







DNA repair glycosylase hNEIL1 triages damaged bases via competing interaction modes

Menghao Liu ^{1,2,9}, Jun Zhang ^{3,4,9}, Chenxu Zhu ^{5,9}, Xiaoxue Zhang^{1,2}, Weide Xiao⁵, Yongchang Yan^{1,2}, Lulu Liu^{1,2}, Hu Zeng ⁵, Yi Qin Gao ^{3,4,6,7}✉ & Chengqi Yi ^{1,5,8}✉

DNA glycosylases must distinguish the sparse damaged sites from the vast expanse of normal DNA bases. However, our understanding of the nature of nucleobase interrogation is still limited. Here, we show that hNEIL1 (human endonuclease VIII-like 1) captures base lesions via two competing states of interaction: an activated state that commits catalysis and base excision repair, and a quarantine state that temporarily separates and protects the flipped base via auto-inhibition. The relative dominance of the two states depends on key residues of hNEIL1 and chemical properties (e.g. aromaticity and hydrophilicity) of flipped bases. Such a DNA repair mechanism allows hNEIL1 to recognize a broad spectrum of DNA damage while keeps potential gratuitous repair in check. We further reveal the molecular basis of hNEIL1 activity regulation mediated by post-transcriptional modifications and provide an example of how exquisite structural dynamics serves for orchestrated enzyme functions.

¹Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. ²Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ³Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing, China. ⁴Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, China. ⁵State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. ⁶Beijing Advanced Innovation Center for Genomics, Peking University, Beijing, China. ⁷Biomedical Pioneering Innovation Center, Peking University, Beijing, China. ⁸Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing, China. ⁹These authors contributed equally: Menghao Liu, Jun Zhang, Chenxu Zhu. ✉email: gaoyq@pku.edu.cn; chengqi.yi@pku.edu.cn

The genome integrity of living organisms is constantly subject to a wide range of threats, such as metabolic reactive oxygen species (ROS) and external ionizing irradiations, which inevitably result in various alternations to the chemical structures of canonical nucleobases¹. These damaged bases are mutagenic and/or cytotoxic to cells and their accumulation can eventually lead to severe diseases including cancer, aging, and neurodegenerative disorders^{2,3}. Thus, cells have evolved the base excision repair (BER) pathway to correct the majority of abnormal bases⁴. BER is initiated by lesion-specific DNA glycosylases, which catalyze the cleavage of cognate base substrates and mediate the recruitment of downstream repair machinery to the damage sites^{5–7}.

Human endonuclease VIII-like 1 (hNEIL1) is a bifunctional DNA glycosylase of the Fpg/Nei family, which catalyzes the removal of damaged bases and subsequent incision of the newly generated abasic sites^{8–10}. Unlike traditional enzymes which are commonly referred to as highly specific biocatalysts, one special feature for hNEIL1 is its unique capability to recognize a wide range of substrates, including formamidopyrimidines, hydantoin lesions, oxidized, and dihydro-modified pyrimidine bases^{11–13}, derived from all four canonical bases (Supplementary Fig. 1). Interestingly, there is an editing event occurring on hNEIL1 pre-mRNA which leads to a K to R recoding at residue 242, and these two forms of hNEIL1 enzyme (K242 and R242 for the unedited and edited form, respectively) are shown to have different lesion specificity and glycosylase activity¹⁴. Because hNEIL1 is increased during the S phase of the cell cycle and can remove base lesions in ssDNA, bubble, and forked structures^{15,16}, it has been proposed that hNEIL1 is involved in the process of replication-associated DNA repair^{17–19}. Hence, hNEIL1 is an essential enzyme for guarding the genome stability in mammalian cells. In line with this, hNEIL1 abnormality has been related to metabolic syndrome, immunity disorders, and carcinogenesis^{20–24}. In addition, emerging kinds of literature have evidenced that hNEIL1 may have a regulatory role in the active DNA demethylation pathway^{25–27}.

Many DNA lesions differ from normal DNA by only one or a few atoms; thus, repair enzymes face the daunting task of distinguishing their cognate substrates from the vast excess of normal bases²⁸. While there are specific glycosylases like uracil-DNA glycosylase (UNG)^{29,30}, 8-oxoguanine DNA glycosylase 1 (hOGG1)³¹, and alkylpurine glycosylase D (AlkD)³², DNA glycosylases (for instance, hNEIL1, AlkA [*E. coli* 3-methyladenine DNA glycosylases]³³, and AAG [human 3-methyladenine DNA glycosylases]³⁴) can also have remarkably broad substrate ranges²⁸. Such broad specificity of repair is believed to be at odds with exclusion of normal DNA. In fact, there is evidence that repair enzymes are not infallible: several NER and BER enzymes have been shown to act on normal DNA^{35–38}. Because the repair substrate is DNA, the carrier of genetic information, even low levels of mistaken activity could jeopardize the integrity of the genome. However, our understanding of the molecular basis of how broad-specificity repair enzymes, including hNEIL1, minimize such risks, i.e. recognize various substrates while suppressing erroneous cleavage of normal bases, is very limited.

Here, by integrating structural, computational, and biochemical investigations, we show that hNEIL1 uses a triage mechanism to achieve broad specificity and reduce futile repair of normal bases. This mechanism involves two mutually competing interaction states: an activated state that commits catalysis and BER, and a quarantine state that temporarily separates and prevents the flipped base from catalysis via auto-inhibition. The dominant interaction state varies for different types of damage and is largely determined by the chemical and structural properties of the flipped bases. Such a mechanism enables hNEIL1 to process a

series of structurally diverse substrates while minimizes the risk of erroneous cleavage of normal bases. The equilibrium between the two states can also be modulated by a naturally existing hNEIL1 protein variant or artificially introduced mutations. Furthermore, based on our understanding of such a mechanism, we can reactivate the catalytically incompetent hNEIL1 mutants via rational design.

Results

hNEIL1 uses a tautomerization-dependent mechanism to recognize various substrates. We previously introduced a P2G mutation in the active site of hNEIL1(CA95) truncation to reduce the catalytic activity, which allowed us to obtain the Tg-containing complex structure³⁹. While we also crystallized the complex structures for more lesions in this study, we were unable to obtain intact substrates in the active site of hNEIL1, presumably due to the residual glycosylase activity. Hence, we introduced an extra E3Q mutation to the truncated hNEIL1 (CA95 P2G) (see the “Methods” section) and succeeded in the determination of complex structures containing 5-hydroxyuracil (5-OHU), spiroiminodihydantoin 1 (Sp1), (S)-guanidinohydantoin (Gh), dihydrothymine (DHT) and dihydrouracil (DHU), respectively (Fig. 1a, Supplementary Fig. 2 and Supplementary Table 1).

Despite the very different chemical structures of these substrates, they are recognized similarly by hNEIL1. The damaged bases are fully extruded from the DNA helix and inserted into a relatively loose binding pocket, which allows the accommodation of substrates with varying sizes and shapes. One edge of the flipped base points towards the protein and forms close contact with the side chain of K242 (Fig. 1b). For a naturally existing hNEIL1 variant (R242)¹⁴ in complex with dsDNAs containing OHU, Sp1, or Gh, the damaged bases are in close contact with the guanidine group of R242 (Fig. 1c). Such conformation of the lesion recognition loop resembles that in the hNEIL1-Tg structure (Fig. 1d), where the positively charged side chain of residue 242 is contiguous to the solvent-exposed flipped base while the aromatic residue Y244 is laid away from the active site (Supplementary Fig. 3a) (thus called “242-in” conformation).

Noteworthy, across all these structures, we observed an unusually short distance (~3.0 Å) between an outwards heterocyclic nitrogen atom of the nucleobases and the terminal nitrogen atom of residue 242. While repulsion is expected for the two groups if both are fully protonated (i.e., an –NH group in the nucleobase and a positively charged side chain of residue 242), we previously showed in the hNEIL1-Tg complex that hNEIL1 utilizes a tautomerization-dependent mechanism to recognize the flipped base³⁹. In this mechanism, the base undergoes a keto-enol tautomerization event where the originally protonated heterocyclic nitrogen atom becomes a hydrogen bond acceptor and hence can be hydrogen-bonded to residue 242 (Fig. 1e). Additionally, this tautomerization-dependent hydrogen bond is crucial for the subsequent catalysis by activating the substrate and stabilizing the transition state³⁹. Based on the similar overall conformation and detailed atomic interactions, we conclude that such a tautomerization-dependent mechanism may be a general strategy for hNEIL1 to recognize not only Tg but also 5-OHU, Sp1, Gh, DHT, and DHU. Hence, tautomerization can serve as a chemical check to interrogate the tendency to tautomerize for a flipped nucleobase, but not its exact structure.

A cryptic nucleobase-binding state in hNEIL1. Surprisingly, in the crystal structures of the edited hNEIL1(R242) bound to DHT or DHU, we observed that hNEIL1 can also interact with the substrates in a very different mode (Supplementary Fig. 3b, c).

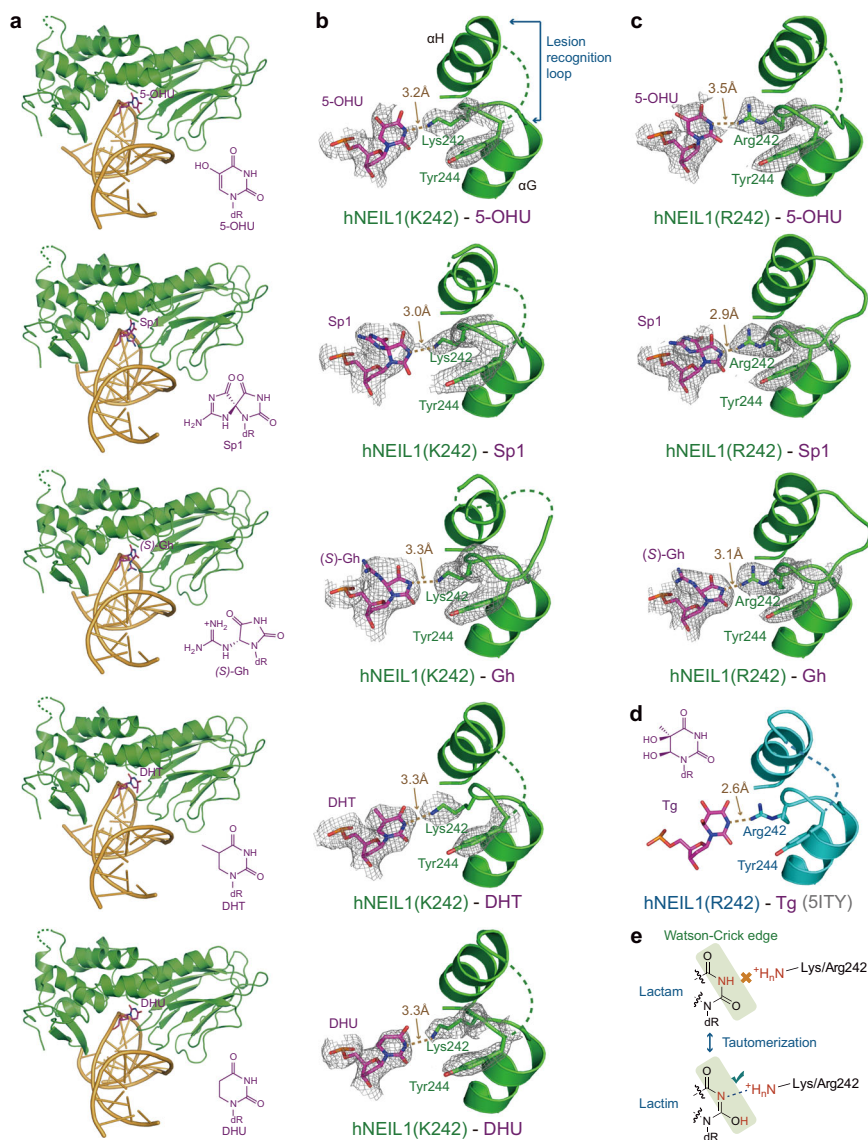


Fig. 1 Tautomerization-dependent recognition of diverse substrates by hNEIL1. **a** Overall views of hNEIL1(K242) bound to double-strand DNAs containing a 5-OHU, Sp1, (S)-Gh, DHT, or DHU. The chemical structures of individual damaged bases are shown in purple. **b** Zoom-in views of the interaction mode for hNEIL1(K242) in complex with various damaged bases. The lesion recognition loop located between α G and α H is indicated. The 2Fo – Fc simulated-annealing composite omit electron density map is shown in gray mesh and contoured at 1.0 σ for both damaged bases and key loop residues. The gold dashed line represents hydrogen bond interaction. The dashed curve stands for an unbuildable short region of the lesion recognition loop due to ambiguous electron density. **c** The 242-in loop conformation was again observed in the crystal structure of a naturally existing variant hNEIL1(R242) in complex with 5-OHU, Sp1, and (S)-Gh, respectively. **d** The previously reported hNEIL1(R242)-Tg structure (PDB code: 5ITY) also shows a 242-in conformation, consistent with the all above described hNEIL1-DNA complexes. **e** The tautomerization-dependent hydrogen bonding between the flipped bases and residue 242 of hNEIL1. The base tautomers are denoted with the lactam (upper) and lactim (lower) forms.

Different from the 242-in loop conformation described previously, in the structures of hNEIL1(R242)-DHU and hNEIL1(R242)-DHT, R242 residue moves away from the flipped bases while Y244 moves in and stacks against DHU/DHT, burying the nucleobase in a hydrophobic enclosing (Fig. 2a and Supplementary Fig. 3d, e) (termed as the “244-in” conformation). Such a conformational change is reminiscent of the “DFG flip” well-known for protein kinases, in which key amino acid residues in a flexible loop can switch their relative positions^{40,41}. The 244-in conformation of lesion recognition loop is also different from that in the hNEIL1 free protein (*apo* conformation), where the loop rotates for $\sim 40^\circ$ (using Gly240 and Gly249 as turning points) to stay far away (>9 Å) from the position of the flipped base

(Fig. 2a). To rule out the possibility that the observed interaction is caused by mutations introduced to hNEIL1 for crystallization experiments, we used an orthogonal approach in which a modified nucleobase substrate is bound by the native protein containing the wild type active site. We synthesized a 2'-fluoro-substituted stable mimic of DHU (FDHU), which allows binding but not catalysis by hNEIL1 (Supplementary Figs. 4 and 5). Indeed, in the hNEIL1(R242)-FDHU structure, residues R242 and Y244 again swaps positions (Fig. 2b, c), recapitulating the observed 244-in conformation in the hNEIL1(R242)-DHU and hNEIL1(R242)-DHT structures (Supplementary Fig. 3b, c).

Unlike the 242-in conformation which corresponds to an activated state priming the glycosylase catalysis reaction³⁹, what

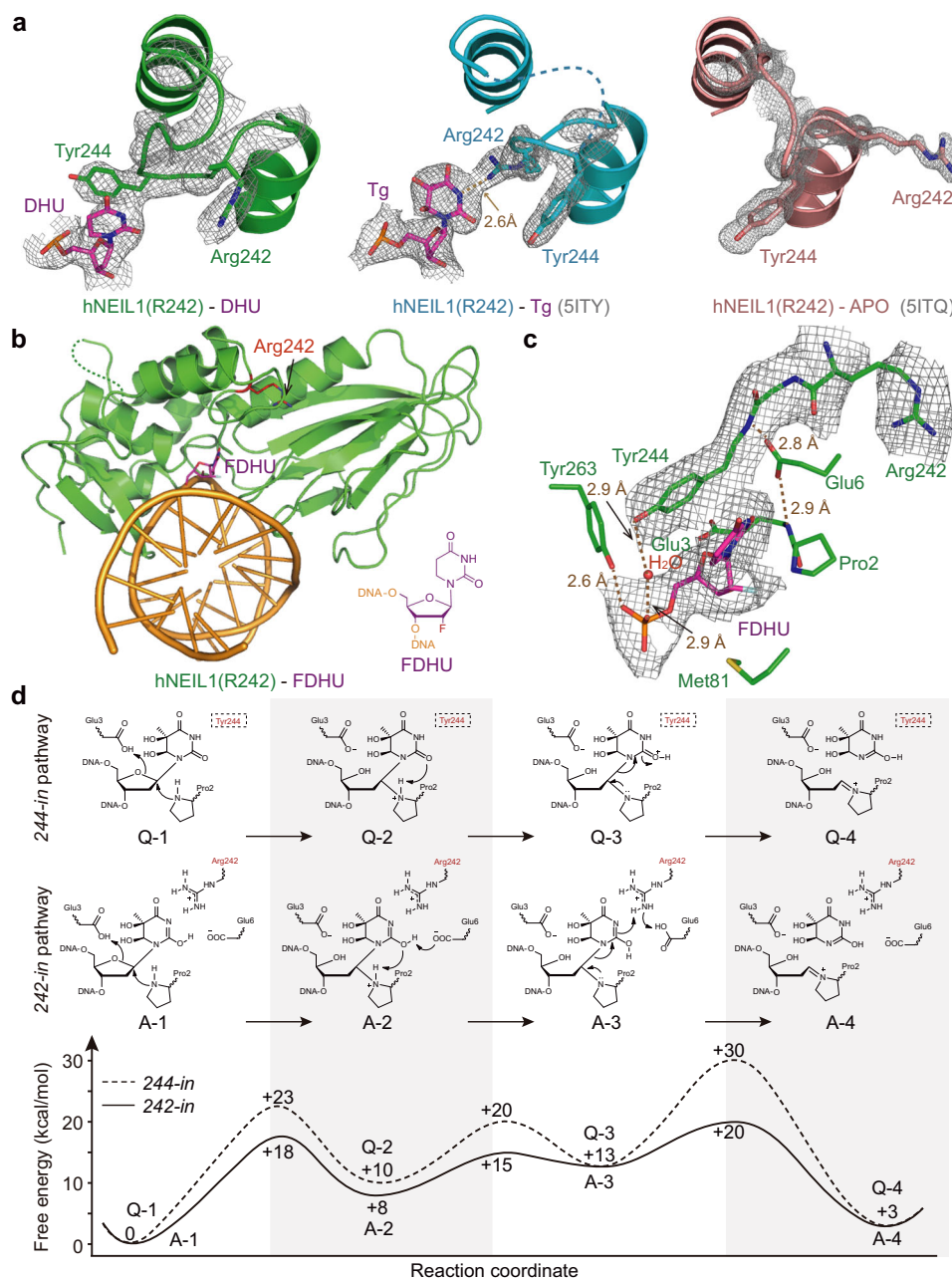


Fig. 2 Characterization of a newfound hNEIL1-DNA interaction mode. **a** Crystal structure of hNEIL1(R242) bound to DHU (left) reveals a distinct conformation of lesion recognition loop, compared with the hNEIL1(R242)-Tg structure (middle) or the hNEIL1(R242)-APO structure (right). In the DHU-bound structure, Arg242 occupies the same position which is originally taken by Tyr244 of the Tg-containing structure, while Tyr244 stacks against the flipped DHU base. In addition, Arg242 of the APO structure is placed far away from the damaged bases and fully exposed to the solvent. The gray 2F_o-F_c simulated-annealing composite omit electron density map is contoured at 1.0 σ for both damaged bases and lesion recognition loop. **b** Overall structure of hNEIL1(R242) bound to dsDNA-containing FDHU. The chemical structure of FDHU deoxynucleotide is shown in purple and residue 242 is colored in red. **c** Zoom-in view of the interactions between hNEIL1(R242) and FDHU. The brown dashed lines represent hydrogen bond interactions and red ball stands for water molecules. **d** QM/MM calculations reveal that the free-energy profile of base cleavage reaction pathway for 244-in conformation displays an extremely high energy barrier, comparing to a more reasonable 242-in conformation.

role the 244-in conformation plays in the function of hNEIL1 remains elusive. Apparently, the hydrogen bond interaction between residue 242 of hNEIL1 and the substrate which is present in the 242-in conformation is absent in the 244-in structures, indicating the loss of a key stabilizing factor of binding and catalysis. Thus, we hypothesized that 244-in conformation is not suitable for executing the glycosylase chemistry. To verify this hypothesis, we performed QM/MM simulations of the ribose-activated catalysis pathway^{39,42} initialized from the hNEIL1

(R242)-DHU structure, where the reaction is commonly assumed to be triggered by a nucleophilic attack to C1' of the ribose by Pro2, followed by the breakage of the glycosidic bond undergoing a highly polarized intermediate (Supplementary Fig. 6). Compared to the 242-in conformation where a delicate hydrogen bond network is formed in the active site³⁹ (Fig. 2d), there lacks catalytically helpful interactions between the substrate and enzyme in the 244-in conformation, causing several drawbacks hindering the catalysis: (i) the position of the flipped base and

ribose becomes more flexible and C1' can stay farther away from Pro2 (Supplementary Fig. 7), hence the reaction is more difficult to be triggered; (ii) the “proton relay” which facilitates the catalysis in *242-in* conformation is no longer applicable; (iii) the produced highly polarized intermediate cannot be stabilized by a hydrophobic Y244. All these factors contribute to the observed high barrier along the reaction free-energy profile (Fig. 2d and Supplementary Fig. 8). Compared to the *242-in* reaction profile which is in line with other ribose-protonated DNA repair studies^{28,42}, the *244-in* pathway could be slowed down by a significantly higher barrier at the glycosylic bond-breaking step (i.e., “Step III” in Fig. 2d) by several orders of magnitude. Thus, the *244-in* conformation is incompetent of glycosylase activity and hence represents a distinct interaction mode with potentially special function.

A proposed nucleobase recognition model involving two competitive interaction states. One conformation was captured for one particular combination of hNEIL1 protein and base lesion during our crystallographic experiments; and in total, we observed three different conformations for the GS-rich, flexible lesion recognition loop. Whether or not these different conformations could all be adopted by hNEIL1 to recognize a particular substrate is unknown. To fully explore possible metastable protein conformations (as observed in crystals) during hNEIL1–DNA recognition, we performed adaptive MD simulations to explore possible metastable loop conformations in hNEIL1–DNA complexes. The adaptive MD helped explore the loop conformations by interpolating between different crystal structures. We can then identify the metastable conformations by relaxing the perturbed structures generated during the adaptive MD (Supplementary Fig. 9 and Supplementary Table 2). Specifically, inspecting the obtained simulation samples with respect to the relative positions between R242/Y244 and the flipped base, we identified three conformational zones with relatively high population and stability⁴³, indicating the existence of (at least) three metastable conformations of the lesion recognition loop upon interacting with DNA (Fig. 3a). Intriguingly, we found that the loop conformations in these three local minima (A, B, and C, respectively) coincide with those observed in three types (i.e., *apo*, *242-in*, and *244-in*, respectively) of crystal structures of hNEIL1 (Fig. 3b). Specifically, in local minimum A, the loop conformation resembles the *apo* form which is experimentally observed for the free protein. Local minimum B consists of *242-in* loop conformation, very similar to the “activated state” defined above, except that the flipped base is not tautomerized. Local minimum C corresponds to the *244-in* conformation that was found in the crystal structure of hNEIL1 bound to DHU/FDHU and DHT. Therefore, MD simulations indicate the co-existence of multiple states for hNEIL1 to interact with a particular nucleobase, among which only the relatively stable one is captured in the crystal structure.

Thus, both our structural and computational observations lead to a substrate recognition model involving multiple interaction states (Fig. 3c and Supplementary Fig. 10). First, the free protein favors a state which has the *apo* loop conformation; hNEIL1 may encounter the DNA in this form (“encounter state”). Once bound to DNA, the recognition loop is shifted away from the *apo* conformation, with the loop moving towards the extruded nucleobase. Along this trajectory, the lesion recognition loop can adjust to a *242-in* conformation, where hNEIL1 exploits a tautomerization-dependent interrogation mechanism attempting to establish hydrogen bond(s) with the flipped base. The activated state arrives if the chemical check is passed, i.e., the key hydrogen bond is established, and the substrate is activated and permitted

to enter the catalytic cycle for excision. Alternatively, the lesion recognition loop may also bypass the activated state and transit to the *244-in* conformation. In contrast to the *242-in* conformation, the *244-in* conformation is catalytically inactive and functionally protects the nucleobase from cleavage, in competition with the activated state. The relative dominance of the two interaction states depends on their thermodynamic stability, which could be quantified by their free energy difference (Fig. 3d).

According to the Michaelis–Menten theory, the enzyme–substrate complex (denoted by “ES”) reaches a binding equilibrium prior to the rate-determining catalysis step, so the overall reaction rate r can be described by the product of the catalysis rate constant and the total concentration of the equilibrium ES complex, [ES]. However, in hNEIL1, the existence of a competing catalysis-inactive state leads to an auto-inhibitory mechanism, that is, only a fraction of the ES complex in the activated state [ES]_A can go through the subsequent catalysis. Given that [ES]_A can be approximated using the free energy difference ΔG between the two competing states, we arrived at the following catalysis rate equation:

$$r_{\text{hNEIL1}} = k_{\text{cat}}[\text{ES}]_{\text{A}} \approx \frac{k_{\text{cat}}}{1 + \exp\left(\frac{\Delta G}{RT}\right)}[\text{ES}] = \tilde{k}_{\text{cat}}[\text{ES}] \quad (1)$$

$$\Delta G = \Delta G_{\text{conf}} + \Delta G_{\text{chem}} \quad (2)$$

where ΔG determines the equilibrium population of the protein–DNA complex in the activated state, and k_{cat} is the rate for the glycosylase reaction; ΔG_{conf} accounts for the conformational transition of the recognition loop from *244-in* to *242-in* (Supplementary Fig. 9), while ΔG_{chem} corresponds to the tautomerization event (Fig. 3d). Note that the equilibrium between the two competing states will influence the overall reaction rate by changing the apparent rate constant \tilde{k}_{cat} . Consequently, hNEIL1 can achieve additional substrate discrimination through the thermodynamic competition between the two competing states, which bears significant impact on the overall rate of the catalysis. If the catalytically inactive state dominates over the activated state (ΔG is positive), a smaller initial rate for the catalysis is expected. Therefore, we refer the catalytically inactive state as a “quarantine state”.

Rational manipulation of hNEIL1. According to the above model, we may swap the preferred interaction states between hNEIL1 and particular substrates, if we are able to inverse the relative dominance of the activated and quarantine states. In light of the observation that different hNEIL1–DNA interaction modes are largely resulted from the conformation changes of the lesion recognition loop, we chose to modify the potential key amino acids on this loop (Fig. 4a). In order to destabilize the quarantine state, we rationally designed hNEIL1(R242, G249P) and hNEIL1(R242, Y244H) to disrupt the flexibility of the loop or the aromatic stacking interactions respectively. Instead of the quarantine state originally observed in hNEIL1(R242)–FDHU structure, we found that both two mutants switched to the activated state as expected (Supplementary Fig. 11a–e).

We then attempted to engineer the enzyme activity via manipulation of the activated state. According to our model, the hydrogen bond(s) between residue 242 and the flipped base is crucial for recognition and catalysis, and disruption of this interaction would destabilize the activated state. As expected, hNEIL1(R242A) or hNEIL1(R242H) demonstrate significantly reduced biochemical activity (Fig. 4b and Supplementary Fig. 11f). As a further test of the proposed mechanism, we also sought to “rescue” the activated state by mutating the lesion-recognition loop. Because a nearby positively charged amino acid side chain is

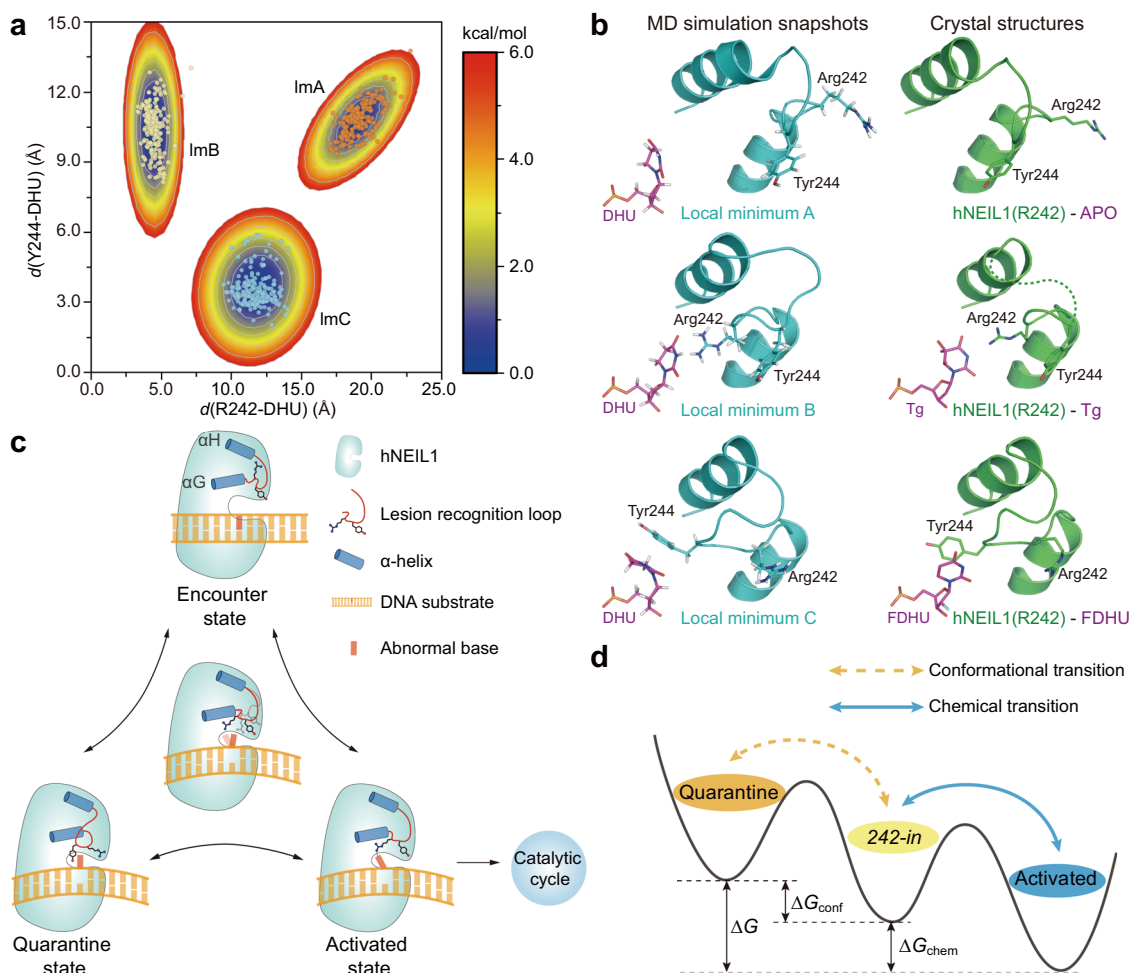
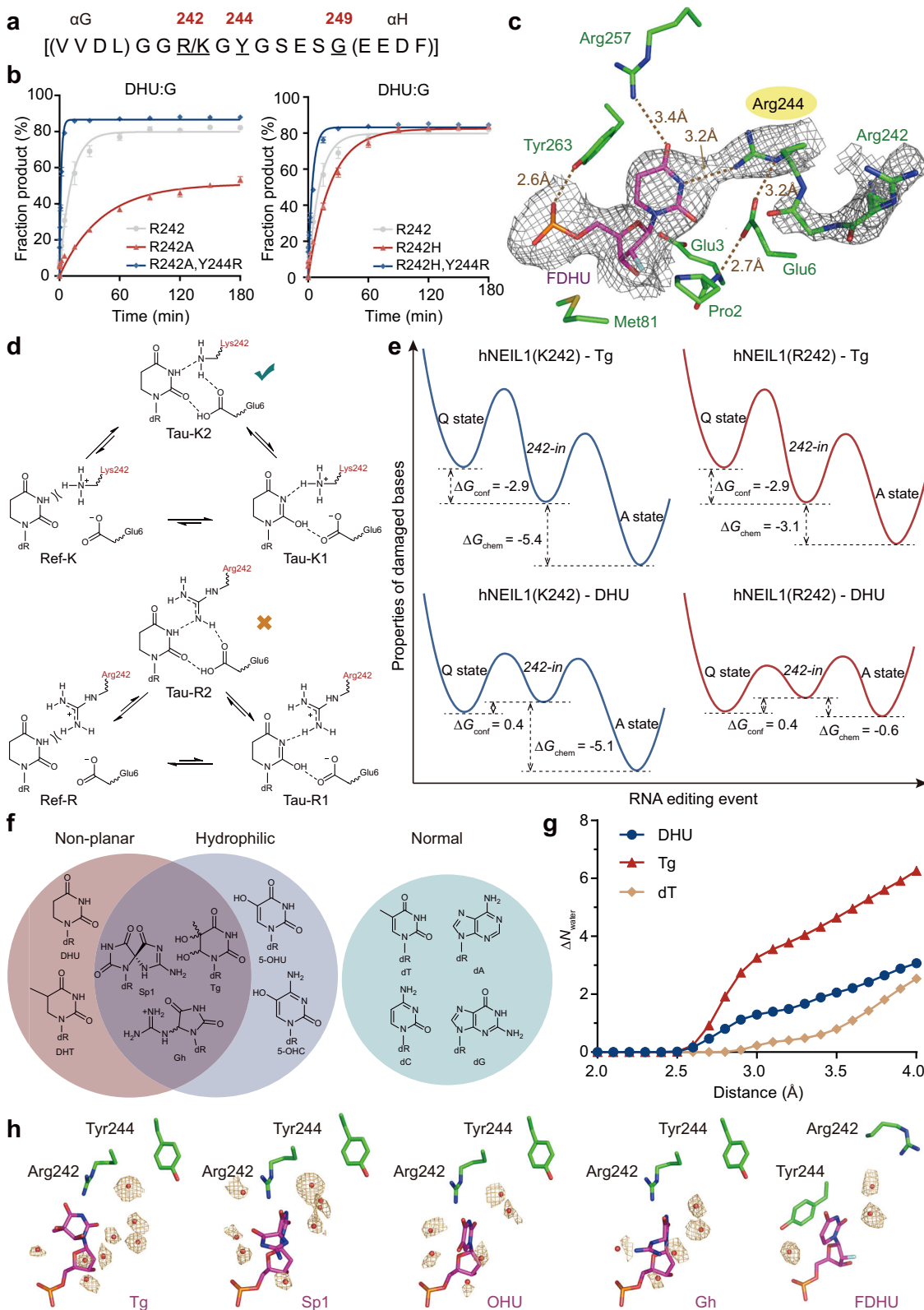


Fig. 3 A proposed nucleobase-recognition model of hNEIL1 by computational and structural evidences. **a** Adaptive MD simulations reveal the existence of three conformational states of hNEIL1 when interacting with the substrate DNA (DHU) with relatively enriched population. ImA local minimum A, ImB local minimum B, ImC local minimum C. The dots superimposed are randomly selected MD configurations from each metastable conformation and colored accordingly. **b** The loop conformations of snapshots from adaptive molecular dynamics simulations are in good coincidence with those observed in the crystal structures. **c** Substrate-recognition model of hNEIL1. Upon encountering the DNA substrates, the lesion recognition loop of hNEIL1 shifts towards the flipped bases from the originally preferred *apo* conformation (Encounter state). Once the tautomerization-dependent hydrogen bond interaction between damaged base and residue 242 is established, the substrate would be activated (Activated state) and permitted to enter the catalysis cycle. Alternatively, if the base does not pass the chemical check, the loop would transit to a 244-*in* conformation (Quarantine state), temporally restricting the nucleobase from catalysis via competing with the activated state. **d** The relative dominance of the quarantine state and activated state depends on the tug of war between their thermodynamic stability. ΔG denotes free energy difference quantifying the relative stability of the activated state over the quarantine state. ΔG can be decomposed into two terms, ΔG_{conf} (Supplementary Fig. 9) and ΔG_{chem} , respectively. The gold dashed line represents a process involving conformational transition and the blue solid line means a chemical transition.

important for catalysis, we attempted to introduce a charged residue in close proximity to the flipped base. Considering that Y244 is able to closely contact the base in the quarantine state, we envisioned that replacing Y244 with a positively charged residue might re-establish an active-state-like interaction between the mutated residue 244 and the flipped base, thus possibly rescuing the catalysis activity. To our delight, introduction of Y244R to hNEIL1(R242A) or hNEIL1(R242H) mutant resulted in both mutants to restore their glycosylase activity; in fact, we even found comparable or higher activity for hNEIL1(R242A, Y244R) and hNEIL1(R242H, Y244R) than the wild-type hNEIL1 for DHU and Tg (Fig. 4b, Supplementary Fig. 11f and Supplementary Table 3). We further solved the crystal structure of hNEIL1 (Y244R) bound to FDHU (Fig. 4c), and found that although the loop remains at the 244-*in* conformation, Arg244 forms a key hydrogen bond to the flipped base, recapitulating the key interaction in the activated state rather than the quarantine state.

On top of the artificially designed mutations above, the naturally existing hNEIL1 variant, caused by an RNA-editing event at position 242¹⁴, provides a “native” example of perturbation to the enzyme–substrate interaction. Particularly, the chemical property of 242 side chain can directly influence the “chemical check” (Fig. 4d). By undergoing a base–acid equilibrium with the adjacent Glu6, K242 can interact with the flipped base as a hydrogen bond acceptor (denoted as “Tau-K2”), forgoing the pre-requisite of the base-tautomerization (“Tau-K1”). Through QM/MM calculations we found that, due to the moderate basicity of K242 (pKa~10)⁴⁴, “Tau-K2” is indeed a stable configuration of the activated state in addition to the base-tautomerized “Tau-K1”. In contrast, R242 can hardly utilize this alternative chemistry because the guanidine group exhibits strong basicity^{44,45}, and no stable configuration of the speculative “Tau-R2” form can be obtained in simulations (Fig. 4d). Accordingly, K242 is more likely to retain the base in the activated state compared to R242, or in



other words, the activated state is more favored by the genomically encoded hNEIL1(K242) compared to its RNA-edited counterpart, which is supported by our QM/MM calculations (Fig. 4e and Supplementary Table 4). Thus, although K and R are commonly regarded as similar basic residues, our findings reveal that RNA editing on hNEIL1 can lead to different biochemical outcomes. Indeed, these two forms of hNEIL1 have been shown to have

different glycosylase activities where the unedited hNEIL1(K242) is faster than edited hNEIL1(R242) on several substrates^{14,39}.

Molecular basis of substrate recognition. Are the amino acid residues on the key lesion recognition loop fully responsible for the stability of the enzyme–substrate interaction? The fact that different recognition states were crystallographically captured for

Fig. 4 Both key loop residues and base per se contribute to the state transition. **a** The amino acid sequence of lesion recognition loop. Three key residues at position 242, 244 and 249 are indicated. **b** Biochemical assays show that hNEIL1(Y244R) is capable of rescuing the glycosylase activities of 242 mutants on DHU. Values represent mean \pm s.e.m. ($n = 3$). See Supplementary Table 3 for calculated rate constants. **c** Zoom-in view of the interactions between FDHU and hNEIL1(Y244R). The simulated-annealing 2Fo–Fc composite omit electron density map is shown in gray mesh and contoured at 1.0σ . **d** An additional stable protonation state (“Tau-K2”) can be stabilized by Lys242, but not by Arg242, to help retain the flipped base in the activated state (illustrated with DHU here). **e** The free energy profile of state transition for hNEIL1(K242) and hNEIL1(R242) bound to duplex DNA containing DHU or Tg. The calculated ΔG_{conf} and ΔG_{chem} values are determined in kcal/mol. Q state denotes quarantine state and A state stands for activated state. **f** The flipped nucleobases can be categorized by their chemical properties including non-planarity and hydrophilicity. Loss of aromaticity and increase of hydrophilicity of bases would facilitate the formation of a relatively stable activated state. **g** The number of desolvated water molecules (ΔN_{water}) when the loop transits from 242-*in* to 244-*in* conformation as a function of the distance from the base computed by MD simulations. **h** The observed solvent environments in the vicinity of the flipped bases within the crystal structures. Water molecules within 5 Å distance of the damaged base are shown here. The annealed difference Fo–Fc electron density map is shown in orange mesh and contoured at 2.5σ .

the same edited hNEIL1(R242) variant bound to different base lesions suggests that the flipped nucleobase per se also contributes to the stability of the interaction states (Supplementary Table 5). hNEIL1 repairs a diverse array of base damage, whose aromaticity and hydrophilicity differ greatly. Oxidized substrates of hNEIL1 are often bulky and contain non-planar structures; this disrupted aromaticity is expected to diminish the π – π stacking with Tyr244 and hence destabilize the quarantine state (Fig. 4f and Supplementary Fig. 12). In addition, oxidation of DNA bases often increases their hydrophilicity. Transition from the activated state to the quarantine state requires de-solvation of the flipped nucleobase; therefore, the more hydrophilic the base is, the higher the energetic penalty would be. Indeed, we observed an increased number of solvated water molecules for hydrophilic bases like Tg compared to normal dT in our all-atom MD simulations (Fig. 4g). Moreover, it is expected that the bases with less solvent interactions (e.g. DHU) would favor the formation of a stable quarantine state compared with the hydrophilic bases (Sp1, 5-OHU, Gh, and Tg). Consistent with this, at a comparable resolution (around 2.42–2.60 Å) of crystal structures containing five different types of lesions, we found quite different solvation environments for the flipped bases (Fig. 4h). Thus, the structural alteration of the damaged bases, including increased hydrophilicity and decreased aromaticity, directly influences hNEIL1–DNA interaction.

We refer such non-covalent interactions (i.e., solvation and hydrophobic interactions) as “structural check”. Such a mechanism is distinct to the “chemical check” we described earlier, in which the side chain of residue 242 helps hNEIL1 interrogate the substrate via tautomerization and specific hydrogen-bonding test. Referring to the previous quantitative model (Eqs. (1) and (2)), ΔG_{conf} can be mainly attributed to the structural check, whereas ΔG_{chem} is determined by the chemical check. Both checking mechanisms exploit the chemical and physical properties of the flipped bases: bases prone to tautomerization are more likely to pass the chemical check, whereas bases with decreased aromaticity and/or increased hydrophilicity are more likely to pass the structural check. Usually, damaged DNA bases (e.g., oxidized Tg) are more hydrophilic and simultaneously more prone to tautomerization (Supplementary Table 4). Therefore, these two checking mechanisms work synergistically for hNEIL1 to recognize DNA damage.

Recognition of a normal base by hNEIL1. According to the above substrate-checking mechanism, hNEIL1 interrogates several aspects of the flipped bases, but none of them are specific to any particular substrate. Therefore, hNEIL1 is capable of recognizing and removing a diverse spectrum of DNA damage via some common properties (e.g. non-planarity, hydrophilicity, etc.). However, such a mechanism does not guarantee that an accidentally flipped normal base does not enter the activated state

and thus being spuriously removed by hNEIL1. Take a normal dT for example: on the one hand, its intact aromaticity and hydrophobicity is expected to prevent dT from passing the two checks, while on the other hand, the outcome of the competition between the interaction states is also influenced by the protein, including the key 242 residue. To examine how hNEIL1 recognizes dT, we conducted all-atom MD simulations to quantify the thermodynamic stability of the quarantine and activated states of hNEIL1(K242) and (R242) bound to dT, respectively. In the simulated hNEIL1(R242)–dT complex, we observed that the quarantine state is more stable than the activated state and ΔG for hNEIL1(R242)–dT is also significantly larger than that of hNEIL1(R242)–Tg (Fig. 5a and Supplementary Table 4), which is consistent with our expectation. We further solved the crystal structure of hNEIL1(R242) bound to dsDNA with a T:C mismatch (Fig. 5b). Indeed, the thymine base is “trapped” in the 244-*in* conformation (Fig. 5c), consistent with the quarantine state also observed for hNEIL1(R242)–DHT and hNEIL1(R242)–DHU complexes. These computational and experimental results further support our model that the quarantine state can help hNEIL1 discriminate normal bases from DNA damage.

We next performed simulations for hNEIL1(K242)–dT, which prefers damage recognition using the activated state, to examine whether K242 would differentially influence the recognition of a normal dT. We calculated the relative stability of the two competing interaction states for hNEIL1(K242)–dT. Surprisingly, we found that the two states are of similar stability, hinting that dT is more likely to undergo catalysis in the presence of hNEIL1(K242) (Fig. 5a and Supplementary Table 4). This observation from simulation was further supported by a new crystal structure, where we found that hNEIL1 adopts the 242-*in* conformation and that the thymine base is actually removed by hNEIL1(K242) *in crystallo* (Supplementary Fig. 13). Furthermore, our biochemical data demonstrate that over the course of 72 h, hNEIL1(K242) is capable of cleaving a significantly higher portion of dT than the edited form R242 (using a duplex DNA containing a mismatched T:C pair) (Fig. 5d and Supplementary Fig. 13). Taken together, these results show that the genomically encoded K242 is faster in damage repair but more prone to error, while the edited R242 is slower but safer.

Discussion

hNEIL1 represents an example of broad-specificity DNA glycosylases. A question remains on how a DNA repair enzyme efficiently recognizes so many structurally diverse substrates while remains discriminative towards the normal bases that are in vast excess in the genome. Our study reveals that instead of utilizing a specific pocket to tightly recognize a flipped base via extensive electrostatic and hydrogen bonding interactions, hNEIL1 uses two competing interaction states to achieve substrate recognition. This mechanism is further corroborated by our finding that the

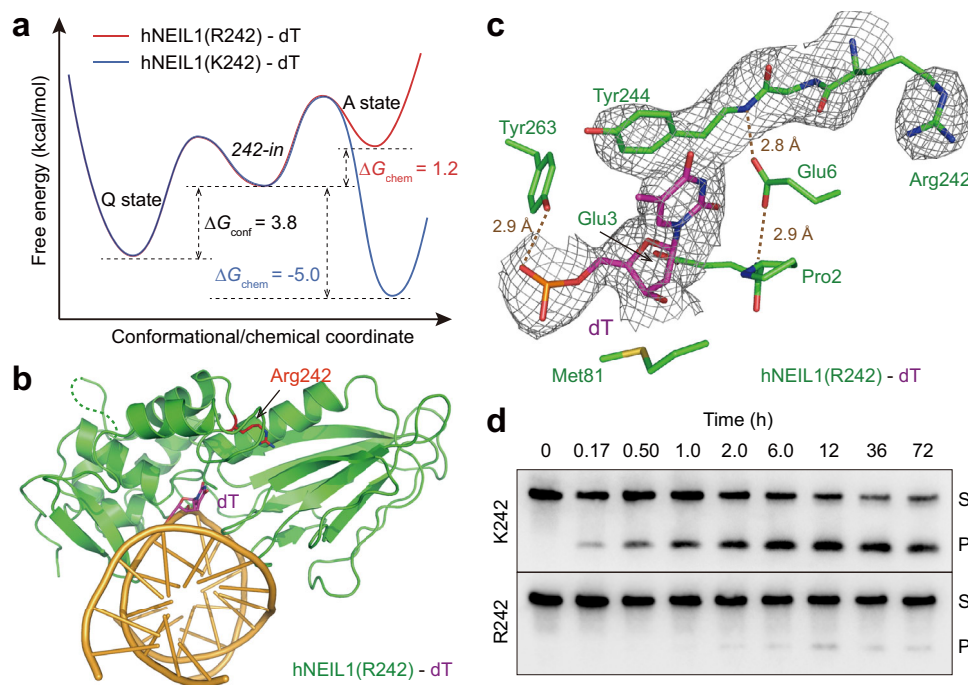


Fig. 5 The interrogation of normal base by hNEIL1. **a** The free energy profile of state transition for hNEIL1(R242) and hNEIL1(K242) bound to a mismatched dT. **b** Overall structure of hNEIL1(R242) bound to double-strand DNA containing a T:C mismatch. The flipped dT is colored with purple and Arg242 is shown in red. **c** Zoom-in view of the interactions between dT and hNEIL1(R242). A quarantine state similar to that identified in the crystal structures of hNEIL1(R242) in complex with (F)DHU or DHT is also observed here. The 2Fo–Fc simulated-annealing composite omit electron density map is shown in gray mesh and contoured at 1.0σ . **d** An electrophoretic assay shows the glycosylase activities of unedited hNEIL1(K242) and edited hNEIL1 (R242) on a T:C mismatch. The experiment was repeated three times independently with similar results. Source data are provided as a Source Data file. Detailed quantitative results can be found in Supplementary Fig. 13.

relative dominance of the competing interaction states can be manipulated by artificially introduced or naturally existing mutations to the key lesion recognition loop. For instance, based on our mechanistic understanding, we rationally introduced an Arginine mutation to the 244 position of the catalytically incompetent hNEIL1(R242H) and hNEIL1(R242A) mutants, thereby creating new gain-of-function mutants. Thus, such investigation also provides an example that a detailed understanding of reaction mechanisms could lead to the rational engineering of enzymes for tailored biological activities and potentially new-to-nature reactions.

Our crystal structures capture two interchangeable conformations for the lesion recognition loop, where residues 242 and 244 swap their relative positions and switch on and off their catalytic activity. Such conformational flexibility is reminiscent of the well-known “DFG flip”, which connects the catalytically active and inactive conformations of protein kinases^{40,41,46}. Based on the DFG flip mechanism, the blockbuster cancer drug imatinib, targeting the “DFG-out” conformation of tyrosine kinases, has been designed for cancer treatment. Given that hNEIL1 abnormality (e.g., single nucleotide polymorphisms) has been identified in metabolic syndrome and several types of cancers^{20,22,47,48}, it would be interesting to design small molecules to selectively target different states of hNEIL1–DNA interaction.

Broadly specific glycosylases are thought to be biologically and even evolutionarily beneficial for cells because DNA damage results in a great structural diversity of lesions. Concomitantly, broad specificity enzymes have been shown to act on undamaged, normal DNA, a phenomenon termed as “gratuitous repair”²⁸. It is considered to be energetically wasteful and results in increased

possibilities of spontaneous mutagenesis²⁸. However, accumulated evidence suggests that gratuitous repair is an inevitable price that broad specificity enzymes must pay^{35–38}. These findings have led to the notion that DNA repair machineries themselves may be a major source of spontaneous mutagenesis. While the literature has documented such phenomena for a long time, our understanding of how cells minimize gratuitous repair lags behind. For instance, AlkA is known to possess a remarkably indiscriminate active site; hence, once a nucleobase is flipped into the active site, the repair is solely dictated by the reactivity of the N-glycosidic bond of different substrates³⁵. In contrast to this catalysis-dependent mechanism, hNEIL1 utilizes a triage mechanism, taking place prior to catalysis, to not only recognize various substrates but also keep potential gratuitous repair activity in check. It remains to be seen to what extent hNEIL1 possesses futile repair activity *in vivo*. In addition, such pre-catalysis checking mechanism also enabled cells to regulate the repair function of hNEIL1, as demonstrated by the RNA editing event that recodes the key residue at position 242: unedited hNEIL1 (K242) is faster but riskier in damage repair, while edited hNEIL1 (R242) is slower but safer. In fact, the RNA editing level at this position is dynamically regulated among different tissues (Supplementary Fig. 13d). Therefore, it remains to be seen whether or not this post-transcriptional modification could be exploited by cells as an adaptive response of DNA repair to DNA damage. For instance, in multiple myeloma, only the unedited hNEIL1 can sufficiently relieve the oxidative burden to the genome under elevated ROS conditions⁴⁹; however, this riskier form of hNEIL1 also exposes the cells to potential threats, as a slower growth phenotype has been observed upon over-expression of hNEIL1 (K242)^{49,50}. Alternatively, it is also conceivable that such a

regulatory mechanism could modulate gratuitous repair and hence spontaneous mutagenesis in an evolution-centric view. Collectively, the triage mechanism proposed for a DNA repair glycosylase in this study provides an example to rethink about the specificity of enzymes in maintaining cellular functions.

Methods

Oligonucleotides synthesis and purification. 5-OHU, DHT, DHU, Tg, and 8-oxo-dG phosphoramidites were purchased from Glen Research. The FDHU phosphoramidite was obtained by chemical synthesis. Damage-containing oligonucleotides were synthesized on an Expedite 8909 DNA synthesizer using standard reagents (Glen Research, Inc.) and deprotected with recommended methods. The resultant oligo DNAs were sequentially purified with GlenPak DNA cartridge and denature PAGE, and desalted on Sep-Pak Cartridge C18 columns (Waters). Complementary strands were purchased from Sangon Biotech (Shanghai). The Gh- and Sp1-containing oligo DNAs were obtained by oxidation of oligomers containing 8-oxoguanine base. All the oligonucleotides were validated by MALDI-TOF-MS. See Supplementary Information for detailed experiment procedures of oligonucleotides preparation and FDHU phosphoramidite synthesis.

Expression and purification of hNEIL1 proteins. Codon-optimized full-length human *neil1* gene CDS was synthesized by GENEWIZ Inc. and cloned into pET30a (+) with C-terminal 6× His-tag. Truncated hNEIL1 protein was constructed based on this plasmid and point mutations were achieved by Fast Mutagenesis System (TransGen Biotech) (Supplementary Table 6). The plasmid was transformed into BL21 (DE3) (TransGen Biotech) competent cells for protein overexpression. Bacterial cells were grown in LB medium with 50 µg/ml kanamycin monosulfate (AMERSCO) at 37 °C until OD₆₀₀ reached 1.0. Isopropyl β-D-thiogalactoside (IPTG, AMERSCO) was added at a final concentration of 1 mM and cells were then induced for protein expression at 16 °C for 16 h. Collected cells were broken by sonication in lysis buffer (50 mM Na₂HPO₄, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, 1 mM PMSE, pH 8.0) and centrifuged at 15,000 rpm for 1 h at 4 °C. The supernatants then passed through a 0.22 µm filter and were ready for subsequent purification. The recombinant proteins were purified at 4 °C with a HisTrap Ni-NTA column (5 ml, GE Healthcare) and Superdex 75 PG chromatography (120 ml, GE Healthcare) using an ÄKTA pure system (GE Healthcare). The equilibration buffer (Buffer A: 50 mM Na₂HPO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0) and elution buffer (Buffer B: 50 mM Na₂HPO₄, 300 mM NaCl, 500 mM imidazole, pH 8.0) were used for affinity chromatography; the gel filtration buffer (10 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5) was used for size exclusion column. The obtained proteins (about 10–15 mg from 2 L of bacterial cells) were dissolved into gel filtration buffer containing a final concentration of 30% glycerol (v/v) and stored at –80 °C after a fast freezing with nitrogen liquid.

Crystallization of hNEIL1-dsDNA complexes. Damage-containing oligonucleotides were firstly mixed with the 13-mer complementary strand in a ratio of 1:1.2 and annealed in 1× TN buffer (10 mM Tris, 100 mM NaCl, pH 7.5) to a final concentration of 1 mM. Before the crystallization step, hNEIL1 proteins were first diluted with gel filtration buffer (with no DTT) and concentrated to 8–10 mg/ml to remove the glycerol component by buffer exchange using a 10-kDa ultrafiltration centrifugal tube (Amicon Ultra, Millipore). 24 nmol double-strand DNA was added into 20 nmol hNEIL1 protein and mixed gently by pipetting to give a final protein concentration of ~6 mg/ml. The mixture was then placed on ice for 30 min to form the protein–DNA complex. hNEIL1(Δ95 R/K242) proteins were used to co-crystallize with double-strand DNA containing FDHU or dT. For DNA with 5-OHU, Gh, Sp1, DHU or DHT base damage, hNEIL1(Δ95 P2G, E3Q, R/K242) proteins with a GS linker (203–222: GSSGGG) were used for co-crystallization. Orthorhombic rod-shaped crystals grew in 1–2 weeks at 4 °C, in hanging drops containing 2.0 µl of complex solution and 0.8 µl of reservoir solution (100 mM NaCl, 50 mM MgCl₂, 100 mM cacodylic acid, pH 6.5, 18% PEG 8 K, and 10 mM spermidine). Crystals were transferred into cryoprotectant solution containing reservoir solution plus 25% glycerol (v/v) for equilibration of 1–2 h and eventually stored in liquid nitrogen for data collection.

Data collection and structure determination. X-ray diffraction data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamlines BL17U, BL18U1, and BL19U1 with Mar CCD⁵¹. The raw data were processed with HKL-2000/3000^{52,53} and phased by molecular replacement with Phaser using the previously determined hNEIL1–DNA complex structure (PDB code: 5ITY) as a searching model. The resulting initial models were built in COOT⁵⁴ and refined with the program REFMACS from CCP4 suite⁵⁵. The data collection and refinement statistics for all the structures are listed in Supplementary Table 1. Simulated-annealing composite omit electron density map⁵⁶ of corresponding structures were generated by phenix⁵⁷ and molecular graphic figures were created with PyMOL (<https://www.pymol.org>).

DNA cleavage assays. The DNA glycosylase activity of hNEIL1 proteins was evaluated with single-turnover cleavage experiments. Briefly, the 30-mer damage-containing oligonucleotides were firstly annealed with corresponding complementary strands as described above. Then, 20 nM of substrate DNA was mixed with 200 nM active enzyme at 16 °C in a 140 µl of total volume containing 20 mM Tris–HCl, pH 7.5, 10 mM EDTA, 0.1 mg/ml BSA, and 150 mM NaCl. 10 µl of aliquots were removed from the reaction mixture at different time points (0.083, 0.25, 0.50, 1.0, 2.0, 5.0, 15, 30, 60, 90, 120, 150, and 180 min). In view of a relatively slow cleavage rate of hNEIL1 on dT, the sample collection time window was extended to 72 h at 37 °C for DNA substrate containing a T:C mismatch. The reaction was immediately stopped with 10 µl formamide DNA-loading buffer (98% formamide, 10 mM EDTA, pH 8.0, 0.025% bromophenol blue, 0.025% xylene cyanole) followed by heating to 95 °C for 5 min. Inactivated samples were kept on ice until all samples for the time series were obtained. The reaction products were finally separated by 15% PAGE containing 8 M urea and visualized by ChemiDoc MP imaging system (Bio-Rad). The fraction of product for each time point was calculated by dividing the intensity of the product band by the sum of intensities of both product and substrate bands in each lane. Three independent experiments were performed for all the substrate–enzyme combinations. To calculate the catalytic turnover rate k_{obs} (min^{–1}), the following nonlinear regression (one-phase association) model was used to fit the curves in GraphPad Prism 7.0: fraction product = $A(1 - \exp(-k_{\text{obs}}t))$, where A represents the amplitude, k_{obs} is the rate constant, and t is the reaction time (in minutes).

Calculation overview. The all-atom models for 15 different hNEIL1–DNA complexes (Supplementary Table 2) were built on the basis of the crystal structures. In some crystal structures where the loop regions were less resolved, we always consulted the high-resolution crystal structures which exhibit the corresponding conformations when possible. The all-atom complex structures were modeled by AMBER-ff99SB force field⁵⁸ and saturated with explicit water models. Before productive MD simulations, all of the all-atom models first underwent relaxations to reduce and equilibrate the improper local structures. The tautomerization energy of the active pocket (i.e. ΔG_{chem}) in all the atomistic models and the free energy profile of 244-in catalysis pathway were achieved using QM/MM simulations, where the semi-empirical SCC-DFTB^{59,60} method was employed to describe the QM-treated regions of the atomistic models. The free energy catalysis pathway was calculated by means of umbrella sampling⁶¹ along prescribed reaction coordinates following the related studies³⁹. Force-field simulations were performed to adaptively sample the metastable conformations of the recognition loop, and to compute the free energy of loop conformational transitions (i.e. ΔG_{conf}) via enhanced sampling. Besides, brute-force force-field simulations were also conducted to investigate the structural fluctuations of the active pocket (including the Y244-base stacking and P2-base distance, etc.). All the simulations were conducted on AMBER18 molecular simulation package⁶².

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6LWA, 6LWB, 6LWC, 6LWD, 6LWF, 6LWG, 6LWH, 6LWI, 6LWJ, 6LWK, 6LWL, 6LWM, 6LWN, 6LWO, 6LWP, 6LWQ and 6LWR. Previously reported hNEIL1(R242)-Tg structure and hNEIL1(R242)-APO structure are available from the PDB under accession ID 5ITY and 5ITQ, respectively. The RNA-seq data that support the findings of this study are available in the NCBI Sequence Read Archive under accession number SRP039090. Source data are provided with this paper.

Received: 24 December 2020; Accepted: 14 June 2021;

Published online: 05 July 2021

References

- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Hoeijmakers, J. H. DNA damage, aging, and cancer. *N. Engl. J. Med.* **361**, 1475–1485 (2009).
- Roos, W. P., Thomas, A. D. & Kaina, B. DNA damage and the balance between survival and death in cancer biology. *Nat. Rev. Cancer* **16**, 20–33 (2016).
- Krokan, H. E. & Bjoras, M. Base excision repair. *CSH Perspect. Biol.* **5**, a012583 (2013).
- David, S. S., O’Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**, 941–950 (2007).
- Hegde, M. L., Hazra, T. K. & Mitra, S. Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells. *Cell Res.* **18**, 27–47 (2008).

7. Mullins, E. A., Rodriguez, A. A., Bradley, N. P. & Eichman, B. F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* **44**, 765–781 (2019).
8. Bandaru, V., Sunkara, S., Wallace, S. S. & Bond, J. P. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair* **1**, 517–529 (2002).
9. Hazra, T. K. et al. Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl Acad. Sci. USA* **99**, 3523–3528 (2002).
10. Morland, I. et al. Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.* **30**, 4926–4936 (2002).
11. Rosenquist, T. A. et al. The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. *DNA Repair* **2**, 581–591 (2003).
12. Krishnamurthy, N., Zhao, X., Burrows, C. J. & David, S. S. Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochem.-Us* **47**, 7137–7146 (2008).
13. Vik, E. S. et al. Biochemical mapping of human NEIL1 DNA glycosylase and AP lyase activities. *DNA Repair* **11**, 766–773 (2012).
14. Yeo, J., Goodman, R. A., Schirle, N. T., David, S. S. & Beal, P. A. RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. *Proc. Natl Acad. Sci. USA* **107**, 20715–20719 (2010).
15. Dou, H., Mitra, S. & Hazra, T. K. Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* **278**, 49679–49684 (2003).
16. Albelazi, M. S. et al. The biochemical role of the human NEIL1 and NEIL3 DNA glycosylases on model DNA replication forks. *Genes* **10**, 315 (2019).
17. Dou, H. et al. Interaction of the human DNA glycosylase NEIL1 with proliferating cell nuclear antigen. The potential for replication-associated repair of oxidized bases in mammalian genomes. *J. Biol. Chem.* **283**, 3130–3140 (2008).
18. Hegde, M. L. et al. Prereplicative repair of oxidized bases in the human genome is mediated by NEIL1 DNA glycosylase together with replication proteins. *Proc. Natl Acad. Sci. USA* **110**, E3090–E3099 (2013).
19. Hegde, P. M. et al. The C-terminal Domain (CTD) of Human DNA Glycosylase NEIL1 Is Required for Forming BERosome Repair Complex with DNA Replication Proteins at the Replicating Genome: DOMINANT NEGATIVE FUNCTION OF THE CTD. *J. Biol. Chem.* **290**, 20919–20933 (2015).
20. Shinmura, K. et al. Inactivating mutations of the human base excision repair gene NEIL1 in gastric cancer. *Carcinogenesis* **25**, 2311–2317 (2004).
21. Vartanian, V. et al. The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. *Proc. Natl Acad. Sci. USA* **103**, 1864–1869 (2006).
22. Forsbring, M. et al. Catalytically impaired hMYH and NEIL1 mutant proteins identified in patients with primary sclerosing cholangitis and cholangiocarcinoma. *Carcinogenesis* **30**, 1147–1154 (2009).
23. Mori, H. et al. Deficiency of the oxidative damage-specific DNA glycosylase NEIL1 leads to reduced germinal center B cell expansion. *DNA Repair* **8**, 1328–1332 (2009).
24. Canugovi, C. et al. Endonuclease VIII-like 1 (NEIL1) promotes short-term spatial memory retention and protects from ischemic stroke-induced brain dysfunction and death in mice. *Proc. Natl Acad. Sci. USA* **109**, 14948–14953 (2012).
25. Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
26. Muller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* **42**, 8592–8604 (2014).
27. Slyvka, A., Mierzejewska, K. & Bochtler, M. Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Sci. Rep.* **7**, 9001 (2017).
28. O'Brien, P. J. Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chem. Rev.* **106**, 720–752 (2006).
29. Slupphaug, G. et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92 (1996).
30. Parker, J. B. et al. Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* **449**, 433–437 (2007).
31. Banerjee, A., Yang, W., Karplus, M. & Verdine, G. L. Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* **434**, 612–618 (2005).
32. Mullins, E. A. et al. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. *Nature* **527**, 254–258 (2015).
33. Labahn, J. et al. Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**, 321–329 (1996).
34. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl Acad. Sci. USA* **97**, 13573–13578 (2000).
35. Bernal, K. G., Johansen, R. F. & Seeberg, E. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J.* **17**, 363–367 (1998).
36. Branum, M. E., Reardon, J. T. & Sancar, A. DNA repair excision nuclease attacks undamaged DNA. A potential source of spontaneous mutations. *J. Biol. Chem.* **276**, 25421–25426 (2001).
37. O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279**, 9750–9757 (2004).
38. O'Brien, P. J. & Ellenberger, T. The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *J. Biol. Chem.* **279**, 26876–26884 (2004).
39. Zhu, C. et al. Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair. *Proc. Natl Acad. Sci. USA* **113**, 7792–7797 (2016).
40. Nagar, B. et al. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859–871 (2003).
41. Shan, Y. et al. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl Acad. Sci. USA* **106**, 139–144 (2009).
42. Sadeghian, K. et al. Ribose-protonated DNA base excision repair: a combined theoretical and experimental study. *Angew. Chem. Int. Ed. Engl.* **53**, 10044–10048 (2014).
43. Zhang, J. et al. Deep representation learning for complex free-energy landscapes. *J. Phys. Chem. Lett.* **10**, 5571–5576 (2019).
44. Haynes, W. M. *CRC Handbook of Chemistry and Physics* 96th edn (CRC Press, 2016).
45. Kobayashi, K. in *Encyclopedia of Astrobiology* (eds Gargaud, M. et al.) 85–85 (Springer, 2011).
46. Klein, T. et al. Structural and dynamic insights into the energetics of activation loop rearrangement in FGFR1 kinase. *Nat. Commun.* **6**, 7877 (2015).
47. Roy, L. M. et al. Human polymorphic variants of the NEIL1 DNA glycosylase. *J. Biol. Chem.* **282**, 15790–15798 (2007).
48. Prakash, A., Carroll, B. L., Sweasy, J. B., Wallace, S. S. & Double, S. Genome and cancer single nucleotide polymorphisms of the human NEIL1 DNA glycosylase: activity, structure, and the effect of editing. *DNA Repair* **14**, 17–26 (2014).
49. Teoh, P. J. et al. Aberrant hyperediting of the myeloma transcriptome by ADAR1 confers oncogenicity and is a marker of poor prognosis. *Blood* **132**, 1304–1317 (2018).
50. Anadon, C. et al. Gene amplification-associated overexpression of the RNA editing enzyme ADAR1 enhances human lung tumorigenesis. *Oncogene* **35**, 4422 (2016).
51. Wang, Q.-S. et al. Upgrade of macromolecular crystallography beamline BL17U1 at SSRF. *Nucl. Sci. Technol.* **29**, 68 (2018).
52. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
53. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution-from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
55. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
56. Hodel, A., Kim, S.-H. & Brunger, A. T. Model bias in macromolecular crystal structures. *Acta Crystallogr. Sect. A* **48**, 851–858 (1992).
57. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D* **75**, 861–877 (2019).
58. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
59. de, M. S. G., Walker, R. C., Elstner, M., Case, D. A. & Roitberg, A. E. Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package. *J. Phys. Chem. A* **111**, 5655–5664 (2007).
60. Elstner, M. et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998).
61. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
62. Case, D. A. et al. *AMBER 2018* (University of California, 2018).

Acknowledgements

The authors would like to thank Prof. Junyu Xiao for helpful advice, Lining Lu, Jinyi Bai, Xing Che, and Zhaozhu Sun for technical assistance. We thank the staffs from BL17U1/BL18U/BL19U1 beamline of National Facility for Protein Science Shanghai (NFPS) at Shanghai Synchrotron Radiation Facility for assistance during data collection. We are also

grateful to National Center for Protein Sciences at Peking University in Beijing, China for facility support. This work was supported by the National Natural Science Foundation of China (nos. 91953201, 21825701 to C.Y., and nos. 92053202, 22050003, 21821004 to Y.Q.G.), the Fok Ying Tung Education Foundation (no. 161018 to C.Y.) and the National Key Research and Development Program of China (no. 2017YFA0204702 to Y.Q.G.).

Author contributions

M.L., J.Z., C.Z., Y.Q.G., and C.Y. designed the research. M.L. and C.Z. performed the structural studies with the help of L.L. and H.Z.; J.Z. performed the computational stimulations; M.L. synthesized and characterized the FDHU phosphoramidite; M.L. and X.Z. performed the biochemical experiments with the assistance of W.X. and Y.Y.; M.L., J.Z., Y.Q.G., and C.Y. wrote the manuscript with contributions from all authors. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24431-y>.

Correspondence and requests for materials should be addressed to Y.Q.G. or C.Y.

Peer review information *Nature Communications* thanks Aishwarya Prakash and other, anonymous, reviewers for their contributions to the peer review of this work.

Reprints and permission information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021