



## OPEN The 'very moment' when UDG recognizes a flipped-out uracil base in dsDNA

Vinnarasi Saravanan<sup>1,4</sup>, Nessim Raouraoua<sup>1,4</sup>, Guillaume Brysbaert<sup>1</sup>, Stefano Giordano<sup>2</sup>, Marc F. Lensink<sup>1</sup>, Fabrizio Cleri<sup>3</sup> & Ralf Blossey<sup>1</sup>✉

Uracil-DNA glycosylase (UDG) is the first enzyme in the base-excision repair (BER) pathway, acting on uracil bases in DNA. How UDG finds its targets has not been conclusively resolved yet. Based on available structural and other experimental evidence, two possible pathways are under discussion. In one, the action of UDG on the DNA bases is believed to follow a 'pinch-push-pull' model, in which UDG generates the base-flip in an active manner. A second scenario is based on the exploitation of bases flipping out thermally from the DNA. Recent molecular dynamics (MD) studies of DNA in trinucleosome arrays have shown that base-flipping can be readily induced by the action of mechanical forces on DNA alone. This alternative mechanism could possibly enhance the probability for the second scenario of UDG-uracil interaction via the formation of a recognition complex of UDG with flipped-out base. In this work, we describe DNA structures with flipped-out uracil bases generated by MD simulations which we then subject to docking simulations with the UDG enzyme. Our results for the UDG-uracil recognition complex support the view that base-flipping induced by DNA mechanics can be a relevant mechanism of uracil base recognition by the UDG glycosylase in chromatin.

The recognition and the subsequent cleavage of wrongly incorporated or chemically induced uracil bases in DNA by repair enzymes is the initial step in the base-excision repair (BER) process. It is performed by the enzyme Uracil-DNA glycosylase (UDG), a protein of 313 amino acids (UniProt P13051 UNG\_HUMAN) that was the first glycosylase to be structurally characterized in complex with damaged DNA<sup>1</sup>; for a review of structural and functional properties of the UDG superfamily of glycosylases, see<sup>2</sup>. UDG structures were repeatedly resolved experimentally in interaction with DNA and made available through the Protein Data Bank (PDB, <https://www.rcsb.org>).

UDG initiates the BER pathway for uracil excision by hydrolyzing a glycosylic bond between a uracil base and the deoxyribose sugar in order to produce a free uracil and an abasic site in the DNA strand<sup>2</sup>. Human UDG also removes uracil from single-stranded DNA, but it is not active against uracil in RNA<sup>3</sup>. Uracil excision requires the base to be flipped out of the DNA double strand so that it can be processed in the enzyme active site<sup>1</sup>.

DNA glycosylases must be extremely efficient to inspect all bases in the genome ( $3 \times 10^9$  base pairs) and detect and repair damaged bases (around 1 in  $10^6$  bases per cell daily) before the genetic code is permanently altered in the next cycle of DNA replication. Given that there are on the order of  $10^5$  copies of glycosylase molecules in the nucleus, each such protein excluding redundancy should be able to sample around 70,000 base pairs of DNA in every cell cycle (about 12 to 24 h)<sup>4</sup>. The fact that, in practice, glycosylases can achieve their job much faster points to the existence of better than purely stochastic mechanisms.

How UDG actually finds the wrongly incorporated base is a question that cannot be conclusively answered by crystallography, as it obviously is a highly dynamic process and X-ray only allows for access to final stable states of the interacting species. In order to cleave the base, UDG must first gain access to it. Several structural studies found that different DNA repair proteins can actively flip out their target base extra-helically (the so-called 'enzymatic' base flipping, see e.g.<sup>5-7</sup>); the efficiency of the excision process is sequence-dependent as demonstrated in<sup>5</sup>; see also the recent work by Orndorff et al.<sup>8</sup>. The mechanisms involved in amino-acid assisted base flipping have received ample attention in the literature also for DNA repair proteins other than UDG, see, e.g.<sup>9-11</sup>. Recent single-molecule experiments have provided evidence that UDG works in a 'scanning' mode, i.e.

<sup>1</sup>University of Lille, CNRS, UMR8576, Unité de Glycobiologie Structurale et Fonctionnelle (UGSF), F-59000 Lille, France. <sup>2</sup>CNRS, Centrale Lille, Univ. Polytechnique Hauts-de-France, UMR 8520 - IEMN - Institut d'Electronique, de Microélectronique et de Nanotechnologie, University of Lille, 59000 Lille, France. <sup>3</sup>Institut d'Electronique, de Microélectronique et de Nanotechnologie (IEMN CNRS UMR8520) and Département de Physique, University of Lille, 59652 Villeneuve d'Ascq, France. <sup>4</sup>These authors contributed equally: Vinnarasi Saravanan and Nessim Raouraoua. ✉email: ralf.blossey@univ-lille.fr

processing along DNA, combined with a ‘peeking’-mode, which involves the insertion of an amino acid into the DNA stack<sup>12</sup>. Already before experimental evidence had been provided for hopping and sliding motions of UDG on DNA<sup>13</sup>, a detailed ‘pinch-push-pull’ scenario had been formulated, in which the non-specific scanning step of UDG along DNA is followed by highly specific actions of UDG on DNA<sup>3</sup>. In this scenario, UDG first induces a kink in the DNA backbone, allowing for the formation of specific amino acid contacts with the base, which then leads to a backbone compression that ‘pinches’ the DNA. Subsequently, the intercalation loop of UDG penetrates the minor groove and flips out the uracil base. In the final ‘pull’ step, the glycosidic bond is cleaved. Importantly, apart from the non-specific diffusion or hopping along DNA, the amino-acid-DNA interaction is actively pursued by UDG. This ‘pinch-push-pull’ scenario thus starts from an initially pristine DNA, and requires UDG to very actively search for the ‘hidden’ base<sup>3</sup>. There is good reason to doubt that search processes are the only, or even the dominant mechanism in locating wrong bases for excision in DNA. Based on available data from in-vitro experiments, a simple Monte Carlo (MC) random search simulation could well accommodate the experimental situation; however, it remains unrealistic by several orders of magnitude for realistically estimated chromatin densities in the eukaryotic nucleus (see [Supplementary Material](#)). Such considerations therefore motivate the investigation for alternative mechanisms, by which defect-containing DNA fragments may become highly visible, and more easily identifiable by the repair proteins compared to serendipitous random search.

It is still under debate whether the spontaneous base flipping, which normally occurs at random all along the DNA because of thermodynamical fluctuations<sup>14–17</sup>, could be an even stronger attractor for glycosylases and other similarly functioning enzymes<sup>18–20</sup>. Given the much longer lifetimes measured by NMR for extrahelical flipping (in the  $\mu\text{s}$  to ms range) compared to intrahelical flip, and the fact that even partially flipped bases may become accessible<sup>21–23</sup>, spontaneous flipping must therefore be properly investigated as a possible first-step, or possibly the very rate-limiting event, in the glycosylase search for damaged DNA sites. Base flipping in pristine DNA in fact happens stochastically and an encounter with UDG could therefore also happen randomly; it is however easy to estimate that the probability of a purely random encounter (product of two independent random events), even with a substantial concentration of UDG molecules, will be generally too low to be of practical importance. DNA in the cell nucleus, however, is subject to numerous untargeted interactions, which might indeed lead to a large increase in the rate of base flipping, compared to the thermally-assisted one, thereby largely facilitating uracil recognition by UDG. In particular, mechanical constraints acting on the DNA structure have been indicated as an additional source of localized bending, twisting and kinking of the double helix, all which are favorable conditions for increasing the base-flipping rate. Uracil differs from thymine by missing the extra methyl at C5 carbon facing the major groove, which creates a distortion in the standard B-DNA structure, and leaves room for further defect evolution, such as clustering of intrastrand cross-links<sup>24</sup>. It has been well established that e.g. a strong curvature of DNA drastically increases its flexibility<sup>25</sup>. Interestingly, in a recent work<sup>26</sup> we modelled by fully-atomistic molecular dynamics (MD) simulations an array of three nucleosomes, under applied mechanical forces in a range reminding the compressive regime that cells may be typically subject during their lifetime. These simulations have shown that the resulting compression of the structure can induce kink instabilities in the linker DNA which facilitate the opening of the double-stranded linker DNA and the flipping-out of the bases<sup>26</sup>. This mechanical mechanism leads to the formation of localized denaturation bubbles through a mechanical twist of the DNA double helix<sup>27</sup>. It can therefore be concluded that the ubiquitous mechanical action on DNA could therefore present a partially or fully exposed uracil base to UDG. This observation thus raises the question whether DNA mechanics can play, among other effects, a relevant role in the recognition process of uracil bases by repair enzymes. Starting from the insights gained in<sup>26</sup>, in this paper we characterize the interaction of UDG with exposed uracil bases and attempt to find the instant in which the recognition complex is formed between the flipped-out base and UDG. In other words, we aim at finding the ‘very moment’ (a notion inspired by<sup>28</sup>) when UDG can recognize a flipped-out base without having previously intervened in its formation. Here we study this problem by a combination of the controlled induction of uracil base-flipping via MD and protein-DNA docking simulations. Our finding of this ‘very moment’ of encounter entails the question of the quality of the complex formed: we compare our simulation results to the crystal structure of the UDG-DNA complex, which is assumed as representative for the action of the enzyme on uracil.

The flipping of nucleic bases has been studied by MD simulations by several authors, see, e.g.<sup>29–32</sup>. Studying the spontaneous base flipping is challenging, as the free energy penalty for the extrahelical state of individual DNA bases is about 10 kcal/mol<sup>33,34</sup>. According to previous nuclear magnetic resonance (NMR) studies<sup>14,35</sup>, the lifetime of the extrahelical state of a flipped DNA base is on the order of microseconds ( $\mu\text{s}$ ). In contrast, the intrahelical state lasts from milliseconds to hundreds of milliseconds, depending entirely on the stability of distinct base pairs in the dsDNA. This significant difference between the timescales accessible in all-atom molecular dynamics (MD) simulations and those of base flipping makes it difficult to obtain converged statistics in computer simulations. Therefore, the rare-event computational methodology is required for spontaneous base flipping mechanism. Várnai and Lavery<sup>29</sup>, Huang et al.<sup>36</sup> and Law and Feig<sup>37</sup> have used external forces to induce individual base flipping through umbrella sampling and replica exchange simulation methodologies. In this work, we have used collective variable (CV) based metadynamics<sup>38</sup> simulations to study the base-flipping mechanism.

Having the base-flipped DNA structures at hand, either from the dedicated simulations described here, or from our earlier work on the compression of trinucleosomes<sup>26</sup>, we use protein-DNA docking methods to study the formation of the recognition complexes as a function of the flipping of the base. Contrary to protein-protein docking, where standard methods working in most of the cases are established, protein-DNA docking requires special attention to the algorithm used, in particular when DNA bears uracil. It is even more essential to assess the efficiency of these different methods for our system as it corresponds to a non-standard case: the nature of the DNA strands that are docked to UDG here is highly altered, as the treatment done to induce base flipping is causing deformations on the strand, and the flipped-out base represents a significant conformational change of

the pristine dsDNA. In order to determine the best-suited protein-DNA docking method for our case, we had to test several available algorithms, and describe in detail our quantification of the quality of the encounter complex through protein-DNA docking.

## Results

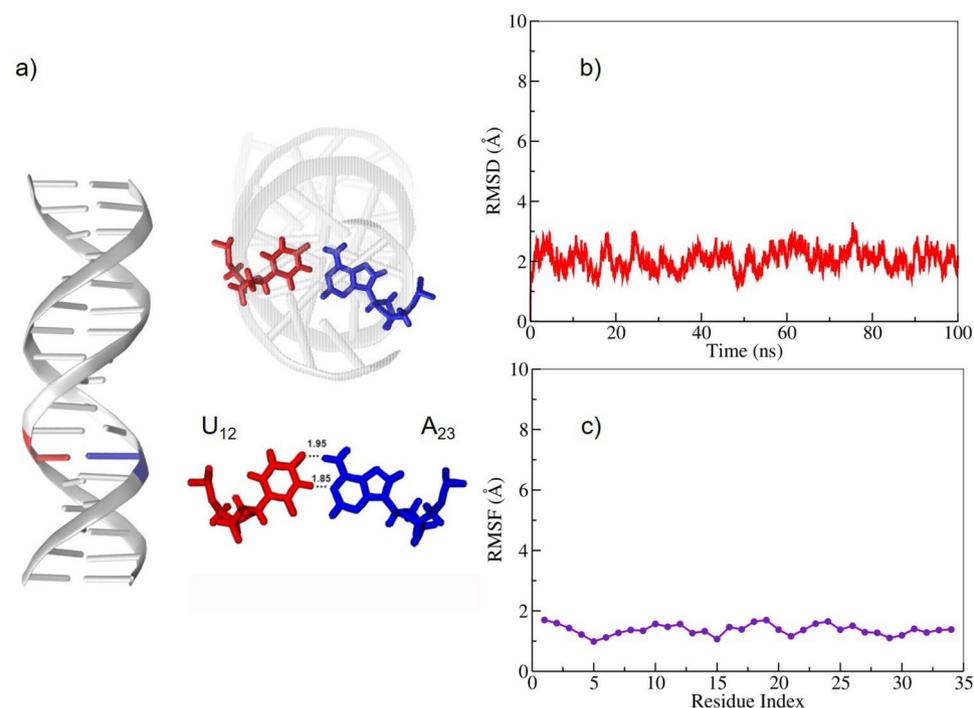
### Mechanisms of uracil base flipping and associated structural deformations

We have obtained dsDNA conformations with flipped uracil bases in two ways. The first approach makes use of the structures generated during the compression MD simulations of tricnucleosome arrays performed in our earlier work<sup>26</sup>. How these structures were selected and prepared for our docking simulations is discussed in “Methods”.

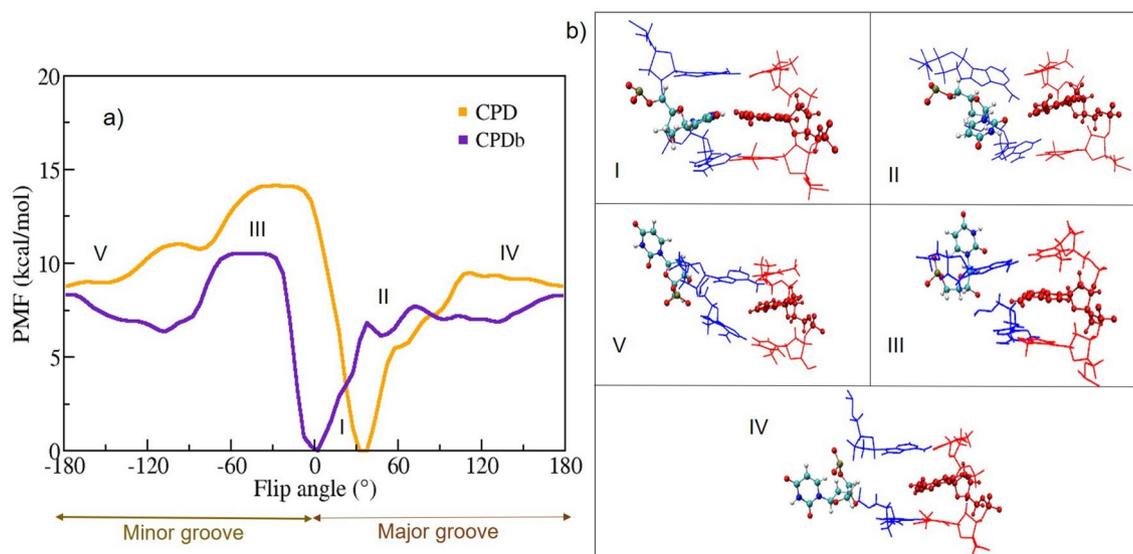
In a second approach, we have generated flipped-out bases in a controlled way from mutated dsDNA by using the metadynamics technique. For this we generated a dsDNA 17-mer in the standard B-DNA form given by the sequence 5′ – CAGGATGTATATATCTG – 3′. The thymine nucleotide T at position 12 was mutated to a uracil base with its cartesian coordinates generated by the Web 3DNA server<sup>39</sup>. The targeted central base pair for flipping was U<sub>12</sub>:A<sub>23</sub>, with U<sub>12</sub> as the target nucleotide for base flipping; the geometry is illustrated in Fig. 1a). First, an all-atom conventional MD simulation was performed to stabilize the dsDNA system over approximately 100 ns. The root-mean-square deviation (RMSD) was used to monitor the overall stability of the dsDNA, as shown in Fig. 1b). The overall RMSD fluctuation was around 1–3 Å, which shows that the relative structural changes are very small. Moreover, the root mean-square fluctuation (RMSF) values calculated for the uracil-mutated dsDNA reveal very small fluctuations at the individual base level, see Fig. 1c).

During the base flipping simulation, an external history-dependent bias potential is applied to the system which can be expressed as a sum of Gaussians along the collective variables (CV) to enhance the sampling efficiency. The CVs for the uracil base flip have been chosen following existing literature, *i.e.* different versions of the center-of-mass pseudo-dihedral angle (CPD), see<sup>40</sup> and CPDa/b<sup>31</sup>. For further details, see “Methods”.

The metadynamics simulations were conducted for the uracil mutated dsDNA system for about 100 ns. The pseudo-dihedral angles (CPD and CPDb) were used as the coordinates for free energy mapping to describe the uracil flipped-in and flipped-out states using the potential of mean force (PMF) analysis as shown in Fig. 2a). The base opening pathway snapshots are shown in Fig. 2b). In this Figure, notable similarities and differences between the two energy profile schemes are observed. The global minima of uracil embedded states at approximately 10° in CPDb correlate well with previous theoretical studies<sup>41</sup>, whereas in the CPD scheme, it is approximately 40°. Here, the preselected CPD reaction coordinates scheme yields slightly different results. Both PMF profile schemes display a separation into two distinct regions: the major groove and the minor groove. A nucleobase within the dsDNA aligns with CPDb > 0° when flipping into the major groove pathway, whereas it aligns with CPDb < 0° when flipping into the minor groove pathway. In Fig. 2a), the spontaneous flipping of uracil through the major groove has a lower free energy barrier than through the minor groove in both CPDb and CPD schemes, respectively. The lower free energy barriers and their corresponding base opening



**Fig. 1.** (a) Uracil-mutated dsDNA system with red and blue colors representing uracil and adenine residues, respectively. (b) Root mean square deviation (RMSD) and (c) Root mean square fluctuation (RMSF) of the uracil-mutated dsDNA system during the first 100 ns of the MD trajectory.



**Fig. 2.** (a) Free energy profile (Potential of Mean Force, PMF) as a function of the Center Of Mass (COM) of the pseudo-dihedral angle for the uracil-mutated dsDNA; (b) The representative snapshots for the uracil base flipping steps [here region I is the Watson-Crick base pair, II, III, V regions are the intermediate steps, and IV is the fully flipped state of uracil].

angles were about 7.5 kcal/mol and  $30^\circ$  for CPDb and 6 kcal/mol and  $60^\circ$  for CPD schemes, respectively. This observation is in good agreement with the previous meta-eABF simulation<sup>41</sup>. The flipped-out pseudo dihedral angle uracil was about  $180^\circ$ , and the corresponding free energy barriers were about 8.2 kcal/mol and 9 kcal/mol for CPDb and CPD schemes, respectively. The transition from the flipped-in to the flipped-out state of uracil has a higher free energy barrier in the CPD than in the CPDb schemes. The obtained results suggest that the observed differences could be attributed to the effect of the reaction coordinate definition on the details of the PMF. Furthermore, the results indicate that uracil flips out more frequently through the major groove pathway than through the minor groove pathway.

In order to understand the structural properties of uracil base opening, analyses of the hydrogen bond distances ( $N1 \cdots H3$  and  $O4 \cdots H6'$ ) and the center of mass (COM) of distance between uracil and adenine were performed for the metadynamics simulation trajectories, as displayed in Fig. 3.

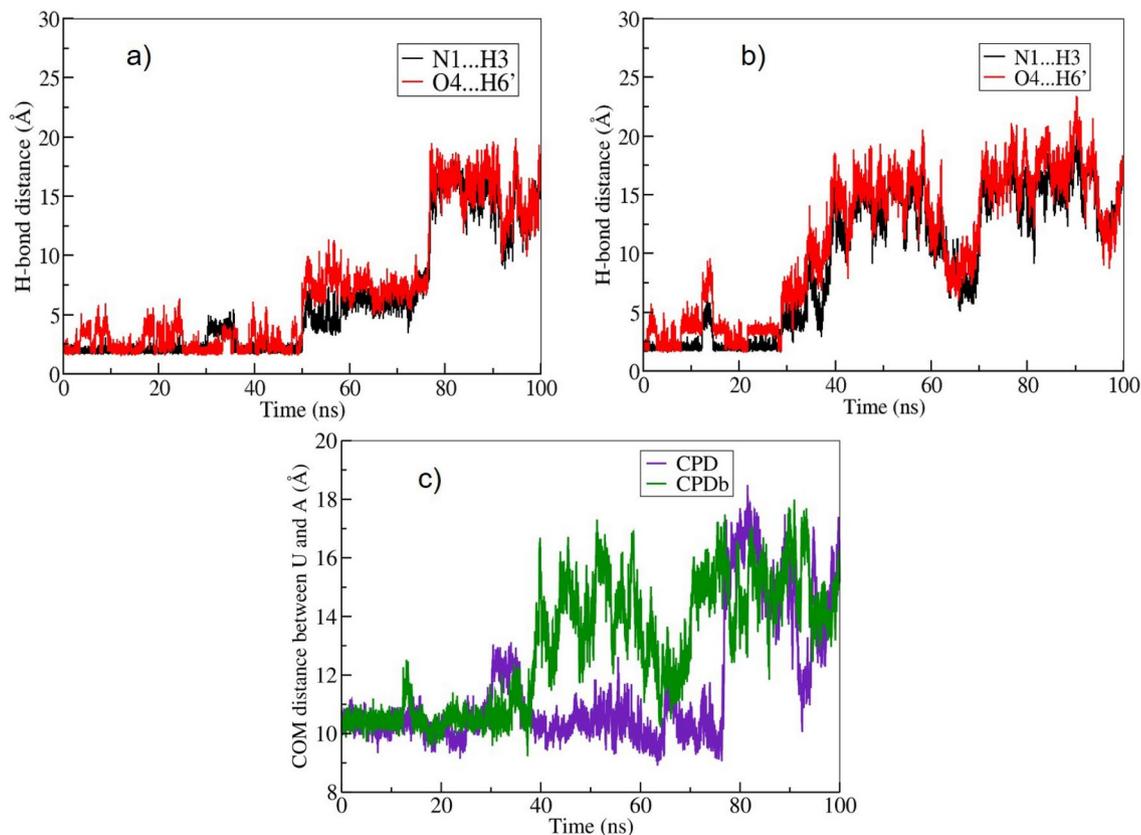
At the beginning of the simulation, the COM distance between uracil and adenine bases were about 10.5 Å. During the simulation between 30–40 ns, this distance increases slightly to 12.5 Å. After 75 ns (Fig. 3c), the uracil base is completely flipped-out, causing the distance to increase to 17 Å, whereas the hydrogen bond (H-bond) distance between  $N1(A) \cdots H3(U)$  and  $O4(U) \cdots H6'(A)$  increased from 1.82 Å to 15.9 Å and from 2.12 Å to 18.4 Å, respectively; see Fig. 3a). In the CPDb (Fig. 3b) scheme, the COM distance increases from 10.5 to 15 Å, whereas the distances of the H-bonds  $N1 \cdots H3$  and  $O4 \cdots H6'$  have increased from 1.82 Å to 16.2 Å and from 2.08 Å to 16.4 Å, respectively.

The RDF analysis shows that, in both CPD and CPDb trajectories, a water molecule closely approaches the  $N6-H6 \cdots O4$  bond between the uracil and adenine bases at an approximate distance of 1.9 Å during the initial simulation time. In contrast, no water molecules are found near the  $N3-H3 \cdots N1$  bond between the uracil and adenine bases (Supplementary Fig. S2). The RDF probability values of  $N6-H6 \cdots O4$  for CPD and CPDb are approximately 0.2, which is further verified by the Independent Gradient Model (IGM)<sup>42</sup>, which provides insight into non-covalent interactions (NCI) between the molecules. During the initial stages of the simulation (0–30 ns), a transient water molecule approaches the  $O4$  atom of the uracil base at a distance of 1.9 Å, forming a weak interaction (shown in Supplementary Fig. S3). The corresponding IGM isosurface value is 0.035, indicating a very weak interaction. Notably, this weak interaction does not influence the disruption of the  $N6-H6 \cdots O4$  bond. Therefore, solvent effects do not play an important role in uracil base flipping under the studied conditions.

Further, we have analysed the uracil-mutated dsDNA structural deformations. For this we used the Curves+ program<sup>43</sup>. The results are given in Fig. 4 and the Supplementary Fig. S4. The intrabase parameters (buckle, opening) and the interbase parameter (tilt) for the flipped-in to flipped-out transition of uracil in the UA base pair ranged from  $-2.5^\circ$  to  $-57.4^\circ$ ,  $10.7^\circ$  to  $168.2^\circ$ , and  $-1.6^\circ$  to  $16.0^\circ$ , respectively. The total dsDNA bending angle increased from  $3.6^\circ$  to  $68.6^\circ$ . In particular, the major and minor groove values of flipped-out uracil dsDNA were about 11 Å and 7 Å, respectively. Since a wider major groove was confirmed more favourable for base flipping, the observed increase in the major groove width favors a structural adaptation that facilitates uracil flipping in the DNA sequence<sup>37</sup>.

### Docking UDG to flipped uracil bases in dsDNA

Two isoforms of human UDG are known to originate from the alternative splicing of the UNG gene. The canonical isoform is UNG2 (Uniprot P13051-1) and has a length of 313 aa. It is expressed in the mitochondrion



**Fig. 3.** (a) and (b) the hydrogen bond distances between N1...H3 and O4...H6' for the CPD and CPDb schemes during the course of the simulation, respectively; (c) Likewise, the calculated average centre of mass of the distance between the uracil and adenine residues for CPD and CPDb schemes.

and is the one referred in the introduction as the first glycosylase characterized in complex with damaged DNA. The second isoform is UNG1 (Uniprot P13051-2, 304 aa), which is expressed in the nucleus and hence the isoform we used for our docking simulations. The difference between these two isoforms lies in their N-terminal beginning portions: UNG1 from residue 36 to its C-terminal end (304) is identical to UNG2 from residue 45 to its C-terminal end (residue 313). The structure of the predicted UNG1 structure we use is shown in Fig. 5a).

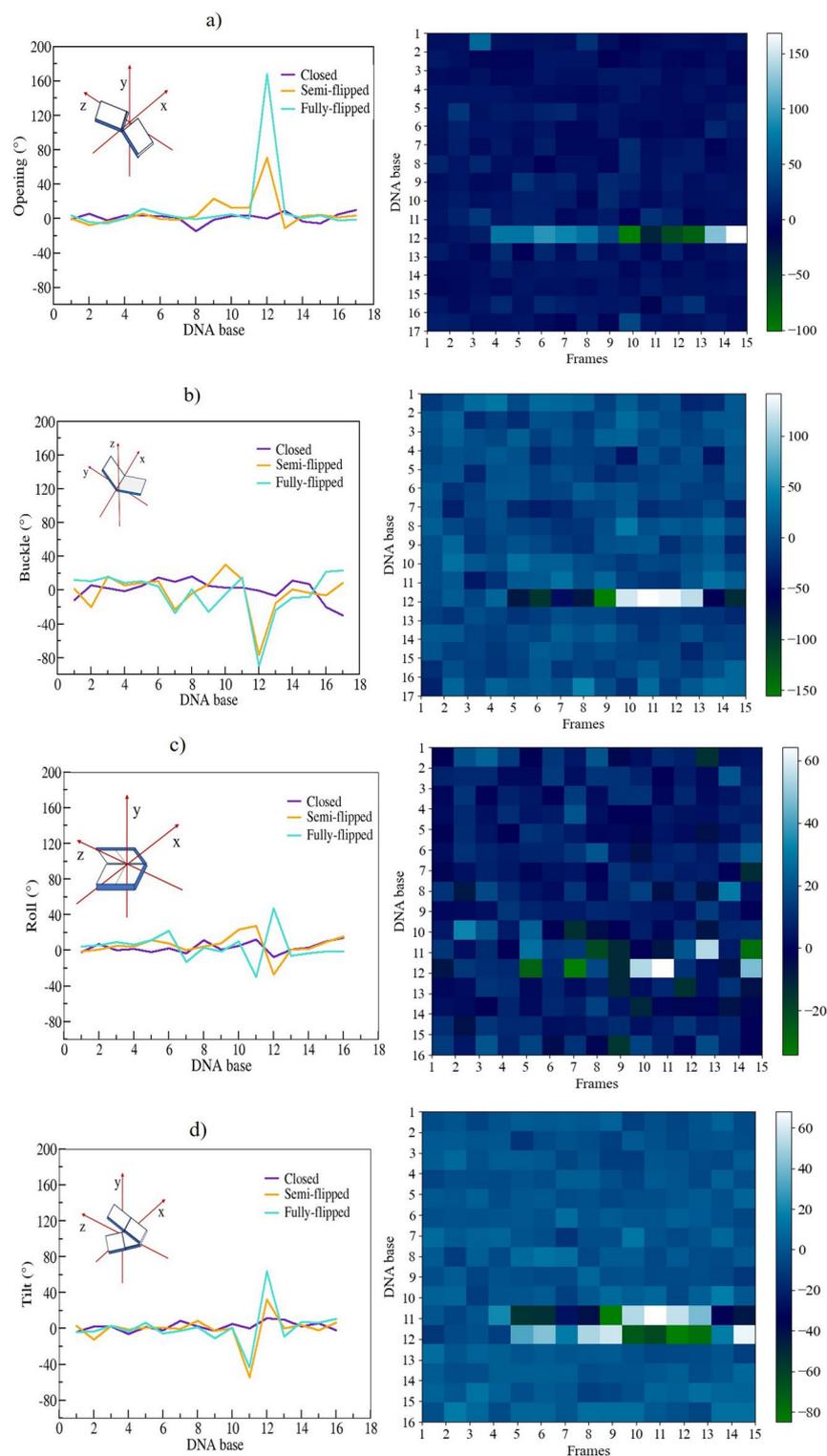
As a first step in studying the docking process we have investigated the flexibility of UDG, since the docking of proteins to DNA is clearly influenced by the flexibility of both docking partners. The results of our investigation are reported in Fig. 5b). The flexibility information obtained with the MEDUSA webserver<sup>46</sup>(see “Methods”) reveals that the portion containing UDG’s catalytic site is predicted as mostly rigid. Upstream of this rigid section the protein is less rigid, even ending on a disordered and highly flexible N-terminal tail.

To study protein-DNA interactions, several docking algorithms and programs have been developed and are mostly available either from a webserver or with its code made publicly available, see “Methods”<sup>47–52</sup>. We have tested these docking programs on the crystal structure PDB:1EMH which we take as our reference structure for the recognition complex. The results of these docking attempts are summarized in Supplementary Fig. S6. We found that pyDockDNA performed best because it manages to discriminate between uracil and thymine, and was therefore selected for our docking simulations (see “Methods”).

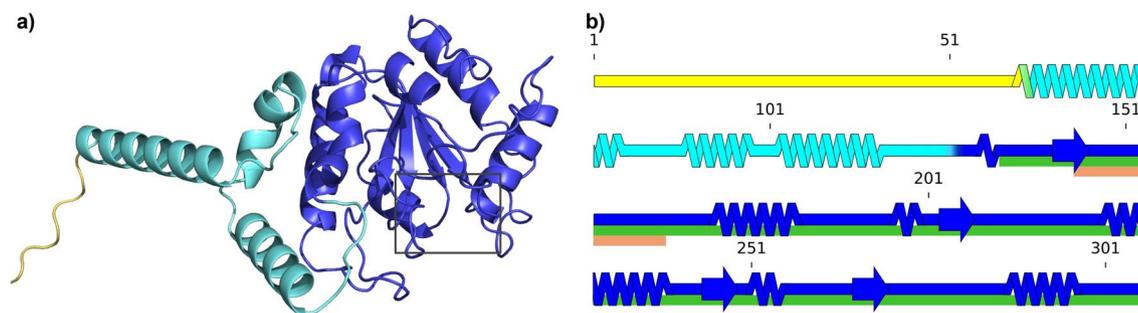
In order to have a global view of UDG’s encounter with the damaged DNA strand, we took snapshots from the MD simulations of either the trinucleosome structure, as obtained in our previous work<sup>26</sup>, or from the MD base-flipping simulations described above. We denote these structures by either a ‘T’ for trinucleosome or ‘O’ for the oligomer (17-mer) dsDNA in the following. Table 1 collects the results of the measurement of suitable angles and distances to characterize the base-flip structures selected for the docking process. We characterized the docked structure by a distance and an angle measurement which is described in “Methods”.

A common feature appearing for every successful docking, as shown in Table 1, is a sufficiently wide opening angle of the flipped-out uracil base and a reasonable distance measure (the lesser being 13 Å) to the opposite DNA strand’s backbone. Both conditions are to be fulfilled for UDG to have enough space to reach the damaged base. These two conditions appear to be the main factors for a high success in docking as compared to the less successful conformations.

To analyse the docking results, we first filtered the conformations sampled by pyDockDNA. From the initial 10,000 docked structures that pyDockDNA produces with FTDOCK, it selects the 100 best according to a scoring function, which in this case is the one with no desolvation. According to the Supplementary Fig. S7a), this associated score is not sufficient to differentiate between the good and the bad docking solutions.



**Fig. 4.** Intra-base pair parameters: **(a)** Opening and **(b)** Buckle, and inter-base pair parameters **(c)** Roll and **(d)** Tilt of uracil-mutated dsDNA for the CPDb scheme were calculated using the CURVES+ program. The plots on the right hand side are the respective heat maps of Opening, Buckle, Roll, and Tilt with respect to the trajectory frames. Here, the dark shades (blue to green) indicate negative values, and the light shades (cyan to white) indicate positive values of these quantities for the uracil-mutated dsDNA.



**Fig. 5.** (a) Cartoon representation of UDG tertiary structure predicted with Alphafold2 (AF2) as described in “Methods”. The structure is represented with PyMOL<sup>44</sup>. Its residues are colored according to the backbone qualitative flexibility evaluation described in the main text and in “Methods”. The three flexibility levels are: *yellow* for flexible, *blue* for rigid and *cyan* for intermediate level between flexible and rigid. In (a), the black frame at the bottom right designates UDG’s catalytic pocket; (b) 2D representation of UDG secondary structure with the same color code as in (a), drawn with SSDraw<sup>45</sup>. Notable regions are highlighted under the concerned residues: the catalytic site is shown in *orange* (InterPro IPR018085) and the conserved UDG domain in *green* (InterPro IPR005122).

	PDB ID:1EMH	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Dihedral angle (°)	-171.75	-5.85	2.61	-4.71	-3.16	-8.66	-66.12	-96.58	-83.18	-108.82	145.12
$d_{U-groove}$ (Å)	20.8	11.8	12.9	14.2	11.5	12.9	12.0	10.4	10.8	11.1	8.3
Success	<b>High</b>	None	None	None	None	None	Low	Medium	Medium	Medium	Medium
	T11	T12	O1	O2	O3	O4	O5	O6	O7	O8	
Dihedral angle (°)	-119.20	175.9	41.8	29.8	151.9	-173.2	-103.6	151.7	103.9	160.5	
$d_{U-groove}$ (Å)	12.8	10.4	14.4	15.2	13.2	16.4	10.9	15.8	10.0	13.3	
Success	Medium	Medium	None	None	<b>High</b>	<b>High</b>	Medium	<b>High</b>	Low	<b>High</b>	

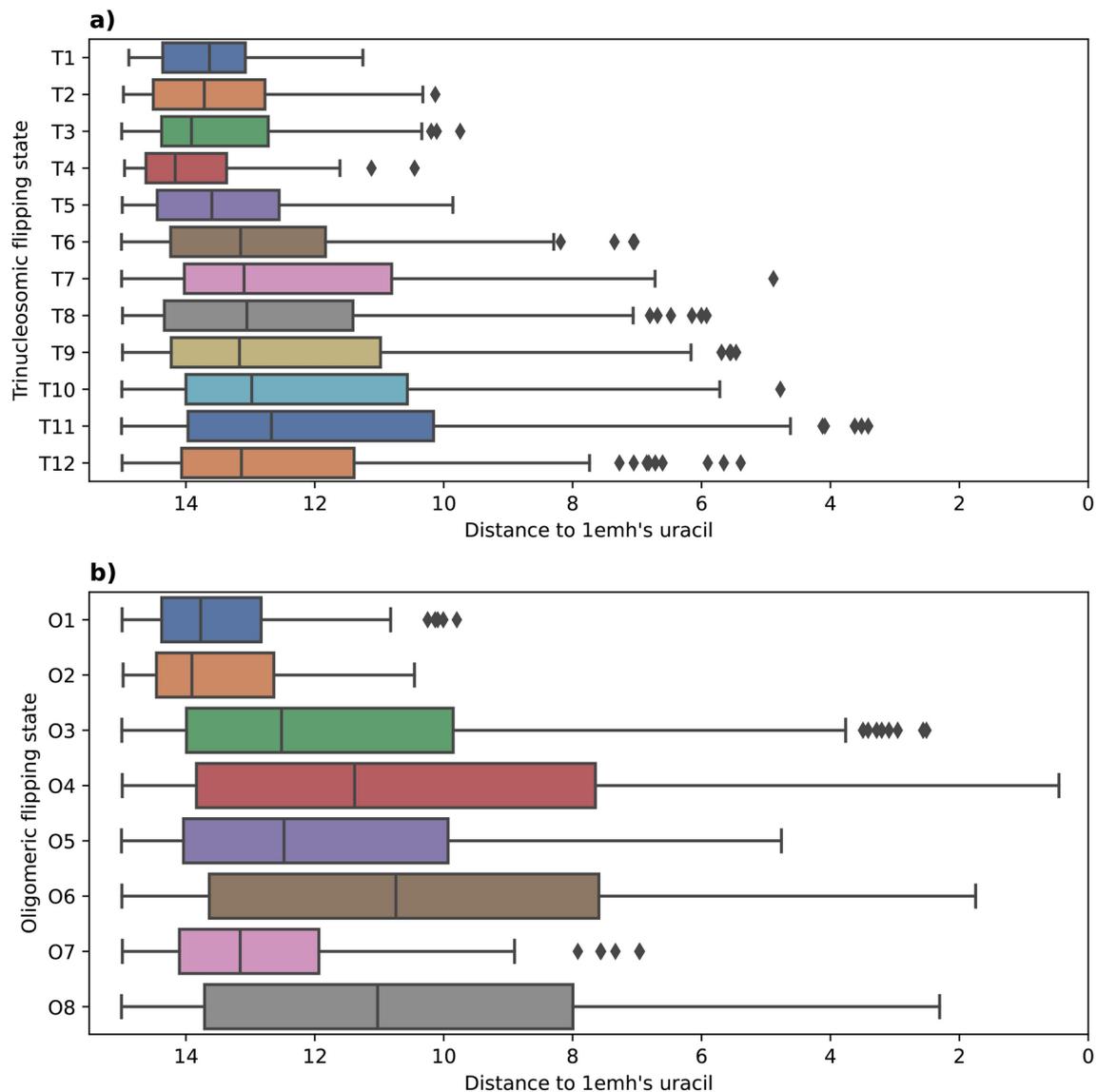
**Table 1.** Quantitative description of impacting features for the structures ‘T’ from the trinucleosome simulations and ‘O’ from the 17-mer dsDNA oligomers, resulting from a gradual base flipping with their docking success evaluation. The distance  $d_{U-groove}$  measurements and *Dihedral angle* metrics are described in Supplementary Fig. S5. The angle evaluates the extent of the base flipping, the distance indicates the minimal distance between the base and both edges of the groove. The closer the base is from the backbone of the opposite strand, the more sterically hindered it is. The sign “-” in dihedral angles implicates flipping towards the minor-groove.

In order to sort the predictions by relevance, i.e. by how well the obtained interface corresponds to the one observed in PDB ID:1EMH, we went back to the initial 10,000 sampled conformations and devised a protocol described in “Methods” which measures the similarity of the uracil position between the docking structure and PDB ID:1EMH used as our reference structure. This measure, represented in Fig. 6 is then translated to an indicator listed in Table 1 as ‘success’. According to Fig. 6, the DNA strand with the highest success with a near perfect similarity to PDB ID:1EMH is ‘O4’. Its docking success coincide with the structure having the highest value in both the opening angle and the groove width according to Table 1.

According to these measures, the highest docking success of both ‘T’ dsDNA and ‘O’ dsDNA (see Fig. 6), respectively ‘T11’ and ‘O4’, were selected for illustration purposes in Fig. 7. It denotes the high similarity of these docking successes to the PDB ID:1EMH structure. ‘T12’ has a way wider flipping angle while being less successful than ‘T11’ (Fig. 6a) for docking to UDG. The determining factor making ‘T11’ the most prone among the ‘T’ dsDNA to interacting with UDG seems to be the large distance between its uracil and the nearest groove’s end. On the other hand, ‘O4’ has at the same time the widest angle of base flipping and the most space in the groove making it the top pick for a successful docking.

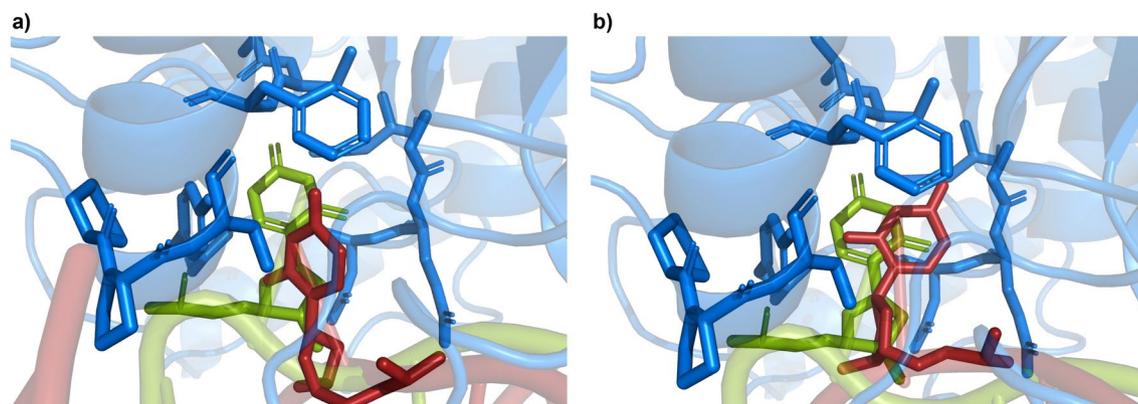
### Structural dynamics of uracil-flipped dsDNA bound to UDG enzyme

The optimal docking conformation of uracil-flipped dsDNA with the UDG enzyme (structure ‘O4’) was further explored using all-atom molecular dynamics simulations for 500 ns. The superposition of the docked complex and the final snapshot of the MD simulation are shown in Supplementary Fig. S9. The overall structural stability, compactness, and the flexibility of the complex system were evaluated using RMSD, radius of gyration ( $R_g$ ), and RMSF analyses. The mean RMSD values of the protein, the protein within the complex, and the protein-dsDNA complex were computed as 2.77 Å, 2.35 Å, and 3.14 Å, respectively, as shown in the Fig. 8a). The compactness of all three systems was assessed using  $R_g$  and the probability distribution of  $R_g$ , as illustrated in the Fig. 8b and c. UDG initially exhibited higher fluctuations, around 20 Å, between 0 and 100 ns. After 100 ns, these

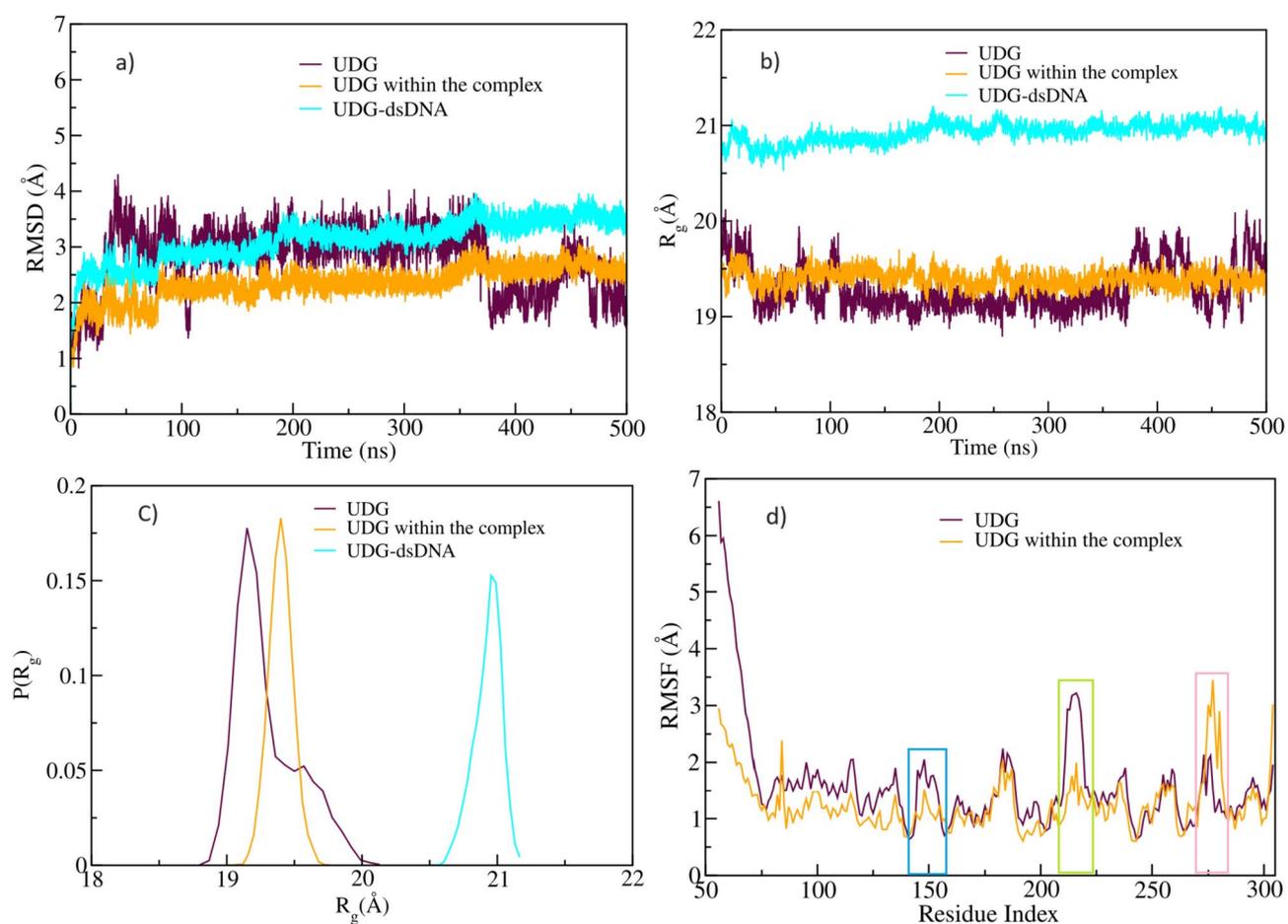


**Fig. 6.** Distance of the docking-predicted uracil relative to the one of PDB ID: 1EMH, depending on the DNA structure used for docking with trinucleosomic strands in (a) and oligomeric strands in (b). The distance (Å) is computed after superimposing both UDG from the docked structure and from PDB ID: 1EMH. Only distances under 15 Å are conserved. X-axis goes from 15 Å to 0 Å. Distances are computed as described in “Methods” and illustrated in Supplementary Fig. S8.

fluctuations began to decrease to approximately 19.5 Å. However, after 350 ns, the UDG experienced increased fluctuations, primarily due to the loop present in the catalytic site (residues 210–220), which became more flexible and fluctuated significantly during the simulation. Consequently, the probability distribution of  $R_g$  adopted a bimodal conformation. In contrast, the  $R_g$  values for the UDG within the complex and the UDG-dsDNA system remained stable throughout the simulation, with values of 19.4 Å and 21 Å, respectively. The flexibility of the systems was further analysed using RMSF, as shown in the Fig. 8d. The flexibility of the loop present in UDG is higher than that in the complex system, as shown in the Fig. 8d. Notably, the loop residues (residues 145–158) located at the catalytic sites exhibit slight flexibility, with an RMSF of approximately 2.4 Å. The presence of uracil at the catalytic sites in the complex stabilizes the loop, rendering it more rigid due to the formation of hydrogen bonds with GLN144 (URA(O4)···H), ASP145 (URA(H5)···O), and PRO167 (URA(O2)···H) which are shown in Supplementary Fig. S10. The loop residues 210–220 in the protein system are highly flexible, with an RMSF range of approximately 3.0–3.8 Å. These residues are further stabilized by the dsDNA during complex formation. In the complex, amino acids 272–278 exhibit higher fluctuations because the LEU272 residue in the loop intercalates into the nucleic acid strand to fill the gap created by uracil flipping into the catalytic site of the UDG enzyme, which correlates well with previous experimental findings<sup>2</sup>. As for the initial stages of base flipping we do not find a relevant influence of the solvating waters on the binding of UDG to the flipped uracil base.



**Fig. 7.** Docked structure for the highest success of both types of induced base flipping ('T' and 'O'), identified according to Fig. 6. **(a)** 'T11' DNA strand (red) docked to UDG (blue; cartoon) with a focus on its catalytic pocket (blue; sticks). The uracil (red; sticks) from the docked DNA is compared to the one from PDB ID:1EMH (green; sticks) as the reference. Following the same color code, **(b)** shows the 'O4' DNA strand docked to UDG with PDB ID:1EMH uracil as a reference to assess how well docking predicts the entrance of uracil into the catalytic site of UDG.



**Fig. 8.** Structural analyses for the UDG-dsDNA complex. **(a)** Root mean square deviation (RMSD), **(b)** radius of gyration ( $R_g$ ), **(c)** probability distribution of  $R_g$ , and **(d)** root mean square fluctuation (RMSF). Here, the blue box highlights the catalytic site of the UDG enzyme, the green box indicates the fluctuation of loop residues (residues 210 to 220), and the pink box indicates the fluctuation of LEU272 intercalated into the dsDNA within the UDG-dsDNA complex.

## Discussion

In this work we present arguments that the mechanism by which UDG identifies uracil bases in dsDNA cannot rely alone on a random search combined with the ‘pinch-push-pull’-model. We purport that Uracil bases can present themselves to UDG via stochastic, thermally-induced base-flipping, but even more readily due to the ubiquitous presence of forces acting on dsDNA. As we have shown earlier<sup>26</sup>, this leads to instabilities of the dsDNA in the course of which bases flip out and become accessible to UDG. Here, we have studied whether this scenario is in accord with the base-flipping process of uracil from dsDNA, and whether the resulting structures can be properly recognized by UDG, i.e. whether a recognition complex between uracil and UDG can be formed. Our approach consisted in studying the base-flipping of uracil from dsDNA oligomers, using a collective variable/metadynamics approach within molecular dynamics simulations. In these simulations, uracil-flipped states were generated, passing through different stages of the flipping process in time. In parallel, we have selected base-flipped structure from our earlier large-scale MD simulations of trinucleosome arrays, in which base-flipping has occurred in the linker DNA of the arrays in the course of a mechanical instability in the compressed dsDNA. The opening angles of all of these structures have been quantified.

Subsequently, we have performed a study of rigid docking of UDG to the base-flipped uracil in dsDNA. While several protein-DNA docking programs have been developed in recent years, the study of protein-DNA docking is generally much less developed than the methodologies for protein-protein docking. In addition, the case of base-flipping requires the docking to be performed not towards a pristine, but rather a dynamically perturbed structure. This requirement has both necessitated a detailed study of available protein-dsDNA docking softwares, as well as a careful investigation of the docking success, i.e. the formation of a ‘proper’ recognition complex. We found pyDockDNA to be the best suited software for our task. For the evaluation of the quality of the docking for our predicted recognition complexes, we took the crystal structure of UDG-dsDNA from PDB ID:1EMH as our reference structure. Our careful evaluation of the docking quality for both the oligomeric and trinucleosomic dsDNA is documented in Supplementary Fig. S6. Overall, we identify two crucial factors in obtaining UDG-dsDNA recognition complexes, both of which are related to the opening angle of the uracil base. Firstly, we find that only a nearly fully flipped base (meaning between an angle of 150° and 180°) allows for recognition complex formation. The second crucial factor is the increase in width of the groove into which the base opens. The latter fact is well-known from earlier experimental studies of uracil embedded in nucleosomal DNA<sup>53</sup>. If these conditions are met we find that the formation of a high-quality recognition complex according to our criteria, as given in “Methods”, is possible. This is supported by MD simulations we performed on the best obtained complex.

Therefore, our study provides evidence that, besides the binding of UDG to dsDNA and its possible participation in a ‘pinch-push-pull’ mechanism, mechanical instabilities of dsDNA in the crowded nuclear environment that favor a destabilization of the dsDNA, and therefore enhance the probability of base flipping beyond a merely random, thermally assisted event, may represent a relevant, alternative or complementary pathway allowing the identification of uracil by UDG, thereby initializing the base-excision process. Our finding is further supported by the discussion of the effect of DNA deformation on protein-DNA complex formation in<sup>25</sup>.

The initial BER recognition process is in fact occurring in a highly dynamic environment and very closely linked to DNA dynamics, at least in chromatin linker DNA. In future work we intend to look at the role of DNA dynamics in nucleosomes on the BER process.

## Methods

### Selecting and preparing the MD-snapshots ‘T’ from the trinucleosome simulations

We have selected twelve snapshots from the trinucleosome simulations described in<sup>26</sup>. They were chosen from the structure called T183, which was reconstructed by assembling three copies of the 197 nucleosome (PDB file 5NL0)<sup>54</sup>. More details as well as access to the trinucleosome structure can be found in<sup>26</sup>. The sequence of the dsDNA sequence extracted from that structure is given by 5’-TACGUATGG-3’. In the original MD-simulations a fully paired dsDNA structure was used. The mechanically flipped-out base was replaced by uracil using the ‘Mutation’ section in the w3DNA 2.0 webserver<sup>55</sup>.

### MD simulations of uracil flipping in the dsDNA oligomer ‘O’

All atom molecular dynamics (MD) simulations were executed with the NAMD 2.14 package<sup>56</sup> with the CHARMM 36 force fields<sup>57</sup>. A concentration of 0.15 mol/L Na<sup>+</sup> and Cl<sup>-</sup> ions were added at random positions to maintain the charge neutrality of the dsDNA system. The all-atom simulations employed periodic boundary conditions (PBC) and multiple time-stepping wherein local interactions were considered every 2 fs and full electrostatic evaluations were conducted every 2 time steps. The particle mesh Ewald method (PME) was employed for long-range electrostatic calculations<sup>58</sup>. The simulation box size was 79.7 × 40.6 × 40.9 Å<sup>3</sup> with a distance of 10 Å between the box edges and the dsDNA. The cutoff and switching distances were set at 12 Å and 10 Å, respectively. Covalent bonds involving hydrogen were held rigid by the RATTLE<sup>59</sup> and SETTLE algorithms<sup>60</sup>. The dsDNA system was minimized for 5000 steps using the conjugate gradient method. The pressure (constant NPT) was maintained at 1 atm, and the temperature was kept at 310 K. Temperature control was attained through Langevin dynamics for all non-hydrogen atoms, while pressure was controlled using a Nose-Hoover Langevin piston. The trajectories of the simulations were visualized and analysed by the visual molecular dynamic program (VMD 1.9.4)<sup>61</sup> and the dsDNA conformational analysis were calculated by Curves+ program<sup>43</sup>.

### Choice of collective coordinates

In this study, the reaction coordinates have been described by three choices i) the centre-of-mass (COM) of pseudo-dihedral angle between the flipping base, the sugar group of the same nucleotide (group 1), the sugar group of the next nucleotide (group 2), the base of the next nucleotide plus the base of the opposing nucleotide (group 3), and the target base (group 4); ii) the center of the two flanking base pairs (group1), group 2 and 3 are defined by the flanking sugar moieties, and group 4 is defined by the target base for the flipping process; and iii) in the CPDb scheme, the group 1 is same as in CPDa, group 2 and 3 are defined by the flanking phosphate group, and the group 4 is defined by the target base for the flipping process. Among these choices, the uracil base opening event did not ensue with the CPDa scheme. As a result, all calculations described here were conducted using the CPD and CPDb schemes. The potential mean force was calculated for reaction coordinate between  $-180^\circ$  and  $180^\circ$  into windows of  $5^\circ$  degree width; settings for Gaussian hillWeight = 0.001 kcal/mol and hillWidth = 2 bin width were added in the metadynamics simulation.

### Definition of docking partners

Docking here consists in simulating the interaction between UDG and the base-flipped DNA. The DNA strands used in these simulations are from the previous step of producing 'O' dsDNA and snapshot selection for 'T' dsDNA described above. As for 'O' dsDNA, over the 14,000 frames of the simulations, we selected the 8 following frames, ranging that we call O1 to O8, respectively: 6983, 7269, 7712, 8272, 9040, 10145, 10394, 11433. For UDG, the structure has been predicted by AlphaFold2 (AF2) with the isoform sequence from Uniprot ID P13051-2. The 56 first residues were discarded as they are in a disordered region with no confidence in its position. This structure is represented in Fig. 5 with PyMOL<sup>44</sup>.

### Measurements of angles and distances in the base-flipped conformations

The opening angles for the flipped-out uracil bases in Table 1 are determined via the collective variable approach, as defined in the text. The method has been used for both the trinucleosome ('T') and control ('C') structures. The  $d_{U-groove}$  measure is calculated between the flipping out uracil and the nearest groove's extremity. This distance spans between the N1 uracil atom and the nearest phosphate from the opposite strand. For an illustration, see Supplementary Fig. 3a, in which four positions P1, P2, P3, and P4 are defined representing the pseudo-dihedral points used to measure the uracil flipping angle around its backbone during the simulation.

### UDG flexibility

Flexibility was predicted using the MEDUSA webserver<sup>46</sup>, <https://www.dsimb.inserm.fr/MEDUSA/>. It predicts flexibility from the protein amino acids sequence in different modes. Each of these modes assign a probability of belonging to a predefined number of flexibility class per residue. The mode we have used is 3 class prediction; the results are summarized in Fig. 5a and b. The structure in Fig. 5a is represented with PyMOL<sup>44</sup>, Fig. 5b is drawn with SSDraw<sup>45</sup>. The flexibility classes assigned according to MEDUSA are used as coloring factors; the explanation is given in the Figure caption.

### Selection of protein-DNA docking software

In order to determine which was the best-suited available software for UDG-dsDNA docking, we have tested different types of docking software. The complete list is given by: HDock<sup>49</sup>, HADDOCK<sup>62</sup>, pyDockDNA<sup>51</sup>, MDockPP<sup>63</sup>, RoseTTAFold2NA<sup>64</sup> and AlphaFold3<sup>65</sup>. They were all evaluated on reproducing the interaction in the 1EMH crystal structure. From our tests, pyDockDNA performed the best on our use case as it discriminates between Uracil and Thymine, thus being the one selected to run the docking simulations. The summary of the comparison is found in Supplementary Fig. S4.

### Evaluation of the docking structures

This protocol takes the PDB ID:1EMH structure as a reference for the highest docking quality; it is the target which we wanted to reproduce by docking, and putting it more precisely, we measured the similarity with which the uracil enters the UDG compared to PDB ID:1EMH. After aligning PDB ID:1EMH UDG and the UDG of every docking conformation for each DNA strand with different base flipping angle and protocol (10,000 docking poses per simulation), we measured the distance between the N3 atoms of the evaluated uracil (U1) and the uracil analog 2'-deoxypseudouridine of the reference uracil (U2). This  $d_{U1-U2}$  distance (in Å) is a direct quantitative evaluation of how much the docking conformation corresponds to the experimental structure. Figure 6a and b both represent this evaluation performed on the trinucleosomic structure ('T' in a) and the oligomeric structure 'O' in b). The goal of this measure was to highlight which DNA strand features are impacting UDG action the most, thus the most successful docking conformation for each strand was taken and its quality was classified into 'None', 'Low', 'Medium' and 'High' quality according to the thresholds:

- 'None':  $d_{U1-U2} > 9 \text{ \AA}$
- 'Low':  $9 \text{ \AA} > d_{U1-U2} > 6 \text{ \AA}$
- 'Medium':  $6 \text{ \AA} > d_{U1-U2} > 3 \text{ \AA}$
- 'High':  $d_{U1-U2} < 3 \text{ \AA}$

## Data availability

All data and scripts of the article are hosted on this GIT repository: [https://gitlab.in2p3.fr/cmsb-public/udg\\_dna](https://gitlab.in2p3.fr/cmsb-public/udg_dna).

Received: 23 September 2024; Accepted: 21 February 2025

Published online: 07 March 2025

## References

- Slupphaug, G. et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92. <https://doi.org/10.1038/384087a0> (1996).
- Schormann, N., Ricciardi, R. & Chattopadhyay, D. Uracil-DNA glycosylases: Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **23**, 1667–1685. <https://doi.org/10.1002/pro.2554> (2014).
- Parikh, S. S., Putnam, C. D. & Tainer, J. A. Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res./DNA Repair* **460**, 183–199. [https://doi.org/10.1016/S0921-8777\(00\)00026-4](https://doi.org/10.1016/S0921-8777(00)00026-4) (2000).
- Friedman, J. I. & Stivers, J. T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957–4967. <https://doi.org/10.1021/bi100593a> (2010).
- Stivers, J. T. Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry* **38**, 952–963. <https://doi.org/10.1021/bi9818669> (1998).
- Bruner, S. D., Norman, D. P. G. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866. <https://doi.org/10.1038/35002510> (2000).
- Yang, C.-G. et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961–965. <https://doi.org/10.1038/nature06889> (2008).
- Orndorff, P. B. et al. Uracil-DNA glycosylase efficiency is modulated by substrate rigidity. *Sci. Rep.* **13**, 3915. <https://doi.org/10.1038/s41598-023-30620-0> (2023).
- Jena, N. R. & Bansal, M. Mutagenicity associated with O6-methylguanine-DNA damage and mechanism of nucleotide flipping by AGT during repair. *Phys. Biol.* **8**, 046007 (2011).
- Vivas, P., Velmurugu, Y., Kuznetsov, S. V., Rice, P. A. & Ansari, A. Mapping the transition state for DNA bending by IHF. *J. Mol. Biol.* **418**, 300–315. <https://doi.org/10.1016/j.jmb.2012.02.028> (2012).
- Baral, S. et al. Evidence for intrinsic DNA dynamics and deformability in damage sensing by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res.* **53**, gkae1290. <https://doi.org/10.1093/nar/gkae1290> (2025).
- Lee, A. J. & Wallace, S. S. Hide and seek: How do DNA glycosylases locate oxidatively damaged DNA bases amidst a sea of undamaged bases?. *Free Radic. Biol. Med.* **107**, 170–178. <https://doi.org/10.1016/j.freeradbiomed.2016.11.024> (2017).
- Porecha, R. H. & Stivers, J. T. Uracil-DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl. Acad. Sci.* **105**, 10791–10796. <https://doi.org/10.1073/pnas.0801612105> (2008).
- Moe, J. G. & Russu, I. M. Kinetics and energetics of base-pair opening in 5'-d(cgcgaattcgcg)-3' and a substituted dodecamer containing G.T mismatches. *Biochemistry* **31**, 8421–8428. <https://doi.org/10.1021/bi00151a005> (1992).
- Várnai, P., Canalia, M. & Leroy, J. L. Opening mechanism of G.T/U pairs in DNA and RNA duplexes: a combined study of imino proton exchange and molecular dynamics simulation. *J. Am. Chem. Soc.* **126**, 14659–14667. <https://doi.org/10.1021/ja0470721> (2004).
- Yin, Y., Yang, L., Zheng, G. & Zhao, X. S. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. *Proc. Natl. Acad. Sci. USA* **111**, 8043–8048. <https://doi.org/10.1073/pnas.1400667111> (2014).
- Nicy, C. D. & Wales, D. J. Energy landscapes for base-flipping in a model DNA duplex. *J. Phys. Chem. B* **126**, 3012–3028. <https://doi.org/10.1021/acs.jpcc.2c00340> (2022).
- Fuxreiter, M., Luo, N., Jedlovsky, P., Simon, I. & Osman, R. Role of base flipping in specific recognition of damaged DNA by repair enzymes. *J. Mol. Biol.* **323**, 823–834. [https://doi.org/10.1016/S0022-2836\(02\)00999-3](https://doi.org/10.1016/S0022-2836(02)00999-3) (2002).
- 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.* **11**, 1230–1236. <https://doi.org/10.1038/nsmb864> (2004).
- Parker, J. B. et al. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449**, 433–437. <https://doi.org/10.1038/nature06131> (2007).
- Giudice, E. & Lavery, R. Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc.* **125**, 4998–4999. <https://doi.org/10.1021/ja034095f> (2003).
- Priyakumar, U. D. & MacKerell, A. D. Computational approaches for investigating base flipping in oligonucleotides. *Chem. Rev.* **106**, 489–505. <https://doi.org/10.1021/cr040475z> (2006).
- Bellamy, S. R., Krusong, K. & Baldwin, G. S. A rapid reaction analysis of uracil DNA glycosylase indicates an active mechanism of base flipping. *Nucl. Acids Res.* **35**, 1478–1487 (2007).
- Churchill, C. D. M., Eriksson, L. A. & Wetmore, S. D. DNA distortion caused by uracil-containing intrastrand cross-links. *J. Chem. Phys. B* **120**, 1195–1204 (2016).
- Taranova, M., Hirsh, A. D., Perkins, N. C. & Andricioaei, I. Role of microscopic flexibility in tightly curved DNA. *J. Phys. Chem. B* **118**, 11028–11036. <https://doi.org/10.1021/jp502233u> (2014).
- Cleri, F., Giordano, S. & Blossey, R. Nucleosome array deformation in chromatin is sustained by bending, twisting and kinking of linker DNA. *J. Mol. Biol.* **435**, 168263. <https://doi.org/10.1016/j.jmb.2023.168263> (2023).
- Hwa, T., Marinari, E., Sneppen, K. & Tang, L.-H. Localization of denaturation bubbles in random DNA sequences. *Proc. Natl. Acad. Sci. USA* **100**, 4411–4416. <https://doi.org/10.1073/pnas.0736291100> (2003).
- Toussaint, J.-P. *L'instant précis où Monet entre dans l'atelier* (Les Éditions de Minuit, 2022).
- Várnai, P. & Lavery, R. Base flipping in DNA: Pathways and energetics studied with molecular dynamics simulations. *J. Am. Chem. Soc.* **124**, 7272–7273. <https://doi.org/10.1021/ja025980x> (2002).
- Bouvier, B. & Grubmüller, H. A molecular dynamics study of slow base flipping in DNA using conformational flooding. *Biophys. J.* **93**, 770–786. <https://doi.org/10.1529/biophysj.106.091751> (2007).
- Song, K. et al. An improved reaction coordinate for nucleic acid base flipping studies. *J. Chem. Theory Comput.* **5**, 3105–3113. <https://doi.org/10.1021/ct9001575> (2009).
- Lindahl, V., Villa, A. & Hess, B. Sequence dependency of canonical base pair opening in the DNA double helix. *PLoS Comput. Biol.* **13**, 1–22. <https://doi.org/10.1371/journal.pcbi.1005463> (2017).
- Lemkul, J. A., Savelyev, A. & MacKerell, A. D. Jr. Induced polarization influences the fundamental forces in DNA base flipping. *J. Phys. Chem. Lett.* **5**, 2077–2083. <https://doi.org/10.1021/jz5009517> (2014).
- Dans, P. D. et al. How accurate are accurate force-fields for b-DNA?. *Nucleic Acids Res.* **45**, 4217–4230. <https://doi.org/10.1093/nar/gkw1355> (2017).
- Guéron, M., Kochoyan, M. & Leroy, J.-L. A single mode of DNA base-pair opening drives imino proton exchange. *Nature* **328**, 89–92. <https://doi.org/10.1038/328089a0> (1987).

36. Huang, N., Banavali, N. K. & MacKerell, A. D. Jr. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci.* **100**, 68–73. <https://doi.org/10.1073/pnas.0135427100> (2003).
37. Law, S. M. & Feig, M. Base-flipping mechanism in postmismatch recognition by muts. *Biophys. J.* **101**, 2223–2231. <https://doi.org/10.1016/j.bpj.2011.09.045> (2011).
38. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **1**, 826–843. <https://doi.org/10.1002/wcms.31> (2011).
39. Zheng, G., Lu, X.-J. & Olson, W. K. Web 3DNA - a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **37**, W240–W246. <https://doi.org/10.1093/nar/gkp358> (2009).
40. Banavali, N. K. & MacKerell, A. D. Jr. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **319**, 141–160. [https://doi.org/10.1016/S0022-2836\(02\)00194-8](https://doi.org/10.1016/S0022-2836(02)00194-8) (2002).
41. Ray, D. & Andricioaei, I. Free energy landscape and conformational kinetics of Hoogsteen base pairing in DNA vs. RNA. *Biophys. J.* **119**, 1568–1579. <https://doi.org/10.1016/j.bpj.2020.08.031> (2020).
42. Lefebvre, C., Klein, J., Khartabil, H., Boisson, J.-C. & Hénon, E. Igmpilot: A program to identify, characterize, and quantify molecular interactions. *J. Comput. Chem.* **44**, 1750–1766. <https://doi.org/10.1002/jcc.27123> (2023).
43. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929. <https://doi.org/10.1093/nar/gkp608> (2009).
44. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8 (2015).
45. Chen, E. A. & Porter, L. L. SSDraw: software for generating comparative protein secondary structure diagrams. *Protein Sci.* **32**, e4836. <https://doi.org/10.1002/pro.4836> (2023).
46. Vander Meersche, Y., Cretin, G., de Brevern, A. G., Gelly, J.-C. & Galochkina, T. MEDUSA: prediction of protein flexibility from sequence. *J. Mol. Biol.* **433**, 166882. <https://doi.org/10.1016/j.jmb.2021.166882> (2021).
47. van Dijk, M., van Dijk, A. D. J., Hsu, V., Boelens, R. & Bonvin, A. M. J. J. Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.* **34**, 3317–3325. <https://doi.org/10.1093/nar/gkl412> (2006).
48. van Dijk, M. & Bonvin, A. M. J. J. A protein-DNA docking benchmark. *Nucleic Acids Res.* **36**, e88–e88. <https://doi.org/10.1093/nar/gkn386> (2008).
49. Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373. <https://doi.org/10.1093/nar/gkx407> (2017).
50. Christoffer, C. & Kihara, D. Domain-based protein docking with extremely large conformational changes. *J. Mol. Biol.* **434**, 167820. <https://doi.org/10.1016/j.jmb.2022.167820> (2022).
51. Rodríguez-Lumbreras, L. A., Jiménez-García, B., Giménez-Santamarina, S. & Fernández-Recio, J. pyDockDNA: A new web server for energy-based protein-DNA docking and scoring. *Front. Mol. Biosci.* **9**, <https://doi.org/10.3389/fmolb.2022.988996> (2022).
52. Christoffer, C. & Kihara, D. Modeling protein-nucleic acid complexes with extremely large conformational changes using flexlerd. *Proteomics* **23**, 2200322. <https://doi.org/10.1002/pmic.202200322> (2023).
53. Ye, Y. et al. Enzymatic excision of uracil residues in nucleosomes depends on the local DNA structure and dynamics. *Biochemistry* **51**, 6028–6038. <https://doi.org/10.1021/bi3006412> (2012).
54. Bednar, J. et al. Structure and dynamics of a 197 bp nucleosome in complex with linker histone H1. *Mol. Cell* **66**, 384–397. <https://doi.org/10.1016/j.molcel.2017.04.012> (2017).
55. Li, S., Olson, W. K. & Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34. <https://doi.org/10.1093/nar/gkz394> (2019).
56. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802. <https://doi.org/10.1002/jcc.20289> (2005).
57. Huang, J. & MacKerell, A. D. Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145. <https://doi.org/10.1002/jcc.23354> (2013).
58. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **36**, 996–1007. <https://doi.org/10.1002/jcc.23899> (2015).
59. Andersen, H. C. Rattle: A velocity version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52**, 24–34. [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1) (1983).
60. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962. <https://doi.org/10.1002/jcc.540130805> (1992).
61. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) (1996).
62. van Dijk, M. & Bonvin, A. M. J. J. Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res.* **38**, 5634–5647. <https://doi.org/10.1093/nar/gkq222> (2010).
63. Huang, S.-Y. & Zou, X. MDockPP: A hierarchical approach for protein-protein docking and its application to CAPRI rounds 15–19. *Proteins* **78**, 3096–3103. <https://doi.org/10.1002/prot.22797> (2010).
64. Baek, M. et al. Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* **21**, 117–121. <https://doi.org/10.1038/s41592-023-02086-5> (2024).
65. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* **630**, 493–500. <https://doi.org/10.1038/s41586-024-07487-w> (2024).

## Acknowledgements

This work has been supported by the ANR grant “Dyprosomé” (ANR-21-CE45-0032-02). We thank the Meso-center of the University of Lille for the High Performance Computing resources and services they provided for the calculations performed during this project.

## Author contributions

R.B. conceived the study, V.S. performed the Molecular Dynamics simulations, N.R. performed the docking simulations of UDG on base-flipped structures from trinucleosome arrays and oligomeric DNA, G.B. provided advice for MD simulations and protein docking simulations; F.C. provided the discussion and MC simulations for the UDG search process on DNA described in the Supplementary Material and selected the trinucleosome linker structures from our earlier work. V.S., N.R., G.B., S.G., M.L., F.C. and R.B. analysed and discussed the results. All authors contributed to the writing of the manuscript.

## Declarations

## Competing interests:

The authors declare to have no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-91705-6>.

**Correspondence** and requests for materials should be addressed to R.B.

**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) 2025