-1.6	-1.6	-1.6	-1.6
ALA100	-2.5	-2.6	-2.5	-2.6	-2.5
TYR90	-6.7	-7.2	-8.4	-9.9	-9.0
total Region I	-51.8	-52.6	-49.0	-52.8	-51.2
HIS92	-0.1	-0.2	-1.1	-2.3	-2.2
PRO110	-0.6	-0.3	-1.3	-2.4	-2.7
VAL103	-0.1	-0.1	-0.2	-0.8	-0.6
HIS91	-0.9	-1.0	-2.6	-3.0	-2.2
PRO108	0.0	0.0	-0.1	-0.4	-0.7
PRO111	-0.4	-0.2	-0.5	-0.5	-0.7
PRO213	-1.1	-1.1	-0.6	-0.4	-0.8
LEU214	-0.4	-0.4	-0.2	0.0	-0.1
ASP88	-5.5	-6.1	-6.6	-6.4	-6.3
PRO89	-3.5	-2.8	-4.2	-4.2	-3.9
GLN95	-0.5	-0.3	-0.7	-1.5	-1.7
VAL107	0.0	0.0	-0.1	-0.8	-1.0
PRO211	-2.1	-1.0	-1.6	-0.9	-1.3
SER212	-0.9	-0.8	-0.3	-1.0	-0.2
total Region II	-16.0	-14.2	-20.1	-24.7	-24.5
total protein	-67.6	-67.9	-69.3	-78.1	-76.5

^aThe values of the highest interaction energy for each of the amino acids are shown in bold.

inhibitors. Although alignment analyses show that the enzymes are 40.1% identical and that the binding pockets are highly conserved, a binding structure analysis shows that their interaction with the DNA substrate is very different (Figure 5). A summary of the differences in binding of DNA between the different regions in the two enzymes are given in Table 1.

An analysis of the enzyme bound DNA substrate simulations showed that the sugar–phosphate backbone plays an important role in the binding of the substrate to the UNG enzymes. The interactions between the DNA substrate and the UNG enzymes are comprised primarily of strong nonspecific electrostatic and van der Waals interactions. The average volume of the hUNG binding pocket was computed to be approximately 20 Å³ smaller than that of the hsvUNG binding pocket. Comparing the distance between the uracil head and the ASN204 and ASN147 amino acids (which are considered the base of the binding pocket) in hsvUNG and hUNG reveals that the uracil in hUNG penetrates the binding pocket deeper than in hsvUNG.

To better understand ligand binding to hsvUNG we compared optimized docked poses for the 6-(4-octylanilino)-uracil (i.e., the most effective inhibitor of hsvUNG determined from experiments) and performed 2 ns simulations for the alkyl chains

pointing towards each of the sides of the DNA binding region areas (shown in blue in Figure 5). The area that produced the

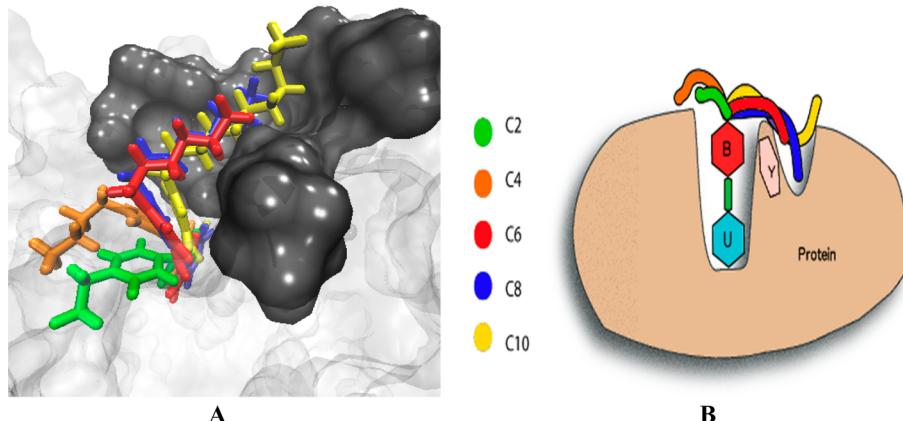


Figure 6. (A) View of inhibitors down into hsvUNG binding pocket. The dark gray surface of the protein represents the amino acids forming the hydrophobic pocket. (B) Schematic of inhibitors in the hsvUNG binding pocket. U, B, and Y represent uracil, benzene, and TYR90, respectively.

Table 4. Summary of Inhibitor-hsvUNG Properties That Determine Binding with RMSD Defined with Respect to the Starting Coordinates of the Simulation

inhibitor	average volume (\AA^3) binding pocket	interaction energy ($\text{kcal}\cdot\text{mol}^{-1}$)	IC_{50} (μM)	distance from binding pocket (\AA)	RMSD (\AA) alkyl chain	binding pocket preference
1	289	-67.6	500	2.83	4.2	DNA-binding
2	294	-67.9	150	2.46	4.3	DNA-binding
3	293	-69.3	30	2.21	2.3	gorge
4	296	-78.1	8	1.85	1.6	gorge
5	298	-76.5	35	2.51	2.1	gorge

strongest interactions with the inhibitors was chosen for all other computations.

Free Energy Perturbation Study of Inhibitor Protein Binding. We compare experimental IC_{50} measurements (Table 2a) with computed free energies of binding ($\Delta\Delta\text{A}$) for the 6-(4-alkylanilino)-uracil inhibitors in hsvUNG and hUNG respectively (Table 2b). In the case of both proteins inhibitor 4 exhibits the strongest inhibition effect. All $\Delta\Delta\text{A}$ are negative, until the simulation transformation going from inhibitor 4 to inhibitor 5 in the hsvUNG enzyme (Table 2b). This indicates a decrease in the inhibition effect which is in agreement with the IC_{50} based experimental results for the hsvUNG enzyme shown in Table 2a. The experiments rank the inhibitors in the order of $4 > 3 > 5 > 2 > 1$ a trend that is well matched with the free energy of inhibitor to hsvUNG binding study (Table 2b). This corroboration of the computed free energy (Table 2b) and the experimental IC_{50} measurements (Table 2a) validate the force field parameters and computational techniques used in this study.

Rationalizing Inhibitor Association with the hsvUNG Enzyme. Previously, a study of this series of inhibitors proposed that the hydrophobic character of the alkyl chains (R) may strongly interact with the hydrophobic area created by the amino acids proline (PRO111 and PRO213) and leucine (LEU214).³²

The average interaction energies between the inhibitors and important amino acids in the binding pocket, gorge, and DNA binding region of hsvUNG were calculated. To simplify the results, we call the uracil binding pocket Region I and the other residues (those in the gorge and DNA binding regions, Region II (shown in Figure S6 A, Supporting Information)). The averages of the calculated interaction energies are shown in Table 3, and the residues in close contact with the inhibitors are shown in Figure 6.

In hsvUNG, the ASN147 amino acid is buried deepest in the binding pocket. Therefore, assessing the strength of association

with the binding pocket can be qualitatively estimated by considering the average distance of the inhibitor from the ASN147 amino acid.²⁷ The high degree of conformational freedom expressed by the alkyl chains necessitates an inclusion of all residues that fall in the gorge area to estimate the binding with the hydrophobic groove.

The bulk of the $-67.6 \text{ kcal}\cdot\text{mol}^{-1}$ interaction energy (averaged over the 2 ns trajectory) between the uracil of inhibitor 1 and hsvUNG is due to the binding pocket interactions and more specifically the 4 hydrogen bonds it forms with the protein. The carbonyl group of ASN147 and the ϵ -nitrogen from the HIS210 ring accepts hydrogen bonds from the hydrogen on N3 and N1 of the uracil head, respectively. PHE101 and GLN87 donate hydrogen bonds to O4 and O2 on the uracil head, respectively. The alkyl tail of inhibitor 1 moves freely in all directions (see Figure 6A) and makes no significant interactions with the hsvUNG enzyme.

Similarly, inhibitor 2 of (averaged) interaction energy $-67.9 \text{ kcal}\cdot\text{mol}^{-1}$ derives the bulk of its association with hsvUNG in the binding pocket with a freely wandering alkyl chain tail that indiscriminately binds to hydrophobic groups about the gorge area. The increase in alkyl chain length from inhibitor 2 to 3 results in the alkyl chain moving with less freedom and being wrapped around TYR90. The overall result is fewer favorable binding pocket interactions with the uracil of inhibitor 3 that is more than made up for by its alkyl chain association with the gorge region leading to a higher average interaction energy of $-69.3 \text{ kcal}\cdot\text{mol}^{-1}$.

The next increment of alkyl chain lengthening occurring in inhibitor 4 leads to a marked increase interaction energy to $-78.1 \text{ kcal}\cdot\text{mol}^{-1}$ with the hsvUNG enzyme. Inhibitor 4 produces the greatest binding affinity of all the inhibitors and is in agreement with experimental values. The alkyl chain fits snugly in the hydrophobic gorge and does not sweep the hydrophobic region

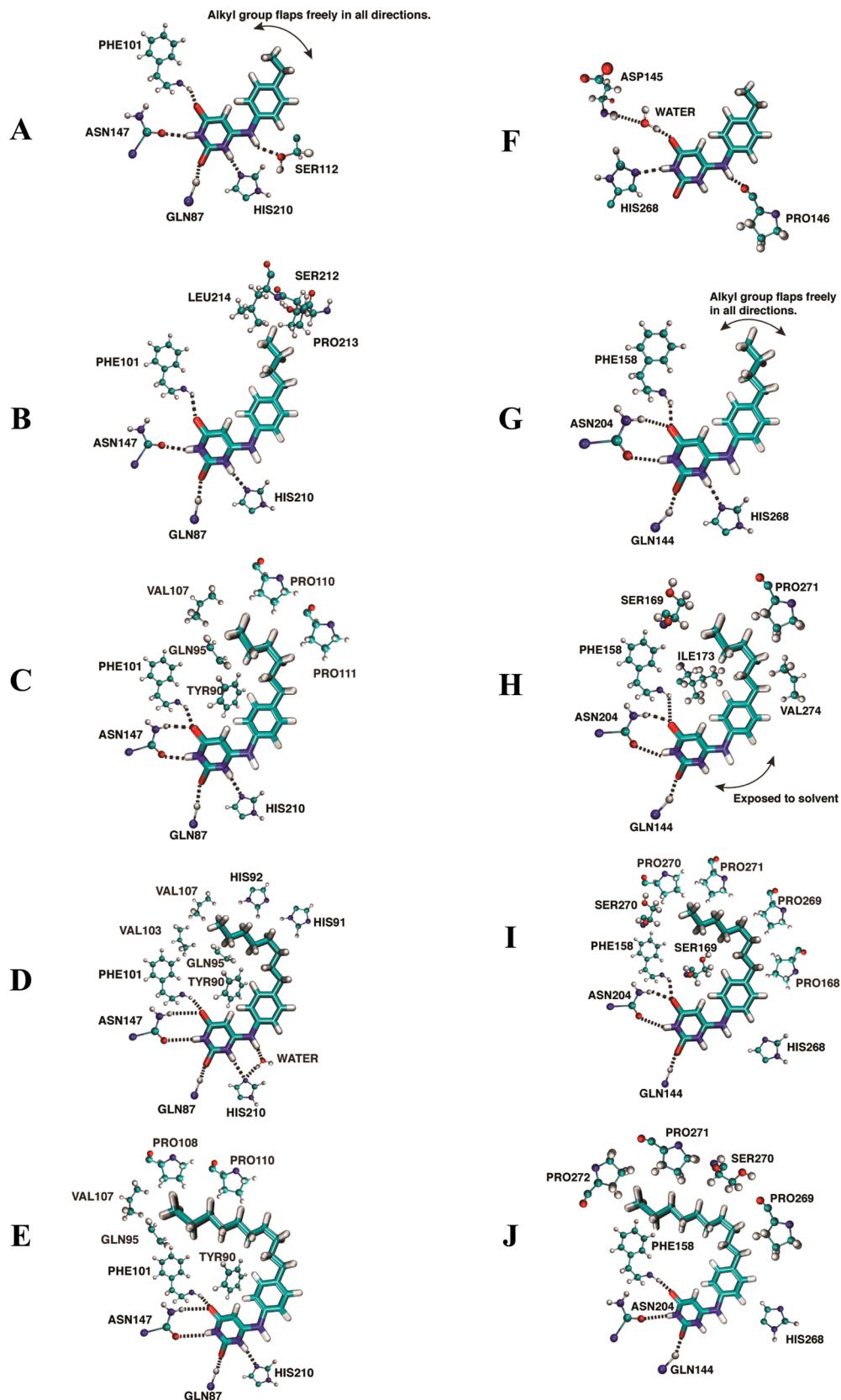


Figure 7. (A–E) Two-dimensional flat view of the hsvUNG enzyme interactions profile with inhibitors 1–5. (F–J) Two-dimensional flat view of the hUNG enzyme interactions profile with inhibitors 1–5.

outside of the binding pocket, as did inhibitors 1 and 2 with the short alkyl chains. Inhibitors 3, 4, and 5 possess the ability to have their alkyl chains form a favorable interaction with the

hydrophobic gorge. The alkyl chains of inhibitors 3, 4, and 5 associate permanently with the hydrophobic gorge of hsvUNG for the duration of the simulations. A common binding pocket

Table 5. Average Interaction Energies between Inhibitors and Key Amino Acid Residues in hUNG

amino acid	hUNG Interaction Energy with Inhibitors				
	1	2	3	4	5
ASN204	0.0	-9.7	-9.1	-8.8	-8.5
CYS157	-0.5	-2.3	-3.3	-2.2	-2.3
ASP145	-4.0	-6.3	-13.4	-7.1	-9.3
PHE158	-0.9	-5.9	-6.2	-7.0	-7.2
TYR147	-6.8	-7.2	-7.0	-7.7	-6.8
GLN144	-1.3	-3.2	-2.3	-2.8	-2.9
GLY143	-0.5	-1.8	-1.6	-1.8	-1.5
HIS268	-2.0	-16.1	-3.9	-1.4	-1.5
PRO146	-4.3	-3.9	-3.5	-3.7	-4.5
total Region I	-20.3	-56.4	-50.2	-42.4	-44.6
PRO269	-0.9	-1.4	-3.3	-1.2	-2.4
PRO168	-0.3	-0.3	-0.6	-0.9	-1.3
PRO167	-0.9	-0.6	-0.3	-0.9	-0.7
HIS148	-5.2	-2.2	-0.2	-1.5	-1.2
SER169	-1.6	-2.9	-4.1	-3.8	-4.3
SER247	0.0	0.0	0.0	-0.2	0.0
ILE173	-0.1	-0.5	-1.4	-0.1	-0.3
SER270	-0.5	-0.5	-1.5	-0.8	-1.5
LEU272	0.0	0.0	-0.1	-0.8	-0.5
SER273	0.0	0.0	-0.2	-0.1	0.0
LEU170	-0.3	-0.6	-1.0	-0.8	-0.5
ASN172	0.0	0.0	-1.4	-0.4	-0.4
PRO271	-0.3	-0.2	-2.1	-1.0	-1.1
GLN152	-1.1	0.0	0.0	-0.5	-0.3
total Region II	-11.2	-9.1	-16.3	-12.9	-14.5
total protein	-36.8	-65.7	-69.1	-66.8	-60.0

affinity is the formation of a water-mediated interaction between the uracil and TYR90. This is indeed the case, except in the case of inhibitor 1, which does not penetrate the binding pocket deep enough.

For the hsvUNG protein, it can be seen that the hydrophobic gorge area shown in Figure 5 plays an important role in explaining the trend in the inhibition effect of the inhibitors for the virus. As the alkyl chain length is increased from two to eight carbons, we see an increase in the interaction energy of the inhibitors. When the transformation in the alkyl chain is increased from 8 carbon atoms to 10 carbon atoms, a decrease in the inhibition effect is observed. This is illustrated in Figure 6 where the increase in alkyl chain length of inhibitor 5 (yellow) is slightly too long to lock into the gorge region. A comparison of properties that may influence the binding affinity of the inhibitors to hsvUNG (Table 4) reveals that the length of the alkyl chain is a major determinant for binding.

Inhibitor Protein Interaction Profile Analyses. Hydrogen bonds, van der Waals, electrostatic, water-mediated, and π -stacking interactions are commonly occurring interactions between proteins and their substrates. A profiling of these

interactions for each of the 10 ns production trajectories was done and is shown in Figure 7 in order to display the variation in position of the inhibitors as their alkyl chain increases in length. The interaction profiles shown are flat 2-dimensional orientated projections of the inhibitors along with the strongly interacting amino acids. These orientations were derived from simulations where the starting points for each inhibitor were based on the same position as that observed for the strongest inhibitor (inhibitor with the highest IC_{50} value). The starting positions for the uracil moiety of each inhibitor are relatively deep inside the binding pocket.

The carboxyl group of the side chain of ASN147 and the ϵ -nitrogen atom from the HIS210 ring accepts hydrogen bonds from the hydrogen on N3 and N1 of the uracil head, respectively. PHE101 and GLN87 donate hydrogen bonds to O4 and O2 on the uracil head, respectively. As already illustrated in Figure 6, we again see (Figure 7A) that the alkyl tail of inhibitor 1 moves freely in all directions leading to various weak interactions with the hsvUNG enzyme.

Rationalizing Inhibitor Behavior in the hUNG Enzyme.

The average interaction energies between the 6-(4-alkylanilino)-uracil inhibitor and important amino acids were calculated and tabulated. The interactions were grouped into interactions with the binding pocket and interactions with the adjacent hydrophobic gorge residues (Table 4). In hUNG, ASN204 is the most buried amino acid in the binding pocket and its interaction with the uracil head of the natural substrate in DNA is a measure of the extent of enzyme DNA binding. Due to the highly conserved nature of the family of uracil-DNA glycosylase enzymes as with the ASN147 counterpart in the hsvUNG enzyme the farther the inhibitor is on average from ASN204, the less likely the inhibitor is to displace the natural substrate in the binding pocket.²⁶ This is considered to be a rough estimate of the binding capability of the inhibitor.

Inhibitor 1 has a weak binding affinity for hUNG (average interaction energy of -36.8 kcal·mol⁻¹) and does not make significant contacts with the catalytic residues (see Table 5). The major contribution to the binding emanates from two hydrogen bonds to HIS268 and PRO146 and a water-mediated interaction between the O4 atom of uracil and ASP145 (Figure 7F).

At the other end a potential binding with the hUNG gorge area comes about sporadically as the alkyl tail does not anchor to the enzyme. It is the weakest inhibitor of hUNG of the five inhibitors.

In the case of inhibitor 2 the binding pocket is penetrated more deeply leading to the formation of a bidentate interaction with the amide functional group of ASN204 (Figure 7G). The alkyl groups in inhibitor 1 and 2 have significant conformational freedom. A water-mediated interaction forms between TYR90 and the uracil head of inhibitor 2 (Figure 7G). Inhibitor 3 has an average interaction energy of -69.1 kcal·mol⁻¹ with the hUNG enzyme and forms a bidentate interaction with amide functional group of ASN204 (Figure 7H). The most preferred hydrophobic pocket for the alkyl chain of inhibitors 3 (Figure 7H), 4 (I), and 5

Table 6. Summary of Properties Determined for the Binding of the Inhibitors for the hUNG Enzyme

inhibitor	average volume (\AA^3) binding pocket	interaction energy (kcal·mol ⁻¹)	IC_{50} (μM)	distance from binding pocket (\AA)	RMSD (\AA) alkyl chain	binding pocket preference
1	265	-36.8	>500	10.42	4.4	DNA-binding
2	269	-65.7	>500	2.4	4.3	DNA-binding
3	274	-69.1	>300	2.03	3.6	DNA-binding
4	272	-66.8	>300	2.05	3.3	DNA-binding
5	275	-60.0	>500	3.8	3.8	DNA-binding

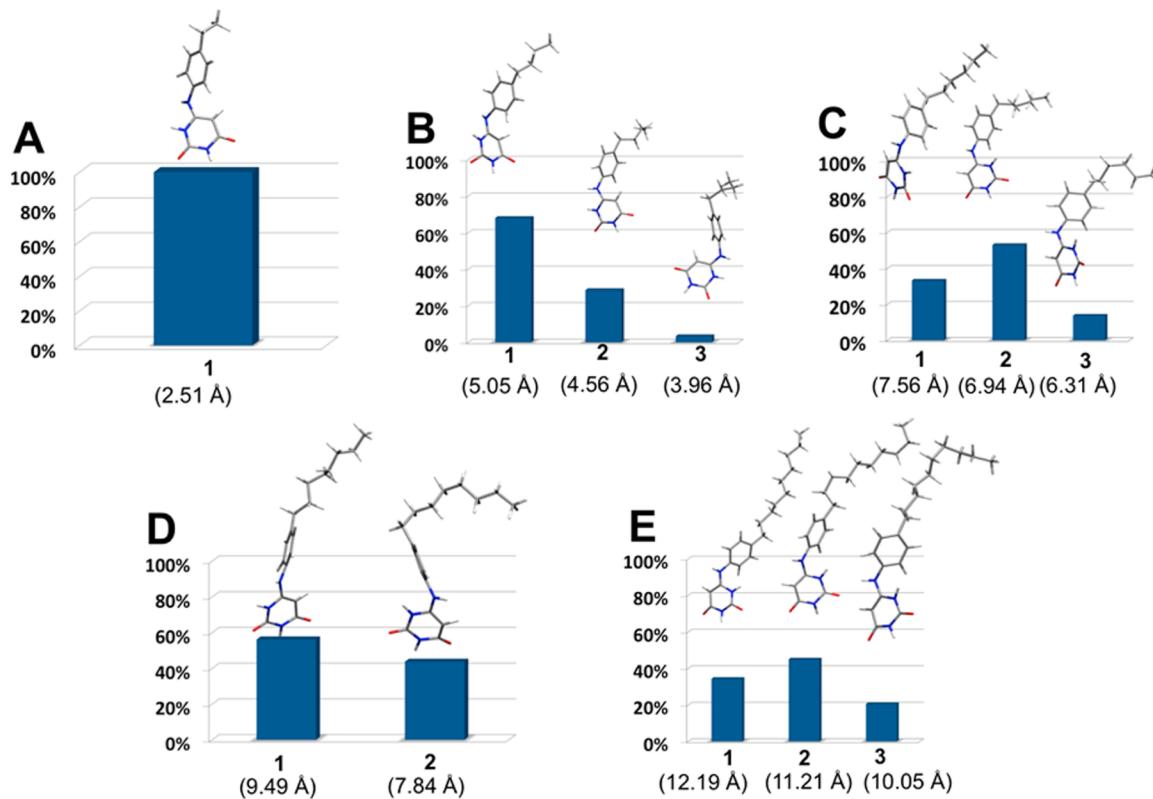


Figure 8. (A–E) Results for a cluster analysis of inhibitors 1–5 in water. The principal component (the distance between the para-carbon of the benzene ring and the last carbon of the aliphatic chain) for each of the cluster centers is shown in parenthesis underneath each figure.

(J) is the pocket lined with the residues SER270, PRO269, and PRO271. This is equivalent to the SER212, PRO213, and PRO214 amino acids that form part of the short DNA binding pocket in hsvUNG which was predicted previously to associate with the alkyl tails of the inhibitors.³² None of the inhibitors associate with the gorge region.

In Table 6 we summarize our results for the inhibitor binding to hUNG, including values for additional properties that were determined in this study.

Ligand Pose Analysis. With the uracil group buried inside the UNG binding pocket the alkyl chains of the alkyl group take on a bent conformation to reach the gorge area. In the virus this bending appears to be assisted by HIS92, which has a relatively similar position to that of GLY149 in the human enzyme. Two models for describing ligand protein binding are the Koshland model and the selected shift model. Koshland's induced-fit model³³ for molecular recognition events dictates that both the ligand and catalytic binding site conformations undergo changes upon binding.

The more recent selected-shift model, or conformational selection mechanism, describes the dynamic response of the protein resulting in the initial binding of a preconfigured ligand conformation to an ensemble of ligand binding sites.³⁴ Recently we investigated the response of an MshB enzyme to the binding of a series of inhibitors using these models as a template and found that the preconfigured solutions conformation of the ligand is central to recognition and binding.³⁵ With this in mind, we investigated the conformations in water for all the inhibitors using the distance from the C4 para carbon of the benzene ring to the terminal carbon on the alkyl group as described in the computational details above (Figure 8). Inhibitor 4 has a larger preference for alkyl chain bending than the rest of the inhibitors.

The solution preconfigured bent ensemble of inhibitor 4 conformations is more likely to access the hsvUNG gorge area compared with the other inhibitors. In the case of hUNG the bent conformers of inhibitor 4 align more readily along the DNA binding site than the gorge area of that enzyme (Table 6).

CONCLUSIONS

The 6-(4-alkylanilino)-uracil inhibitor that has an octyl hydrophobic group (inhibitor 4) is able to selectively inhibit the hsvUNG enzyme despite its structural similarities and sequential identities that it shares with the human enzyme hUNG. Molecular dynamics simulations of the 6-(4-alkylanilino)-uracil inhibitors show that they exhibit an increased propensity for a bent alkyl chain in water with increase in chain length. The bending has its origins in the aliphatic chain self-associating in a manner consistent with the hydrophobic effect. The conformational complementarity between the bent inhibitor 4 and hsvUNG hydrophobic gorge region is superior to that of the human enzyme. The gorge regions in hsvUNG and hUNG are similar in structure and sequence making it theoretically possible that the inhibitor 4 (octyl chain length) should strongly associate with the gorge region in the human enzyme. This is not the case, hence the selective inhibition.

An analysis of the region around the uracil binding pocket using structural alignment algorithms revealed that the α carbon of HIS92 in the virus enzyme is very close to the α carbon of GLY149 in the human enzyme. This lack of a histidine in that position in the case of the human enzyme appears to lead to a less favorable interaction compared with the virus enzyme. An interaction energy analysis further supports this observation, as the interaction with HIS92 is significant compared with other amino acids in Region II. The identification of the hydrophobic

regions particularly the gorge and specific properties of inhibitor 4, i.e., chain length and preconfigured conformational bending, has not been previously made. This provides a rational explanation of the experimental finding that 6-(4-alkylanilino)-uracil with an octyl chain is an efficient and selective inhibitor of hsvUNG.

■ ASSOCIATED CONTENT

Supporting Information

Computational details. Figures S1–S4 and Tables S1–S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kevin.naidoo@uct.ac.za

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is based upon research supported by the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation to K.J.N. U.H. and W.C. thank SARChI for masters and postdoctoral fellowship support, respectively.

■ REFERENCES

- (1) Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362*, 709–715.
- (2) Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutation Research-DNA Repair* **2000**, *460*, 165–181.
- (3) Bellamy, S. R. W.; Krusong, K.; Baldwin, G. S. A rapid reaction analysis of uracil DNA glycosylase indicates an active mechanism of base flipping. *Nucleic Acids Res.* **2007**, *35*, 1478–1487.
- (4) Cao, C.; Jiang, Y. L.; Stivers, J. T. The catalytic power of uracil DNA glycosylase in the opening of thymine base pair. *J. Am. Chem. Soc.* **2006**, *128*, 13034–13035.
- (5) Cao, C.; Jiang, Y. L.; Stivers, J. T.; Song, F. Dynamic opening of DNA during the enzymatic search for a damaged base. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1230–1236.
- (6) Wong, I.; Lundquist, A. J.; Bernards, A. S.; Mosbaugh, D. W. Pseudo steady-state analyses of a single catalytic turnover by Escherichia coli uracil-DNA glycosylase reveals a "pinch-pull-push" mechanism. *J. Biol. Chem.* **2002**, *277*, 19424–19432.
- (7) Verri, A.; Mazzarello, P.; Biamonti, G.; Spadari, S.; Focher, F. The Specific Binding of Nuclear Protein(s) to the cAMP Responsive Element (CRE) Sequence (TGACGTCA) is Reduced by the Misincorporation of U and Increased by the Deamination of C. *Nucleic Acids Res.* **1990**, *18*, 5775–5780.
- (8) Pyles, R. B.; Thompson, R. L. Evidence that the Herpes Simplex Virus Type 1 Uracil DNA Glycosylase is Required for Efficient Viral Replication and Latency in the Murine Nervous System. *Journal of Virology* **1994**, *68*, 4963–4972.
- (9) Chen, R.; Wang, H.; Mansky, L. M. Roles of Uracil-DNA glycosylase and dUTPase in virus replication. *Journal of General Virology* **2002**, *83*, 2339–2345.
- (10) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (11) Nielsen, J. E. Analysing the pH-dependent properties of proteins using pK(a) calculations. *J. Mol. Graph. & Model.* **2007**, *25*, 691–699.
- (12) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (13) Foloppe, N.; MacKerel, A. D. All-atom empirical force field for nucleic acids: 1. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (14) Jorgensen, W. L. Revised TIPs for simulations of liquid water and aqueous solutions. *J. Phys. Chem.* **1982**, *77*, 4156–6516.
- (15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (16) Steinbach, P. J.; Brooks, B. R. Protein hydration elucidation by molecular dynamics simulations. *Proc. Natl. Acad. Sci. (USA)* **1993**, *90*, 9135–9139.
- (17) Frisch, M. J.; T, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Ciosowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, revision E.01; Gaussian, Inc.: Wallingford, CT and Pittsburgh, PA, 2004.
- (18) Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649.
- (19) van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.
- (20) York, D.; Darden, T.; Pedersen, L. G. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (21) Nose, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (22) Hoover, W. G. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (23) Jolliffe, I. T. *Principal Component Analysis*, second ed.; Springer-Verlag: New York, 2002.
- (24) Krusong, K.; Carpenter, E. P.; W, B. S. R.; Savva, R.; Baldwin, G. S. A comparative Study of Uracil-DNA Glycosylase from Human and Herpes Simplex Virus Type 1. *J. Biol. Chem.* **2006**, *281*, 4983–4992.
- (25) Drummond, A. J.; Ashton, B.; Buxton, S.; Cheung, M.; Cooper, A.; Heled, J.; Kearse, M.; Moir, R.; Stones-Havas, S.; Sturrock, S.; Thierer, T.; Wilson, A. *Geneious v5.1*; Biomatters Limited, 2010.
- (26) 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.
- (27) Savva, R.; McAuleyhecht, K.; Brown, T.; Pearl, L. The Structural Basis of the Specific Base-Excision Repair by Uracil-DNA Glycosylase. *Nature* **1995**, *373*, 487–493.

- (28) Kuszewski, J.; Gronenborn, A. M.; Clore, M. G. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.* **2008**, *5*, 1067–1080.
- (29) Nielsen, J. E.; Vriend, G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Proteins-Structure Function and Genetics* **2001**, *43*, 403–412.
- (30) Toukmaji, A. Y.; Board, J. A. Ewald summation techniques in perspective: a survey. *Comput. Phys. Commun.* **1996**, *95*, 73–92.
- (31) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (32) Sun, H. M.; Zhi, C. X.; Wright, G. E.; Ubiali, D.; Pregnolato, M.; Verri, A.; Focher, F.; Spadari, S. Molecular modeling and synthesis of inhibitors of herpes simplex virus type 1 uracil-DNA glycosylase. *J. Med. Chem.* **1999**, *42*, 2344–2350.
- (33) Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. (USA)* **1958**, *44*, 98–104.
- (34) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603.
- (35) Rogers, I. L.; Gammon, D. W.; Naidoo, K. J. Conformational preferences of plumbagin with phenyl-1-thioglucoside conjugates in solution and bound to MshB determined by aromatic association. *Carbohydr. Res.* **2013**, *371*, 52–60.