AlkB dioxygenase preferentially repairs protonated substrates: specificity against exocyclic adducts and molecular mechanism of action*

Agnieszka M. Maciejewska[1], Jarosław Poznański, Zuzanna Kaczmarska, Beata Krowisz, Jadwiga Nieminuszczy, Agnieszka Polkowska-Nowakowska, Elżbieta Grzesiuk and Jarosław T. Kuśmierek

From the Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

* Running title: *AlkB dioxygenase prefers protonated substrates*

To whom correspondence should be addressed: Agnieszka M. Maciejewska, Institute of Biochemistry and Biophysics, Pawinskiego 5A street, 02-106 Warsaw, Poland. Tel.: +48 22 5923512; Fax: +48 22 5922190; E-mail: agniesza@ibb.waw.pl

Keywords: AlkB dioxygenase; DNA repair; exocyclic adducts to DNA bases; etheno adducts to DNA bases; molecular modeling

Background: AlkB dioxygenase removes alkyl and exocyclic lesions *via* an oxidative mechanism restoring the native DNA bases.
Results: AlkB repair efficiency is pH- and Fe(II) -concentration - dependent, which correlates with the substrate pK_a.
Conclusion: AlkB recognizes and repairs protonated substrates.
Significance: This study provides experimental evidence for the molecular mechanism of action of AlkB.

SUMMARY

Efficient repair by *Escherichia coli* AlkB dioxygenase of exocyclic DNA adducts: 3,N^4-ethenocytosine, 1,N^6-ethenoadenine, 3,N^4-α-hydroxyethanocytosine and, reported here for the first time, 3,N^4-α-hydroxypropanocytosine, requires higher Fe(II) concentration than the reference 3-methylcytosine. The pH optimum for the repair follows the order of pK_a values for protonation of the adduct suggesting that positively charged substrates favorably interact with the negatively charged carboxylic group of Asp135 side-chain in the enzyme active centre. This interaction is supported by molecular modeling indicating that 1,N^6-ethenoadenine and 3,N^4-etheno-cytosine are bound to AlkB more favorably in their protonated cationic forms. An analysis of the pattern of intermolecular interactions that stabilize the location of the ligand points to a role of Asp135 in recognition of the adduct in its protonated form. Moreover, also *ab initio* calculations underline the role of substrate protonation in lowering the free energy barrier of the transition state of epoxidation of the ethenoadducts studied. The observed time-courses of repair of mixtures of stereoisomers of 3,N^4-α-hydroxyethanocyto-sine or 3,N^4-α-hydroxypropanocytosine are unequivocally two-exponential, indicating that the respective isomers are repaired by AlkB with different efficiencies. Molecular modeling of these adducts bound by AlkB allowed evaluation of the participation of their possible conformational states in the enzymatic reaction.

Exocyclic DNA base adducts are produced by a variety of bifunctional electrophilic agents of exogenous and endogenous origin. These agents include products of endogenous lipid peroxidation occurring as a consequence of oxidative stress: α,β-unsaturated aldehydes (enals) such as malondialdehyde, acrolein $(ACR)^2$, croton-aldehyde or *trans*-4-hydroxy-2-nonenal (1). Other sources of such compounds are environmental pollutants (tobacco smoke, automobile exhaust), dietary toxins (formed, for example, in over-fried food) or metabolites of the industrial carcinogen vinyl chloride - chloroethylene epoxide and chloroacetaldehyde (CAA).

Structurally, exocyclic adducts are analogous but can differ in ring size, saturation, angularity and nature or location of substituents. Some of them occur as stereoisomers (2). The rings, usually 5- or 6-membered, are formed *via* bridging of exocyclic NH_2 group with an endocyclic nitrogen atom of a purine or pyrimidine base. The only DNA base lacking an

exocyclic NH_2 group and thus resistant to this kind of modification is thymine (3). Reaction of DNA bases with CAA proceeds *via* hydroxyethano intermediates, which are spontaneously dehydrated to an etheno (ε) form (4). We have shown that the most stable among them – 3,N^4-α-hydroxyethanocytosine (HEC), with a $t_{½}$ at 37ºC of about 24 hours (5), plays an independent role in mutagenesis (6,7).

Exocyclic ring modifications are highly mutagenic and cytotoxic DNA lesions, leading to miscoding or replication blocks and generating various degenerative disorders and several types of cancer. Different mechanisms appear to have evolved for the removal of specific exocyclic adducts. These include base excision repair, nucleotide excision repair, mismatch repair and AP endonuclease-mediated repair (for review see (2,8)). Recently, yet another, single-protein AlkB-directed repair system protecting cellular DNA and RNA against alkylating agents has been discovered.

E. coli AlkB is the best known member of the superfamily of α-ketoglutarate (αKG)- and iron-dependent dioxygenases (9) that removes alkyl lesions from DNA bases *via* an oxidative mechanism restoring the native bases in the DNA. Its primary substrates are 1-methyladenine (m^1A) and 3-methylcytosine (m^3C) (10,11). AlkB can also revert bulkier adducts such as ethyl, propyl and hydroxyalkyl groups (12,13) as well as exocyclic etheno (ε) and ethano adducts (14-16). It has been reported that AlkB also removes methyl groups from 1-methylguanine (m^1G) and 3-methylthymine (m^3T) although much less efficiently (17-19). Further studies have revealed that AlkB homologs are present in almost all organisms including human. Besides methylated ssDNA, also dsDNA and RNA can be substrates for AlkB and its homologs (20-22). As do other members of this superfamily, AlkB requires oxygen (O_2) and non-heme Fe(II) as cofactors and αKG as cosubstrate to initiate oxidative demethylation of DNA bases resulting in the formation of succinate and CO_2. The methyl and ethyl groups are hydroxylated and spontaneously released as formaldehyde or acetaldehyde, respectively (10-12). In the case of an etheno bridge the double bond is epoxidised, then spontaneously hydroxylated to glycol and released as glyoxal, leading to regeneration of the natural base (16,23). Recently, we have found that AlkB repairs not only εC, but also its precursor HEC and we also formulated a hypothesis that AlkB more readily recognizes and repairs substrates if they exist in a protonated form (7). This was also previously suggested by Koivisto *et al*. (18) and in crystallographic study by Yi *et al.* (23) both on the basis of differences of repair efficiency between the best AlkB substrates m^1A and m^3C and very poorly repaired non-protonated m^3T.

Here, we provide experimental evidence for the molecular mechanism of action confirming that *E. coli* AlkB dioxygenase preferentially repairs substrates in cationic form. Our study is focused on the specificity of AlkB-directed repair of various exocyclic DNA adducts which differ in their protonation ability. These include two unsaturated, five-membered etheno adducts – 1,N^6-ethenoadenine (εA) and 3,N^4-ethenocytosine (εC), both known AlkB substrates (14,16,24), and two saturated ones: the five-membered HEC and a six-membered acrolein adduct - 3,N^4-α-hydroxypropano-cytosine (HPC) (25,26). Additionally, as a positive control, m^3C was used (Fig. 1). HEC, described by us recently (7), together with HPC reported here for the first time, form a new, chemically distinct group of AlkB substrates. Using HPLC technique and single-stranded pentamer oligodeoxynucleotides containing the modifications studied, we established optimal pH, Fe(II) and αKG concentrations for the AlkB-directed repair reaction. In parallel, basing on the known crystal structure of AlkB in complex with Fe(II), αKG and methylated trinucleotide T(m^1A)T (PDB3I2O) (27), we performed molecular modeling of putative complexes with the modified TXT trinucleotides (where X represents a modified base). The consensus achieved between simulated annealing (SA) procedure coupled with molecular dynamics (MD) studies and screening for the best location for X done with the aid of a restricted docking algorithm allowed building of reliable structures for all of the analyzed trinucleotides in complex with AlkB.

EXPERIMENTAL PROCEDURES

Detailed descriptions of all the methods are presented in Supplemental Materials.

Preparation of AlkB substrates - We reacted pentamer deoxynucleotides, TTXTT, where X is cytosine (C) or adenine (A), with CAA, ACR, or MMS, and obtained derivatives containing modified cytosines: HEC, εC, m^3C and HPC, respectively, or εA. After HPLC purification the modified pentamers were used as AlkB substrates.

AlkB assay - The reaction mixtures in a volume of 20 μL contained: appropriate 50 mM buffer, 1 mM dithiothreitiol, $Fe(NH_4)_2(SO_4)_2$ and αKG at varying concentrations, and 500 pmoles of the substrate. Reactions were started by addition of purified AlkB protein or the component studied, allowed to proceed at 37°C, stopped by adding 230 μl of ice-cold water and frozen at -20ºC to deactivate AlkB and then analyzed by HPLC. Enzyme concentrations and reaction time (5 – 30 min) were screened in preliminary experiments to produce results enabling estimation of adduct repair efficiency. We began with determination of optimal pH for a given substrate, then at that particular pH we established optimal Fe(II), and finally optimal αKG concentration. The time course reactions for rate determinations were carried out under conditions optimal for each adduct and also in so called "physiological" conditions (50 mM HEPES/NaOH buffer pH 7.5, 100 μM $Fe(NH_4)_2(SO_4)_2$ and 50 μM αKG).

Ab initio calculations - The initial structures of modified bases were adopted from the structure of corresponding native nucleoside, where the sugar moiety was mimicked by a methyl group. This was followed by *ab initio* analysis with the aid of the Firefly version 7.1 program (28). The density functional theory (DFT) calculations were performed with the B3LYP functional (29), using the 6-31G(d,p) basis set (30), previously found applicable in the analysis of similar systems (31,32).

Protein-ligand interaction: Simulated Annealing procedure - Initial coordinates of the complexes were adopted from the structure of T(m^1A)T bound to AlkB (PDB3I2O) (27). Structural calculations were performed with the aid of the SA protocol implemented in X-PLOR (33,34).

Molecular dynamics in explicit aqueous solvent - MD analysis was done with the aid of Yasara-Structure package {Krieger, E. http://www.yasara.org} using standard Yasara2 force-field (35) extended for the non-standard ligands by adding their *ab initio*-derived topologies and charge distributions. The starting structures of the complexes were taken for the various ligands directly from the output of the SA procedure. The protein was then solvated by explicit water molecules to give an average solvent density of 1.004 g/ml. Since the structural alignment of all the accessible structures of AlkB (including those with ligands) are identical within the RMSD threshold of 0.3 Å (calculated for Cα atoms of residues located in secondary-structure elements with the only exception of Q48-M57 and G68-S79), coordinates of the protein residues that were separated by more than 6 Å from the modified base were fixed during MD simulations. Similarly, all distant water molecules (threshold 8 Å) were also frozen. For each of the base modification analyzed a 10ps MD simulation was performed to let the protein to tune its structure to accommodate the ligand, and the last 3 ps of the trajectory were then subjected to detailed analysis. The stability of the four complexes of the two possible stereoisomers of HEC and HPC, all in *anti*-conformation on glycoside angle, were additionally tested in the terms of 15ns molecular dynamics. Those simulations were done with the use of Yasara package in explicit water box of a size 60/63/60 Å using yasara2 force-field (35). The simulations were done in constant NTP ensemble, at 298 K, keeping the solvent density of 0.997 $g \cdot cm^{-3}$. To mimic the system close to the experimentally analyzed, all the modified nucleosides where flanked at both sides by deoxythymidine residues, which location was adopted directly from the crystal structure of T(m^3C)T trinucleotide bound to AlkB [3I49 PDB record]. The protein conformations and locations of modified residues were taken directly from the preliminary 10ps simulations. During the first 1ns of MD trace the coordinates of the whole protein, and also those of aromatic bases, were constrained. In the following steps all the side-chains of protein were iteratively unrestricted within the centered at iron atom sphere of a radius growing in four 1ns steps. After that the modified base, αKG, Fe(II) cation, the protein backbone and the flanking deoxythymidine residues were freed in 1ns steps. The last 5 ns of MD simulation were performed for the fully solvated system with the only constraints fixing the stacking orientation of thymine bases.

Docking procedure – estimation of free energy of binding - This was done by *de novo* docking performed for the MD-derived structures with the aid of the Autodock 4.2 program (36) implemented in the Yasara-Structure package. Interactions between ligands and the protein were scored by Van der Waals, electrostatic, hydrogen bonding and desolvation energy terms adopted from the AMBER03 force-field (37).

Numerical analysis of the rates of the repair process - Time-dependent changes in the concentrations of modified (substrate) and unmodified (repaired product) nucleotides were monitored in several independent experiments

performed for various rationally selected mixtures of reagents. For each substrate all the experimental data collected in a series of independent experiments (1…n) were analyzed together, assuming the common model of a second order reaction controlled by a single transition state, described by the following equation

$$dS_i(t)/dt = S_i(t) \cdot k \cdot [AlkB]_i \quad (1)$$

where $[AlkB]_i$ is the enzyme concentration in the i-th experiment, $S_i(t)$ is the time evolution of substrate concentration in that experiment, and k is the universal rate of repair related directly to a given type of substrate. The product $k \cdot [AlkB]_i$ represents the actual pseudo first order rate of the repair in the i-th experiment, k_i, that depends on the enzyme concentration, leading to a simplified pseudo first order equation.

$$dS_i(t)/dt = S_i(t) \cdot k_i \quad (1a)$$

The formal integration of equation (1) leads the formula

$$S_i(t) = S_i^0 - R_i \cdot (1 - e^{-k \cdot [AlkB]_i \cdot (t+\Delta)}) \quad (2)$$

in which S_i^0 and $[AlkB]_i$ are the initial concentrations of substrate and AlkB, respectively; k is the second order repair rate; Δ is a delay related to the setup of experimental procedure, and R_i is the upper limit of substrate that can be repair in the i-th experiment. For each substrate all the kinetic data were analysed together in terms of model (2) consisting of two global parameters (k, Δ) and n experiment-specific R_i values, giving together multi-brunch function (see Fig. 5 in Results and Supplementary Figs S1 and S2). These parameters were then optimized to obtain the best agreement with experimental data, in which every brunch of the curve represent individual experiment. In the case of HEC and HPC, the model (2) was additionally extended to describe the repair of mixture (1:α) of individual stereoisomers with two isomer-specific repair rates (k_1 and k_2),

$$S_i(t) = S_i^0 - R_i \cdot (1 + \alpha - e^{-k_1 \cdot [AlkB]_i \cdot (t+\Delta)} - \alpha \cdot e^{-k_2 \cdot [AlkB]_i \cdot (t+\Delta)}) \quad (3)$$

where α stands for relative ratio of two isomers. All the calculations were done with the aid of R package {http://www.R-project.org}.

RESULTS

Optimal reaction conditions for AlkB directed repair. Pentamers containing modified bases: m^3C, HPC, HEC, εC and εA were studied as potential substrates for AlkB protein. Using HPLC to separate the modified and unmodified (repaired) oligomers (see Fig. 2 as an example), we found that all the modifications were repaired by *E. coli* AlkB dioxygenase. The appearance of the repaired oligomers was also confirmed by mass spectrometry (MS) (Supplementary Table S2).

pH dependence – Optimal pH for m^3C and HPC repair was found to be 7.5, for HEC – 5.8, whereas both εA and εC were most efficiently repaired at pH 4.6. The determined experimentally values of optimal pH for repair reactions strictly followed the order of pK_a of the lesions (see Table 1). The optimal pH for the repair of adducts that display pK_a for protonation higher than 5.8 was generally one pH unit below the pK_a. This clearly indicates much more efficient repair for the substrate in its cationic form. The comparison of repair kinetics of εA determined at neutral and optimal pH is shown in Supplementary Fig. S1.

Fe(II) concentration dependence – Optimal Fe(II) concentration was in the range between 50 μM for m^3C repair at pH 8.0 and 5 mM for εC repair at optimal pH of 4.6 (far away from the physiological condition). It anti-correlates with pK_a and optimal pH for the repair of a given adduct (see Fig. 3). Corresponding values are collected in Table 1.

α-Ketoglutarate concentration dependence - For αKG we did not observe as strong a tendency as for Fe(II). The optimal αKG concentrations varied in the range of 25 to 100 μM, definitely below the commonly used 500 μM. Although we applied the lower, optimal αKG concentrations in further experiments, we did not focus on that subject.

The experimentally observed dependences of the repair rates on pH, iron and αKG concentrations are presented in Supplementary Fig. S3.

Pseudo first order kinetics of AlkB repair (equation 1a). The kinetics of the repair processes for εA and HPC monitored for a long time using relatively high AlkB concentration are shown in Fig. 4. The linearity observed for εA repair in Fig. 4A identifies, in accordance to equation 1a, pseudo first order reaction. In the case of HPC two different asymptotes clearly identify two species visibly differing in their repair rates, however both processes can be individually considered as pseudo first order (see Fig. 4B).

Second order kinetics of AlkB repair (equation 1). The initial rates of the repair of m^3C and εA were estimated for nine individual experiments performed for three rationally selected AlkB

concentrations. These short-time experiments (shown in Supplementary Fig. S4 for εA, as an example) clearly demonstrated the pseudo-first order of the repair kinetics (equation 1a). Moreover, the pseudo first order rates, k_i, are proportional to the AlkB concentration, just confirming a simple second order rate of the repair process (equation 1, see Table 1 for rates estimated for εA and m^3C).

Time course. The kinetics of the repair process measured enabled us to determine the relative efficiency of this process normalized for enzyme concentration. This is expressed further as k_1 (and additionally k_2 for HEC and HPC), representing the rate of repair by 1 nmol of enzyme per second (Table 1). In order to improve the reliability of the rates obtained for a given adduct a set of independent experiments were analyzed in parallel, as described in Experimental Procedures. Results of most of the kinetic experiments performed at both physiological and "optimal" pH conditions were in agreement with the branches of the hypothetical single-exponent concentration-adjusted repair curve (see Table 1 for the rates). The quality of the fit could be judged in the terms of Root of mean squared deviation (RMSD) normalized per degree of freedom (df). Those values fall in the range of 10-20 pmol of substrate (*i.e.* 2-5%), strictly corresponding to the experimental errors estimated in selected experiments (see error bars in Fig. 5 and Supplementary Fig. S2 as an example). The exceptions were HEC and HPC, for which two stereoisomers on the hydroxylated exocyclic carbon atoms were generated upon the chemical reaction. For these two adducts the repair process is visibly better described by two-exponential curves with RMSD reduced from 25 to 20, and from 30 to 15 pmol for HEC and HPC, respectively (see Table 1 for details), clearly indicating that stereoisomers of these two pairs differ significantly in their susceptibility to AlkB-directed repair. In fact, reliable estimates of the rates of repair of individual HEC stereoisomers could only be at the optimal pH of 5.8. Kinetic curves of the repair of HPC by various concentrations of AlkB are presented in Fig. 5 and Supplementary Fig. S2. In general, we have found that the repair efficiency of the adducts studied follows the order: m^3C > HEC(fast) > HPC(fast) > εA > εC > HEC(slow) and HPC(slow) (for more details see Discussion). In Fig. 2 we have shown HPLC profile of HPC repair. Almost complete disappearance of the peak corresponding to HPC confirms that in high AlkB concentration both components of the stereoisomeric mixture are completely repaired.

Ab initio studies of adducts – protomeric and dissociation equilibria. In order to estimate the influence of substrate protonation on the efficiency of AlkB-directed repair various tautomeric/protomeric forms of the adducts studied were tested. A comparison of calculated free energies of protonation, ΔG_{prot}, with the experimentally established values of pK_a was used for verification of the applicability of the used level of theory. The good correlation obtained validated the method used (see Supplementary Fig. S5). Consequently, for each molecule its lowest-energy protomeric state was used in following molecular mechanics simulations. Analysis of the electron density maps clearly evidenced delocalization of the charge through a large parts of the respective aromatic electron systems (Supplementary Fig. S6). Additionally, according to the postulated mechanism of the first step of εA and εC oxidation (16,23), their appropriate epoxides were also analyzed (see Discussion for details).

Structures of enzyme-substrate complexes - To predict the mode of accommodation of the substrates in the active site of the enzyme, preliminary theoretical models of AlkB complexes with modified trinucleotides were prepared using the SA protocol followed by 10-ps MD of the fully solvated system, with the "far" residues fixed. For verification of the quality of the structural models see Supplementary Fig. S7. The resulting trajectories confirmed that AlkB easily accommodates all the adducts studied. A detailed inspection of the resulting structures points to a common pattern of protein-ligand interaction involving Thr51, Tyr55, Met57, Ser58, Trp69, Tyr76, Lys127, Leu128, Ser129, His131, Asp133, Asp135, Gln132 and Arg161, among which Trp69, His131 and Asp135 display the largest contact areas with the modified base.

The detailed analysis of location in the active site of AlkB of both stereoisomers of T(HEC)T and T(HPC)T trinucleotides clearly demonstrated that all the interactions crucial for AlkB activity are strictly preserved. Thus, in accordance with the results obtained for AlkB, the orientation of αKG molecule is stabilized by the interactions with polar side-chains of residues Arg 204, Arg 210 and Asn 120. Among the latter, the hydrogen bond between Asn side-

chain amide and α carbonyl group of αKG could be observed during approximately 80% of a simulation time, while the two others are almost strictly present. During the majority of the simulation time the N4 amino group of modified bases interacts with the side-chain of Asp 135. In all four complexes the Fe(II) cation is within the 6Å distance to the alkyl adducts.

The further analysis concerns the stability of the bound forms, which was estimated by a free-energy scoring function implemented in Autodock (36). This was done for the last 30 snapshots uniformly sampled from the registered MD trace. The obtained structures of the neutral and charged forms of the adducts bound to AlkB indicate that εA and εC are bound more favorably (*ca.*1 kcal/mol) in their cationic forms (see Fig. 6 for structures). An analysis of the pattern of intermolecular interactions that stabilize the location of the ligand points to the role of Asp135 in recognition of N-alkylated base in its protonated form, while Trp69 and His131 sandwich the modified base orienting it in relation to the iron redox center (see Fig. 7 A and C). In general, the preference of AlkB towards cationic substrates could be clearly related to the distribution of electrostatic potential within the binding pocket area (see colored regions inside the pocket accommodating εA in Fig. 7 B and D).

The stereoisomeric pairs of HEC and HPC also differ in their free energy of binding, but it should be pointed out that the observed differences in their repair efficiency should be attributed to the differences in the accessibility of Cβ of HEC and Cγ of HPC (Fig. 1) for the redox attack by Fe(II) rather than to their different binding energies (Fig. 8).

DISSCUSION

The aim of this work was to assign the role of *E. coli* AlkB dioxygenase in the repair of exocyclic DNA adducts and to establish crucial aspects of its mechanism of action. It is known that the lesions studied here can be repaired *via* the base excision repair (BER) pathway. The exocyclic adducts εC, HEC, HPC (26,38,39) and, to some extent, also εA (24,40) are excised from DNA by bacterial Mug glycosylase. It has been shown in the *in vitro* experiments that εA is removed from DNA by AlkA glycosylase (41). However, we found no evidence of this mode of repair *in vivo* (24). Here, we have shown that all the mentioned above modifications are repaired by AlkB. Bearing in mind that HPC was less efficiently removed by Mug from DNA than εC and HEC (26), and that εA is a very poor substrate for *E.coli* glycosylases it seems that AlkB plays a crucial role in the repair of these lesions. Our previous *in vivo* observations (7) strongly argue in favor of AlkB as a principal repair enzyme also for the HEC and εC lesions. It the case of all the above adducts Mug may act only as a back-up enzyme for AlkB. Nevertheless, it cannot be excluded that the enzymes complement each other since AlkB preferentially removes modifications from single stranded substrates, whereas Mug acts on dsDNA. The fact that two independent systems, AlkB direct repair and BER, are involved in the repair of exocyclic adducts strongly emphasizes how dangerous these lesions are for the genome integrity.

Repair of methylated and exocyclic adducts by AlkB protein in vitro. The best AlkB substrates m^3C and m^1A bear a positive charge at physiological pH, whereas neutral m^3T and m^1G are repaired very poorly (17-19). As we have shown in this study, efficient repair of the exocyclic adducts studied also occurs at a pH corresponding to their protonated, positively charged forms. These results suggest that in the AlkB enzyme's active centre the substrate is bound favorably in a cationic form. Table 1 shows absolute repair rates (*i.e.*, k_1, and, when applicable, k_2) estimated for the repair of m^3C and the exocyclic adducts at the physiological pH of 7.5, and at the optimal pH determined for each substrate. In the case of HPC pH 7.5 was also found to be optimal. Not surprisingly, it appeared that m^3C is better repaired at pH 7.5, where the residue is almost fully protonated, whereas at the commonly used pH of 8.0 17% of m^3C remains neutral.

It should be stressed that for εC and εA, whose pK_a's are 3.7 and 3.9, respectively, the optimal pH of 4.6 is substantially higher than their pK_a values. However, the latter value corresponds exactly to the pK_a for the second carboxyl group of αKG (pK_a1 = 2.47; pK_a2 = 4.68 (42)), while the pK_a of the aspartic acid side-chain carboxyl group is usually close to 3.9. Taking into consideration that in similar systems the binding of Fe(II) cation has been shown to be a metal-to-ligand charge-transfer process (43) requiting electron from the coordination center, the existence of lower pH limit for the repair reaction confirms that the doubly dissociated anionic form of αKG is the only form of the co-substrate involved in the catalytic reaction of εC

and εA repair at pH 4.6. In the case of adducts displaying lower pK_a values, the optimal pH for their repair is a compromise between the two factors determined the efficiency of the repair. A decrease of pH may also induce AlkB unfolding. The Fe(II) cofactor binding has also been confirmed to be pH-dependent, but this could be partially compensated for by variation in its concentration. The pH dependence of the optimal iron concentration has to be judged in terms of protonation of the imidazole ring of histidines that are involved in Fe(II) coordination in the AlkB catalytic centre. Thus, the protonation of His131 and His187 at a lower pH needs to be compensated for by an increase of Fe(II) concentration, and consequently an exact (anti) correlation is observed between the optimal values of pH and the logarithm of iron concentration (see Fig. 3).

On the other hand, strong lowering of pH of the reaction medium is accompanied by an increase of the population of the favored cationic form of the adduct. As a consequence, for high-pK_a adducts in the optimal-pH state (one unit below pK_a) the AlkB state is identical with that at the neutral pH, while the adducts are almost fully protonated. In the case of εC and εA, at a pH required for their efficient protonation (*i.e.*, ~2.7) AlkB would be inactive and also the αKG co-factor would only be partially dissociated. The pH optimum of the reaction at *ca*. one unit above the substrate's pK_a thus allows the enzyme to retain activity and αKG to be doubly dissociated, while still producing *ca.* 10% of the substrate in the protonated form.

Conformational changes of AlkB in complex with charged adducts. Inspection of the MD trajectories clearly indicated that conformational preferences of Asp135 are coupled with the ionic state of the ligand, which cationic form favor the Asp135 side-chain rotamer with side-chain carboxyl group located proximal to the protonated base (see Figs 6 and 8). According to the obtained 15ns MD trajectories, this interaction is responsible for the preferences of AlkB towards protonated substrates.

Ab initio analysis of free energy of activation for the oxidation of εA and εC - Oxidation of the double (etheno) bond is the first step of ethenoadducts' repair by AlkB resulting in the formation of a transient epoxide form of the modified base. The following steps, leading to regeneration of the unmodified base, occur spontaneously (16,23).

Although εA and εC have similar pK_a and are best repaired at the same pH, we observed a five-fold difference in their repair rates despite both adducts being (partially) protonated (see Table 1 for details). Therefore, we analyzed the thermodynamics of the repair process calculating with *ab initio* methods the free energy difference between εA and εC and their epoxy derivatives. We found that for the neutral molecule in aqueous solution the putative barrier is higher by ~ 0.4 and 0.7 kcal for εA and εC, respectively, in comparison to the same molecule in the monocationic state. This, in agreement with the experimentally confirmed tendency, indicates that the epoxidation step is 2-3 times faster for the charged substrates than for the neutral ones. The relative epoxidation barrier was estimated to be by 2.3 kcal/mol (for the protonated forms) and 2.5 kcal/mol (for the neutral forms) higher for εC than that for εA, indicating that εA is expected to be repaired ~ 100-fold faster than εC (neglecting the differences in their binding by the enzyme). Both, *in vitro* and *in silico* data, confirm that AlkB more readily repairs ethenoadducts in their cationic form, what at physiological conditions results in substantial decrease of their repair rates. On the other hand, our (7,24) and other group (16,44) results indicate that in *E. coli* cells AlkB does repair ε-adducts quite efficiently. One can thus speculate that *in vivo* some other, yet unknown, factors may be involved in AlkB-mediated direct reversal of ε-adducts.

Kinetics of the repair. We have experimentally proven that at high concentration of the enzyme repair reaction is clearly single-exponential (Fig. 4A) with the evident exception of HPC (Fig. 4B), and putatively HEC (results not as clear as for HPC possibly due to spontaneous HEC dehydratation during the time of the reaction (4)). Moreover, using short-time experiments with different AlkB concentrations (Supplementary Fig. S4 A) we have demonstrated that the series of pseudo first order kinetics of the repair (equation 1a) are in agreement with global model (equation 1) of the reaction (Supplementary Fig. S4 B). This validated the applicability of more general equation 1 for all AlkB concentrations used in long-time kinetics studies. As expected, the repair rates estimated for εA and m^3C with the aid of both approaches are virtually the same (Table 1). However, it should be emphasized that for low-picomolar AlkB concentration only the partial repair of sub-nanomolar substrate was

observed (see Fig. 5 and Supplementary Figs S1 and S2).

The proposed method is equivalent to the commonly used approach. However, the replacement of two-step analysis (*i.e.* time evolution of reagents in individual experiments followed by regression of the estimated rates against AlkB concentration) by one-step analysis (equation 2) reduces the number of required parameters from 3n+1 (*i.e.* k_i, R_i and Δ_i for every individual experiment, and slope k_i vs [AlkB] in the last step) down to n+2 (R_i for each experiment and global values of k and Δ).

In the case of HPC and HEC, which are mixture of stereoisomers, the proposed approach was simply necessarily required for data analysis, since the standard one is based on a very initial step of substrate repair, and thus the slower process cannot be analyzed. We have previously demonstrated that global multi-brunch analysis of a set of individual experiments was valuable in the analysis of circular dichroism, calorimetric and NMR data (46-48).

Mixtures of HPC and HEC stereoisomers. Two-exponential repair kinetics of HPC (Figs. 4 B and 5) and HEC indicates that both stereoisomers are repaired, although the efficiency of the repair differs by approximately one order of magnitude for HPC, and about six-fold in the case of HEC (Table 1).

In the reaction of dCMP with acrolein, an equimolar mixture of product diastereomers was observed (49). As a substrate for AlkB we have used pentadeoxynucleotide with the modified base surrounded by thymine residues. In Fig. 4 B the extrapolated asymptote representing slower HPC isomer repair would cross the vertical axis below 100 pmol. It suggests that the HPC isomers in the pentamer are not equimolar. Indeed, in preliminary NMR ROESY and TOCSY spectra recorded for T(HPC)T trideoxynucleotide we have evidenced two forms of HPC in the ratio 3:1 (unpublished). This demonstrates that, in contrast to free nucleotide (49), a pattern of steric interactions with dT flanking dC create in oligomer a chiral environment that favors one of two possible HPC stereoisomers upon dC modification by acrolein.

Inspection of the proposed structures of the R- and S-isomers of both compounds showed that for both S-isomers the exocyclic hydroxyl group points towards the iron redox center (Fig. 8). This should prevent an efficient oxidation of the exocyclic adduct. However, the unfavorable orientation of the substrate could be compensated for by the *anti* to *syn* transition around the glycosydic bond. As modeled (Supplementary Fig. S8), the relative orientation of the S-*syn* forms resembles that obtained for the R-*anti* ones, with the oxygen atom directed away from the iron atom. A similar orientation of the oxygen atom was recently determined for the ethenoadenine glycol adduct (23), which was identified as a transition state for the repair of εA (16,23).

It is worth mentioning (see Fig. 8 for details) the substantial differences in topology of patterns of intermolecular interactions caused by hydroxyl group of the lesion of HEC (A, B) and HPC (C, D). Thus hydroxyl groups in R stereoisomers, which are pointing towards iron atom, make hydrogen bonds with either Asp 135 or Glu 136 (B, D), that way interfering with the conserved interaction between Arg 210 and Glu 136 observed for S stereoisomers (A, C), and in all crystal structures of AlkB.

Although we confirmed that both stereoisomers of HEC and HPC are repaired (see Fig. 2, and also k_1 and k_2 rates in Table 1), at the moment we have no experimental basis to confirm whether the S forms are preferably repaired in the *syn* or the *anti* conformation. We also have no experimental evidence which of the HEC and HPC stereoisomers is repaired more efficiently, however, molecular modeling suggests that S stereoisomers may be preferably repaired (see discussion above and also Supplementary Fig. S8).

CONCLUSIONS AND PERSPECTIVES

Using experimental and *in silico* approaches, we have confirmed ours (7) and other authors' (18,23) suggestions that *E. coli* AlkB preferentially recognizes and repairs protonated substrates. The best AlkB substrates, m^1A and m^3C, exist in a cationic form at physiological pH. Here, we have experimentally demonstrated that also other AlkB substrates are more readily repaired at a pH corresponding to their cationic form. Employing computational methods we have established that a crucial role in recognition of the positively charged substrate is played by negatively charged Asp135 located in the proximity of the bound modified base. More generally, the negative electrostatic potential distributed over the whole area of the binding site is responsible for the substrate preferences of AlkB (Figs. 6 and 7).

Human AlkB dioxygenase homologs hABH1-3 (12,18,21,50,51) differ in their specificity to recognize ss- and dsDNA or the ability to repair RNA modifications but they share the same fold, including active centre organization with a non-heme Fe(II) cation, and as one can suppose on the basis of their known substrate specificity, may share with the bacterial enzyme the ability to recognize and repair much more readily protonated substrates.

Exocyclic DNA adducts are produced directly by environmental chemical compounds or occur as end-products of cellular oxidative stress. Ample data indicate that they play important roles in the development of several pathological states, including inflammation, chronic infection, metal imbalance and cancer (52). The ever expanding number of such adducts among AlkB substrates emphasizes the likely role of human AlkB homologs in the etiology and prevention of those pathologies.

REFERENCES

1. Bartsch, H. (1999) Keynote address: exocyclic adducts as new risk markers for DNA damage in man. *IARC Sci. Publ.*, 1-16
2. Hang, B. (2004) Repair of exocyclic DNA adducts: rings of complexity. *Bioessays* **26**, 1195-1208
3. Leonard, N. J. (1984) Etheno-substituted nucleotides and coenzymes: fluorescence and biological activity. *CRC Crit. Rev. Biochem.* **15**, 125-199
4. Krzyzosiak, W. J., Biernat, J., Ciesiolka, J., Gornicki, P., and Wiewiorowski, M. (1979) Further studies on adenosine and cytidine reaction with chloroacetaldehyde, a new support for the carbinolamine structure of the stable reaction intermediate and its relevance to the reaction mechanism and tRNA modification. *Polish. J. Chem.* **53**, 243-252
5. Kusmierek, J. T., and Singer, B. (1982) Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. *Biochemistry* **21**, 5717-5722
6. Borys, E., Mroczkowska-Slupska, M. M., and Kusmierek, J. T. (1994) The induction of adaptive response to alkylating agents in *Escherichia coli* reduces the frequency of specific C→T mutations in chloroacetaldehyde-treated M13 *glyU* phage. *Mutagenesis* **9**, 407-410
7. Maciejewska, A. M., Ruszel, K. P., Nieminuszczy, J., Lewicka, J., Sokolowska, B., Grzesiuk, E., and Kusmierek, J. T. (2010) Chloroacetaldehyde-induced mutagenesis in Escherichia coli: The role of AlkB protein in repair of 3,N4-ethenocytosine and 3,N4-α-hydroxyethanocytosine. *Mutat. Res.* **684**, 24-34
8. Krwawicz, J., Arczewska, K. D., Speina, E., Maciejewska, A., and Grzesiuk, E. (2007) Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease. *Acta Biochim. Pol.* **54**, 413-434
9. Aravind, L., and Koonin, E. V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, RESEARCH0007
10. Falnes, P. O., Johansen, R. F., and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178-182
11. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174-178
12. Duncan, T., Trewick, S. C., Koivisto, P., Bates, P. A., Lindahl, T., and Sedgwick, B. (2002) Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U S A* **99**, 16660-16665
13. Koivisto, P., Duncan, T., Lindahl, T., and Sedgwick, B. (2003) Minimal methylated substrate and extended substrate range of *Escherichia coli* AlkB protein, a 1-methyladenine-DNA dioxygenase. *J. Biol. Chem.* **278**, 44348-44354
14. Mishina, Y., Yang, C. G., and He, C. (2005) Direct repair of the exocyclic DNA adduct 1,N^6-ethenoadenine by the DNA repair AlkB proteins. *J. Am. Chem. Soc.* **127**, 14594-14595
15. Frick, L. E., Delaney, J. C., Wong, C., Drennan, C. L., and Essigmann, J. M. (2007) Alleviation of 1,N^6-ethanoadenine genotoxicity by the *Escherichia coli* adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U S A* **104**, 755-760

16. Delaney, J. C., Smeester, L., Wong, C., Frick, L. E., Taghizadeh, K., Wishnok, J. S., Drennan, C. L., Samson, L. D., and Essigmann, J. M. (2005) AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nat. Struct. Mol. Biol.* **12**, 855-860
17. Delaney, J. C., and Essigmann, J. M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U S A* **101**, 14051-14056
18. Koivisto, P., Robins, P., Lindahl, T., and Sedgwick, B. (2004) Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.* **279**, 40470-40474
19. Falnes, P. O. (2004) Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res.* **32**, 6260-6267
20. Aas, P. A., Otterlei, M., Falnes, P. O., Vagbo, C. B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E., and Krokan, H. E. (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859-863
21. Falnes, P. O., Bjoras, M., Aas, P. A., Sundheim, O., and Seeberg, E. (2004) Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res.* **32**, 3456-3461
22. Ougland, R., Zhang, C. M., Liiv, A., Johansen, R. F., Seeberg, E., Hou, Y. M., Remme, J., and Falnes, P. O. (2004) AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol. Cell* **16**, 107-116
23. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C. G., Cui, Q., and He, C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468**, 330-333
24. Maciejewska, A. M., Sokolowska, B., Nowicki, A., and Kusmierek, J. T. (2011) The role of AlkB protein in repair of 1,N6-ethenoadenine in Escherichia coli cells. *Mutagenesis* **26**, 401-406
25. Chenna, A., and Iden, C. R. (1993) Characterization of 2'-deoxycytidine and 2'-deoxyuridine adducts formed in reactions with acrolein and 2-bromoacrolein. *Chem. Res. Toxicol.* **6**, 261-268
26. Borys-Brzywczy, E., Arczewska, K. D., Saparbaev, M., Hardeland, U., Schar, P., and Kusmierek, J. T. (2005) Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts. *Acta Biochim. Pol.* **52**, 149-165
27. Yu, B., and Hunt, J. F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc. Natl. Acad. Sci. U S A* **106**, 14315-14320
28. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. J., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., and Montgomery, J. A. J. (1993) General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347-1363
29. Hertwig, R. H., and Koch, W. (1997) On the parameterization of the local correlation functional: What is Becke-3-LYP? *Chem. Phys. Lett.* **268**, 345-351
30. Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J., and Pople, J. A. (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **77**, 3654-3665
31. Poznanski, J., Najda, A., Bretner, M., and Shugar, D. (2007) Experimental (13C NMR) and theoretical (ab initio molecular orbital calculations) studies on the prototropic tautomerism of benzotriazole and some derivatives symmetrically substituted on the benzene ring. *J. Phys. Chem. A* **111**, 6501-6509
32. Wasik, R., Lebska, M., Felczak, K., Poznanski, J., and Shugar, D. (2010) Relative role of halogen bonds and hydrophobic interactions in inhibition of human protein kinase CK2alpha by tetrabromobenzotriazole and some C(5)-substituted analogues. *J. Phys. Chem. B* **114**, 10601-10611
33. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a

random array of atoms. Circumventing problems associated with folding. *FEBS Lett* **239**, 129-136

34. Nilges, M., Kuszewski, J., and Brunger, A. T. (1991) *Computational Aspects of the Study of Biological Macromolecules by NMR*, J.C. Hoch, ed., New York: Plenum Press
35. Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., and Karplus, K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114-122
36. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comput. Chem.* **19**, 1639-1662
37. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012
38. Saparbaev, M., and Laval, J. (1998) 3,N^4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for *Escherichia coli* double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase. *Proc. Natl. Acad. Sci. U S A* **95**, 8508-8513
39. Jurado, J., Maciejewska, A., Krwawicz, J., Laval, J., and Saparbaev, M. K. (2004) Role of mismatch-specific uracil-DNA glycosylase in repair of 3,N^4-ethenocytosine in vivo. *DNA Repair (Amst)* **3**, 1579-1590
40. O'Neill, R. J., Vorob'eva, O. V., Shahbakhti, H., Zmuda, E., Bhagwat, A. S., and Baldwin, G. S. (2003) Mismatch uracil glycosylase from Escherichia coli: a general mismatch or a specific DNA glycosylase? *J. Biol. Chem.* **278**, 20526-20532
41. Saparbaev, M., Kleibl, K., and Laval, J. (1995) Escherichia coli, Saccharomyces cerevisiae, rat and human 3-methyladenine DNA glycosylases repair 1,N6-ethenoadenine when present in DNA. *Nucleic Acids Res.* **23**, 3750-3755
42. Chen, H., Laurent, S., Bedu, S., Ziarelli, F., Chen, H. L., Cheng, Y., Zhang, C. C., and Peng, L. (2006) Studying the signaling role of 2-oxoglutaric acid using analogs that mimic the ketone and ketal forms of 2-oxoglutaric acid. *Chemistry & Biology* **13**, 849-856
43. Ho, R. Y., Mehn, M. P., Hegg, E. L., Liu, A., Ryle, M. J., Hausinger, R. P., and Que, L., Jr. (2001) Resonance Raman studies of the iron(II)--alpha-keto acid chromophore in model and enzyme complexes. *J. Am. Chem. Soc.* **123**, 5022-5029
44. Kim, M. Y., Zhou, X., Delaney, J. C., Taghizadeh, K., Dedon, P. C., Essigmann, J. M., and Wogan, G. N. (2007) AlkB influences the chloroacetaldehyde-induced mutation spectra and toxicity in the pSP189 *supF* shuttle vector. *Chem. Res. Toxicol.* **20**, 1075-1083
45. Leiros, I., Nabong, M. P., Grosvik, K., Ringvoll, J., Haugland, G. T., Uldal, L., Reite, K., Olsbu, I. K., Knaevelsrud, I., Moe, E., Andersen, O. A., Birkeland, N. K., Ruoff, P., Klungland, A., and Bjelland, S. (2007) Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *EMBO J* **26**, 2206-2217
46. Gozdek, A., Stankiewicz-Drogon, A., Poznanski, J., and Boguszewska-Chachulska, A. M. (2008) Circular dichroism analysis for multidomain proteins: studies of the irreversible unfolding of Hepatitis C virus helicase. *Acta Biochim. Pol.* **55**, 57-66
47. Poznanski, J., Wszelaka-Rylik, M., and Zielenkiewicz, W. (2004) Concentration dependencies of NaCl salting of lysozyme by calorimetric methods. *Thermochim. Acta* **409**, 25-32
48. Zielenkiewicz, W., Marcinowicz, A., Poznanski, J., Cherenok, S., and Kalchenko, V. (2005) Complexation of isoleucine by phosphorylated calix 4 arene in methanol followed by calorimetry, NMR and UV-VIS spectroscopies, and molecular modeling methods. *J. Mol. Liq.* **121**, 8-14
49. Smith, R. A., Williamson, D. S., and Cohen, S. M. (1989) Identification of 3,N4-propanodeoxycytidine 5'-monophosphate formed by the reaction of acrolein with deoxycytidine 5'-monophosphate. *Chem. Res. Toxicol.* **2**, 267-271
50. Westbye, M. P., Feyzi, E., Aas, P. A., Vagbo, C. B., Talstad, V. A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N. B., Slupphaug, G., Otterlei, M., and Krokan, H. E.

(2008) Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* **283**, 25046-25056
51. Ringvoll, J., Moen, M. N., Nordstrand, L. M., Meira, L. B., Pang, B., Bekkelund, A., Dedon, P. C., Bjelland, S., Samson, L. D., Falnes, P. O., and Klungland, A. (2008) AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Res.* **68**, 4142-4149
52. Nair, U., Bartsch, H., and Nair, J. (2007) Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. *Free Radic. Biol. Med.* **43**, 1109-1120

Acknowledgements We would like to thank Jacek Olędzki, M.Sc, of the Laboratory of Mass Spectrometry IBB PAS for the MS analyses and Katarzyna Ruszczyńska Ph.D., of Laboratory of Biological NMR IBB PAS for assistance in setting NMR experiments.

FOOTNOTES

*This work was supported by the Polish-Norwegian Research Fund [PNRF-143-AI-1/07] and partially by the National Science Centre grant [UMO-2012/04/M/NZ1/00068]. Molecular modeling performed by J.P. was partially supported by grant CRP/08/011 founded by the International Centre for Genetic Engineering and Biotechnology, Trieste, Italy.

[S]This article contains supplemental methods, Tables S1 and S2 and Figs S1-S8

[1]To whom correspondence may be addressed: Institute of Biochemistry and Biophysics PAS, Pawinskiego 5A street, 02-106 Warsaw, Poland. Tel.: +48 22 5923512; Fax: +48 22 5922190; E-mail: agniesza@ibb.waw.pl

[2]The abbreviations used are: ACR, acrolein; CAA, chloroacetaldehyde; αKG, alpha-ketoglutarate; ε, etheno, see Fig. 1 for structures and abbreviations of adducts; MMS, methyl methanesulphonate; MD, molecular dynamics; SA, simulated annealing;

FIGURE LEGENDS

FIGURE 1. Structures of studied adducts to DNA bases. Methylated: 3-methylcytosine (m^3C) and lipid peroxidation products - exocyclic unsaturated: 3,N^4-ethenocytosine (εC), 1,N^6-ethenoadenine (εA), and exocyclic saturated: 3,N^4-α-hydroxyethanocytosine (HEC) and 3,N^4-α-hydroxypropanocytosine (HPC). Note that Cα of HEC and HPC is chiral.

FIGURE 2. HPLC profiles of HPC repair by AlkB protein. Panel A, substrate TT(HPC)TT pentamer (retention time of 13.7 min.) containing 12% of unmodified TTCTT (retention time of 11.9 min.); panel B, 4-min reaction, 98% of repair. AlkB reaction condition: 50mM HEPES buffer pH of 7.5; 1mM dithiothreitol; 100 μM $Fe(NH_4)_2(SO_4)_2$; 50 μM αKG; 500 pmole of TT(HPC)TT substrate and 50 pmole of AlkB protein. HPLC peak occurring at 9.6 min. represents dithiothreitol.

FIGURE 3. Interdependence between optimal concentrations of H^+ and Fe(II) cations for repair of various adducts by AlkB protein.

FIGURE 4. Pseudo first order kinetics of AlkB repair. The kinetics of the repair processes for εA (A) and HPC (B). The linearity observed for εA repair identifies, in accordance with equation 1a, pseudo first order reaction. In the case of HPC two evident asymptotes clearly identify two species visibly differing in their repair rates.

FIGURE 5. Kinetics of HPC stereoisomers mixture repair by various amounts of AlkB. Marks of a given type represent sets of experimental data and solid lines follow the single-substrate model (A) and two-substrates model (B) fitted to the same data. The visibly better agreement was obtained for the two-substrates model (see Table 1 for details) pointing to the substantial difference in repair rates of both stereoisomers (see Figure 8 for molecular basis of that effect). Error bars represent standard

deviations estimated for those experiments that were repeated three times. The figure shows initial 30 minutes of reactions. See Supplementary Fig. S2 for full course of the reactions.

FIGURE 6. Proposed structures of AlkB complexed with εA and εC. The coordinates for neutral and charged substrates were obtained after 20 rounds of Simulated Annealing procedure started from the accessible structure of m^1A bound to AlkB (PDB3I2O) (27). Asp135, postulated to be crucial for recognition of substrates in their cationic forms, is shown in stick representation, and iron atom is marked by pink sphere of the appropriate radius.

FIGURE 7. εA location in AlkB active centre. A, C - ribbon model, B, D - protein surface colored, according to electrostatic potential distribution, from red (-250 kJ/mol) to blue (+250 kJ/mol).

FIGURE 8. Stereo view of proposed structures of AlkB complexed with S (A, C) and R (B, D) stereoisomers of charged HEC (A, B) and HPC (C, D) substrates in *anti* conformation. Results were obtained from 15ns Molecular Dynamics simulations. See Supplementary Fig. S8 for proposed structures of complexes with those ligands in *syn* conformation.

TABLE 1. Kinetic parameters repair by AlkB of m^3C and exocyclic DNA adducts at physiological pH of 7.5 and pH optimal for repair of each substrate.

Substrate	pK_a	pH	Cationic form [%]	k [$nmol^{-1}s^{-1}$]	RMSD (d.f.) [pmol]	Optimal Fe(II) [μM]	Optimal αKG [μM]
m^3C	8.7	7.5	94	$3.1 \pm 0.3^{\dagger}$ $2.7 \pm 0.4^{*}$	19 (38) -- (7)	100	-
		8.0	83	$1.8 \pm 0.6^{\dagger}$	39 (8)	50	100
HPC	8.6	7.5	93	$0.21 \pm 0.02^{\dagger}$	30 (33)	100	50
				k_1 $0.60 \pm 0.10^{\ddagger}$ k_2 $0.07 \pm 0.01^{\ddagger}$	15 (32)		
HEC	6.9	7.5	20	$0.18 \pm 0.16^{\dagger}$	28 (9)	-	-
		5.8	93	$0.24 \pm 0.04^{\dagger}$	24 (15)	1000	25
				k_1 $0.75 \pm 0.27^{\ddagger}$ k_2 $0.13 \pm 0.02^{\ddagger}$	20 (14)		
εA	3.9	7.5	< 0.1	$0.051 \pm 0.006^{\dagger}$	6 (9)	-	-
		4.6	17	$0.54 \pm 0.10^{\dagger}$ $0.56 \pm 0.02^{*}$	19 (15) -- (7)	3000	50
εC	3.7	7.5	< 0.1	$\sim 0.001^{\dagger}$	5 (7)	-	-
		4.6	11	$0.13 \pm 0.01^{\dagger}$	7 (27)	5000	50

† estimated according to single-substrate model (equation 2)

‡ estimated according to two-substrates model (equation 3)

*estimated for a restricted set of data according to pseudo first order model (equation 1a)

RMSD – root of the mean squared deviation of the fitted function, d.f. –number of degrees of freedom.

-- not applicable

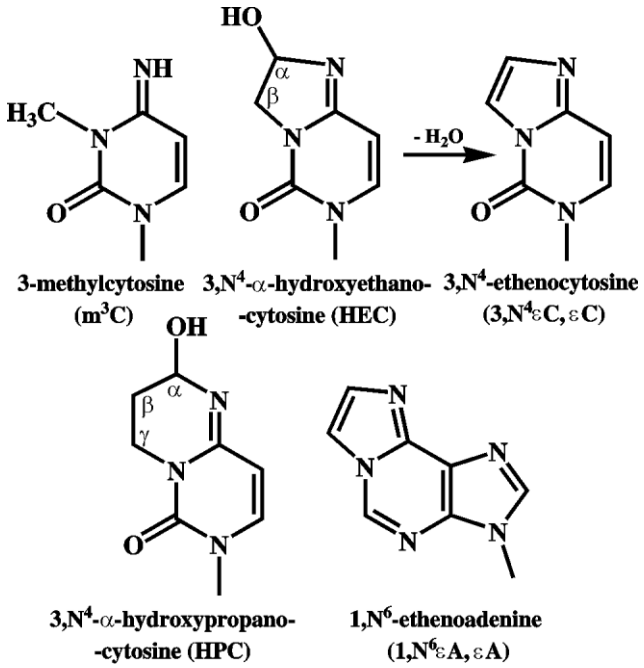
HO
NH
H3C
α
β
- H2O
3-methylcytosine
(m3C)
3,N4-α-hydroxyethano-
-cytosine (HEC)
3,N4-ethenocytosine
(3,N4εC, εC)
OH
α
β
γ
3,N4-α-hydroxypropano-
-cytosine (HPC)
1,N6-ethenoadenine
(1,N6εA, εA)

FIGURE 1.

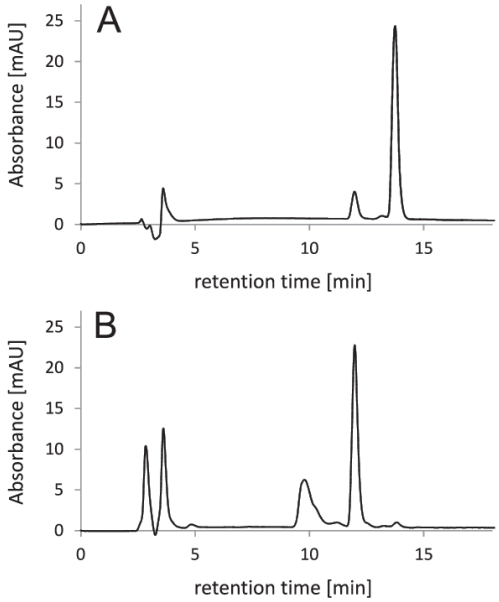
A
Absorbance [mAU]
25
20
15
10
5
0
0
5
10
15
retention time [min]
B
Absorbance [mAU]
25
20
15
10
5
0
0
5
10
15
retention time [min]

FIGURE 2

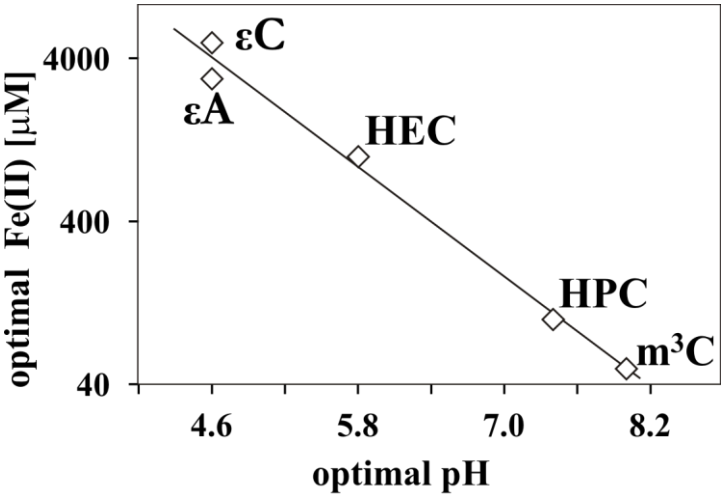
εC
εA
HEC
HPC
m3C
4000
400
40
optimal Fe(II) [μM]
4.6
5.8
7.0
8.2
optimal pH

FIGURE 3.

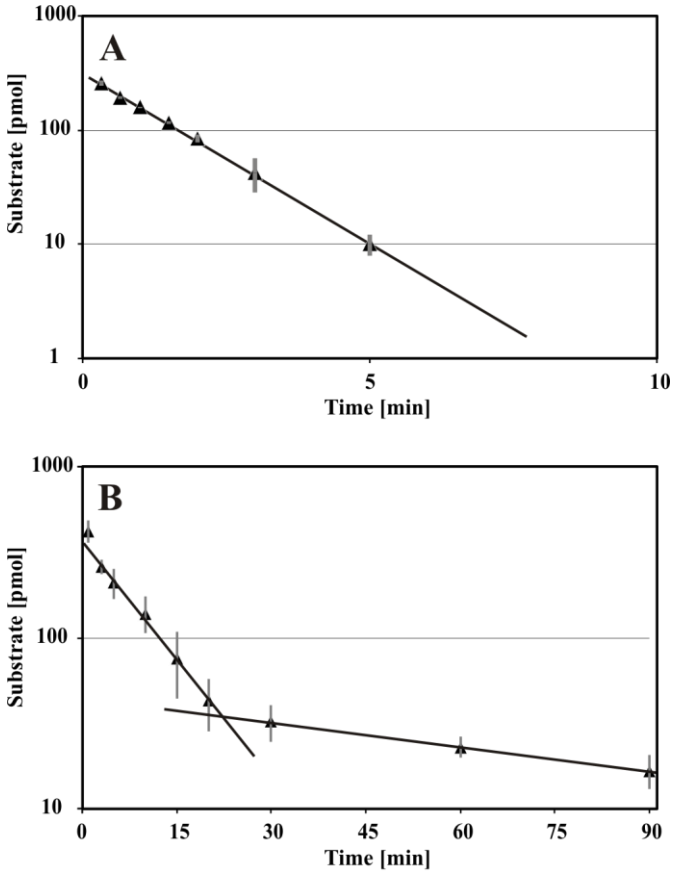
A
1000
100
10
1
Substrate [pmol]
0
5
10
Time [min]
B
1000
100
10
Substrate [pmol]
0
15
30
45
60
75
90
Time [min]

FIGURE 4

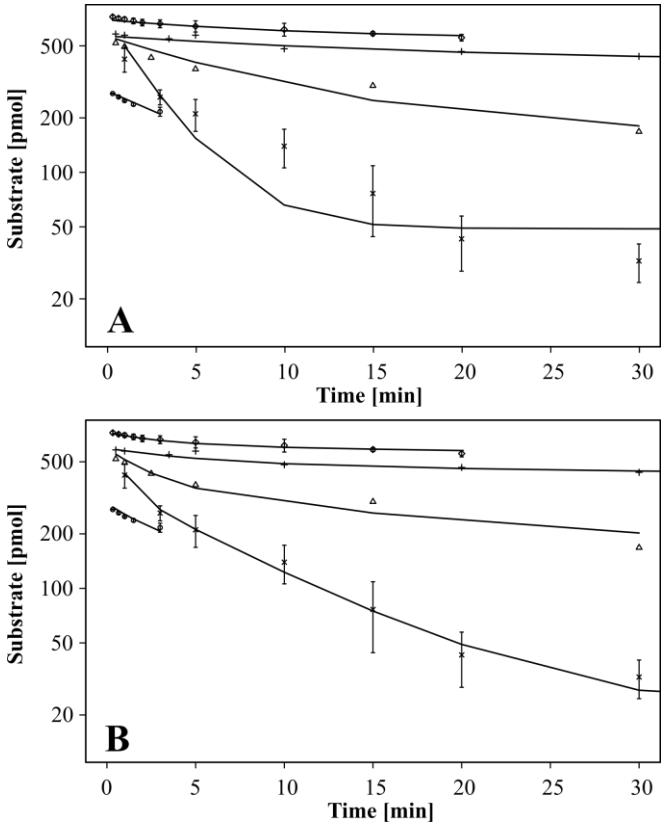
Substrate [pmol]
500
200
100
50
20
A
0
5
10
15
20
25
30
Time [min]
Substrate [pmol]
500
200
100
50
20
B
0
5
10
15
20
25
30
Time [min]

FIGURE 5

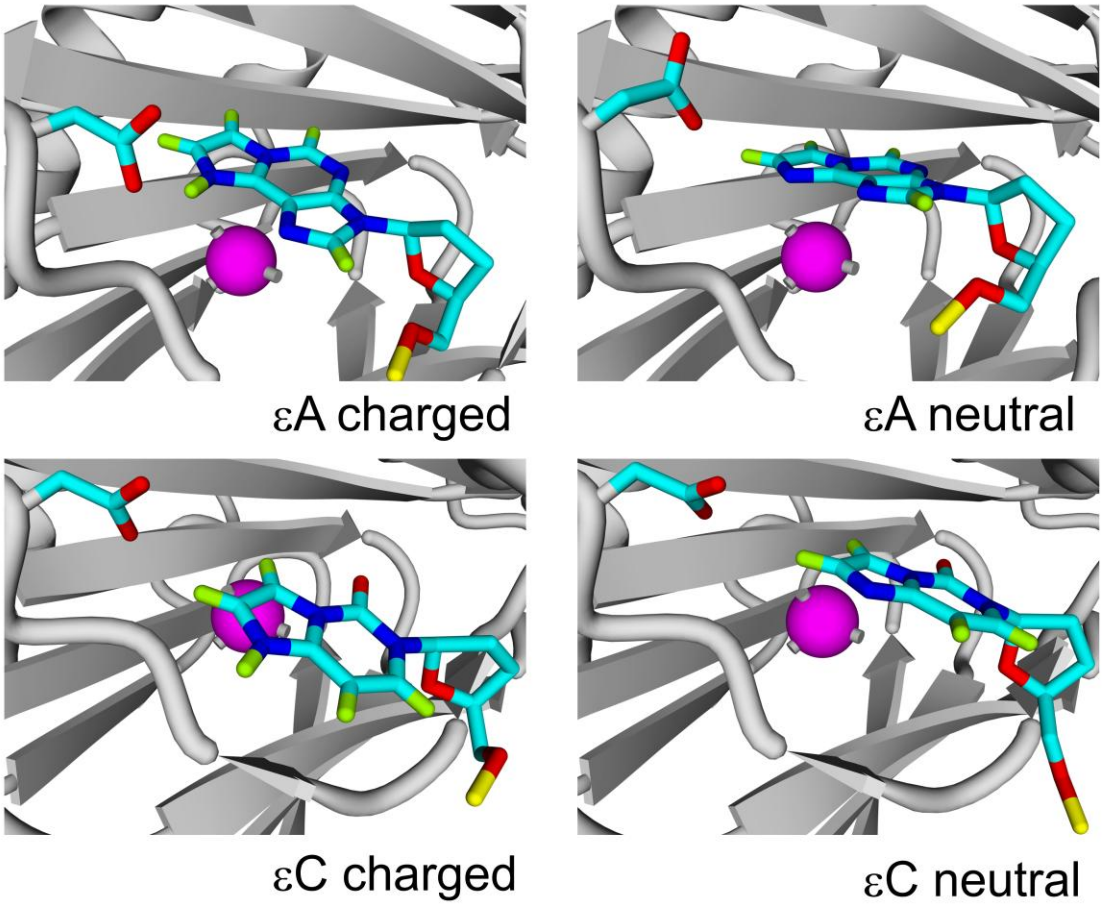

FIGURE 6

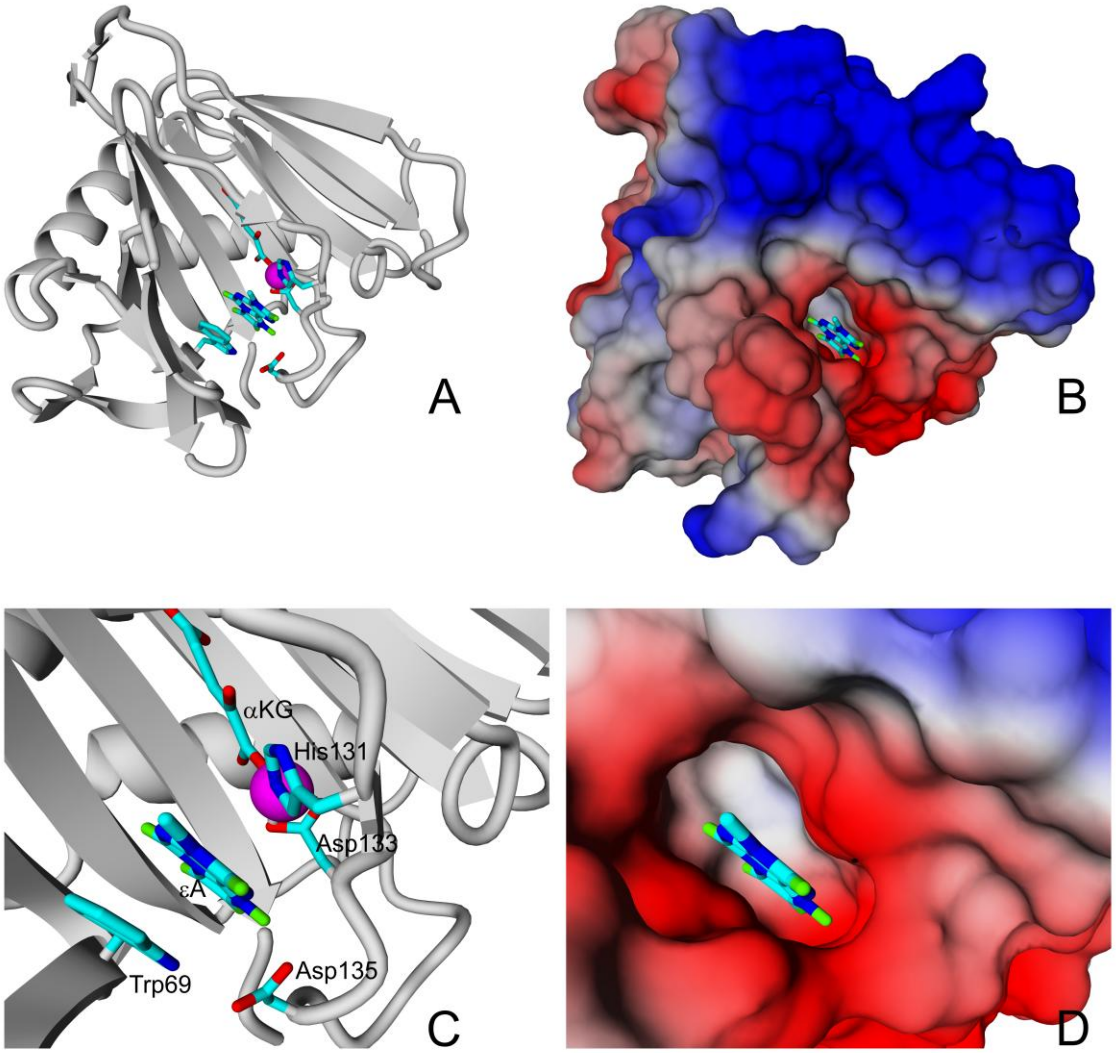

FIGURE 7

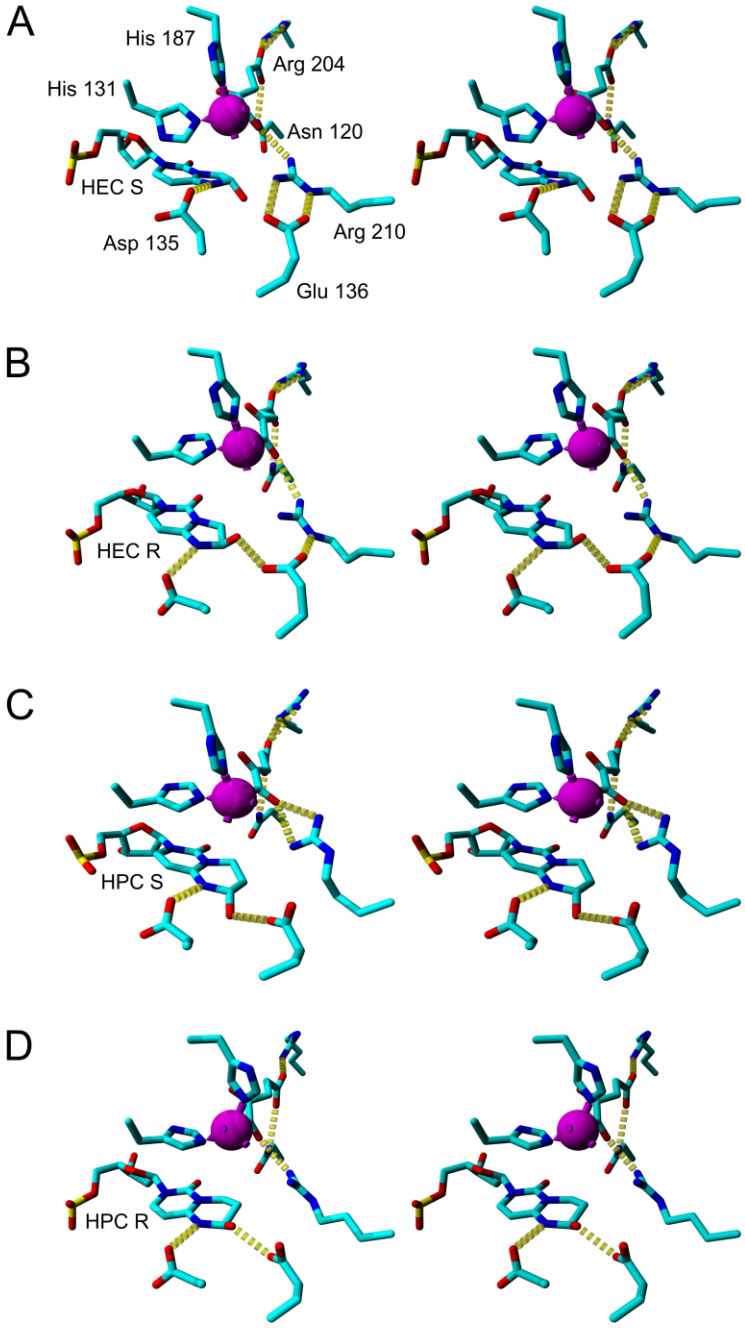
A
His 187
Arg 204
His 131
Asn 120
HEC S
Arg 210
Asp 135
Glu 136
B
HEC R
C
HPC S
D
HPC R

FIGURE 8