BIOCHEMISTRY

 Volume 47, Number 25 June 24, 2008

Articles

The Thymine−DNA Glycosylase Regulatory Domain: Residual Structure and DNA Binding[†]

Caroline Smet-Nocca,[‡] Jean-Michel Wieruszeski,[§] Vicky Chaar,[‡] Arnaud Leroy,[§] and Arndt Benecke*[,‡,||]

Institut de Recherche Interdisciplinaire, USR CNRS 3078, Université de Lille 1, 1 rue du Professeur Calmette, 59021 Lille Cedex, France, Unité de Glycobiologie Structurale et Fonctionnelle, UMR CNRS 8576, Université de Lille 1, Villeneuve d'Ascq, France, and Institut des Hautes Etudes Scientifiques, 35 route de Chartres, 91440 Bures-sur-Yvette, France

Received November 8, 2007; Revised Manuscript Received April 6, 2008

ABSTRACT: Thymine−DNA glycosylases (TDGs) initiate base excision repair by debasification of the erroneous thymine or uracil nucleotide in G·T and G·U mispairs which arise at high frequency through spontaneous or enzymatic deamination of methylcytosine and cytosine, respectively. Human TDG has furthermore been shown to have a functional role in transcription and epigenetic regulation through the interaction with transcription factors from the nuclear receptor superfamily, transcriptional coregulators, and a DNA methyltransferase. The TDG N-terminus encodes regulatory functions, as it assures both G·T versus G·U specificity and contains the sites for interaction and posttranslational modification by transcription-related activities. While the molecular function of the evolutionarily conserved central catalytic domain of TDG in base excision repair has been elucidated by determination of its three-dimensional structure, the mechanisms by which the N-terminus exerts its regulatory roles, as well as the function of TDG in transcription regulation, remain to be understood. We describe here the residual structure of the TDG N-terminus in both contexts of the isolated domain and the entire protein. These studies lead to the characterization of a small structural domain in the TDG N-terminal region preceding the catalytic core and coinciding with the region of functional regulation of TDG's activities. This regulatory domain exhibits a small degree of organization and is implicated in dynamic molecular interactions with the catalytic domain and nonselective interactions with double-stranded DNA, providing a molecular explanation for the evolutionarily acquired G·T mismatch processing activity of TDG.

Base excision repair (BER)[1] pathways are crucial to maintain genomic integrity in face of constant physical and chemical DNA damage with mutagenic potential. Several enzymes with high substrate specificity are required for this type of DNA repair. Among them, uracil−DNA glycosylases (UDGs), which are highly conserved during evolution,

[†] This work has been funded by the Centre National de la Recherche Scientifique (CNRS), the Région Nord-Pas de Calais, the Institut de Recherche Interdisciplinaire, the Institut des Hautes Études Scientifiques, the European Hematology Association−José Carreras Foundation, and the Agence Nationale de la Recherche. The 600 and 800 MHz facility used in this study was funded by the Région Nord-Pas de Calais (France), the CNRS, the University of Lille 1 and the Institut Pasteur de Lille.

* To whom correspondence should be addressed: tel, +33 (0)1 60 92 66 65; fax, +33 (0)1 60 92 66 09; e-mail, arndt@ihes.fr.

[‡] Institut de Recherche Interdisciplinaire.

[§] Unité de Glycobiologie Structurale et Fonctionnelle.

[||] Institut des Hautes Etudes Scientifiques.

[1] Abbreviations: AP, apurinic/apyrimidic; APE1, A/P-endonuclease 1; BER, base excision repair; CAT, TDG catalytic domain; CBP, Creb (cAMP-responsive element binding protein) binding protein; Dnmt3a, DNA methyltransferase 3a; dsDNA, double-stranded DNA; DQF-COSY, double-quantum-filtered correlated spectroscopy; HSQC, heteronuclear single-quantum coherence spectroscopy; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; MBD4/MIG, methyl-binding domain/mismatch glycosylase; MUG, mismatch-specific uracil−DNA glycosylase; NOESY, nuclear Overhauser exchange spectroscopy; RD, TDG N-terminal regulatory domain; SRC-1, steroid receptor coactivator 1; RP-HPLC, reverse-phase high-pressure liquid chromatography; SSP, secondary structure propensity; SUMO, small ubiquitin-like modifier; TDG, thymine−DNA glycosylase; TOCSY, total correlation spectroscopy; TROSY, transverse relaxation-optimized spectroscopy; UDG, uracil−DNA glycosylase.

10.1021/bi7022283 CCC: $40.75
Published on Web 05/31/2008

specifically recognize and initiate removal of uracil bases in DNA (*1–3*). Uracil occurs in DNA principally because of cytosine deamination (*4*). Like every base carrying an exocyclic group, cytosine is sensitive to spontaneous hydrolytic reactions resulting in G•U mismatches. Similarly, methylated cytosine (meC) undergoes spontaneous or enzymatic deamination resulting in G•T mismatches in DNA at an even higher rate than cytosine (*5*). This transition bears even more potential havoc for the cell as not only a G•T mismatch carries an ambiguity as to the original sequence but also epigenetic information stored in the meC base might be lost. Indeed, G•meC to A•T transitions lead to a global loss of CpG islands and modification of DNA methylation profiles found to be associated with some cancers and genetic diseases (*6*). Thymine−DNA glycosylases (TDGs), proteins of the *Escherichia coli* MUG family (*7, 8*), have the ability to detect and repair thymine bases in DNA in the context of G•T mismatches in addition to a G•U repair activity. The human TDG (*9, 10*) and MBD4/MIG (*11, 12*) proteins thereby display lower G•T than G•U repair activity, which indicates, in conjunction with their high homology to UDGs, an adaptive evolution on the structural and functional levels to satisfy the increasing physiological need for G•T repair following the emergence of epigenetic regulation through cytosine methylation.

Interestingly, the acquired thymine specificity in G•T mismatches is encoded by the TDG N-termini which are either not present or highly different from prokaryotic MUG and UDG enzymes (*13*). Unlike the catalytic domains that exhibit an evolutionarily conserved sequence and fold (*14–18*), the N- and C-termini of TDG are more variable and thought to be unfolded. It has been proposed that the low structural definition of the TDG N-terminus confers a higher adaptability and consequently a larger range of recognized substrates. In particular, it has been suggested that the TDG N-termini could directly interact with DNA in a non-sequence-specific manner while scanning for G•T and G•U mismatches, thereby reinforcing DNA binding to allow processing of less favorable substrates at the expense of enzymatic turnover (*19, 20*). This hypothesis is, however, yet unproven but supported by the observation that TDGs also remain tightly associated with the abasic site produced by the glycosylase reaction (*7, 14*). Moreover, this strong binding to AP sites also indicates that human TDG might have other related cellular activities. In this respect the fact that TDG functionally interacts with the retinoic acid (*21*) and estrogen (*22*) receptor transcription factors, as well as with the transcription coregulators SRC-1 (*23*) and CBP/p300 (*24*), is particularly interesting. Indeed, TDG has been shown to positively stimulate transcription activity of these molecules. The mechanism by which TDG exerts its functional role in transcription regulation is today, however, not understood, but again preliminary evidence seems to point toward stabilizing protein−DNA complexes (*21*). Another indicator of the second role of the TDG N-terminus in transcription regulation is the fact that the basic lysine-rich N-terminal domain of TDG is specifically targeted by the universal transcription coregulator CREB-binding protein (CBP). CBP and the highly related p300 protein are capable of specifically binding and, via their histone acetyltransferase activities, acetylating the TDG N-terminus at four distinct lysine residues, thereby inhibiting the CBP/TDG interaction (*24*). Moreover, CBP-dependent acetylation of the TDG N-terminus downregulates the replacement of the cytotoxic abasic site that occurs as the second step in the BER process through the recruitment of the endonuclease APE1 (*24*). This latter observation supports a model in which not only TDG has distinct activities in BER and transcription regulation but also how those activities could be coupled at sites of G•T mispairs and hence sites that had previously undergone epigenetic regulation through cytosine methylation. No direct evidence highlighting such a coupling between epigenetic regulation of transcription and BER was available until recently, when functional interactions between TDG and Dnmt3a were described linking BER with DNA methylation. Dnmt3a was found to enhance TDG glycosylase activity on G•T mismatches; conversely, TDG downregulates Dnmt3a methyltransferase activity, potentially preventing erroneous remethylation of cytosines nearby the restored site. Remarkably, the TDG N-terminus was implicated in Dnmt3a binding as well as parts of the catalytic domain (*25*).

Another posttranslational enzymatic modification of TDG has recently been described and characterized. The SUMO conjugation at a unique K330 consensus site located within the C-terminal region of TDG is involved in a structural modification of the nearby active site (*14, 15*) and seems to have a remote control on the N-terminal conformation and function (*19*). Interestingly, TDG sumoylation has also direct impact on the enzymatic activity by selectively preventing G•T mismatch repair (*26*). The structural rearrangement of the active site upon sumoylation was also found to be associated with an increase in enzymatic turnover during G•U repair, probably by facilitating TDG's dissociation from the abasic site (*14, 26*). Furthermore, since SUMO conjugation, similarly to N-terminal truncation, strongly decreases DNA binding, it has been proposed that the long-range conformational change of the N-terminus upon sumoylation at the C-terminus could modulate the structural and functional properties of the N-terminus with respect to DNA interaction and its influence on enzymatic turnover (*19, 20, 26*). Like many sumoylated proteins TDG can engage in noncovalent interactions with SUMO through two SUMO binding motifs located in the enzymatic domain (*14, 27*). The SUMO-binding capacity could promote self-sumoylation of TDG and/or TDG binding to other sumoylated proteins (*28*). Once more, the N-terminus of TDG interferes with this process through a downregulation of TDG/SUMO-1 interactions (*27*). Interestingly, SUMO conjugation inhibits interactions with CBP whereas it enhances the stimulating effect of APE1 on TDG function in BER (*26*). Therefore, N-terminal acetylation and C-terminal sumoylation could compete with each other to regulate TDG function during DNA repair and transcription.

Taken together, these findings point to a fundamental role of the N-terminus as a key regulator of TDG's different cellular activities. The N-terminus thereby not only serves as a dynamic protein interacting surface but also undergoes reversible enzymatic modifications. These dynamic, context-dependent processes have a determining effect on TDG localization and activity. However, without exception, for none of the different functional roles that can so far be associated to the TDG N-terminus molecular details are known. This is certainly in part caused by the fact that the low structural complexity and possible conformational dynamics of this part of TDG have so far eluded structural

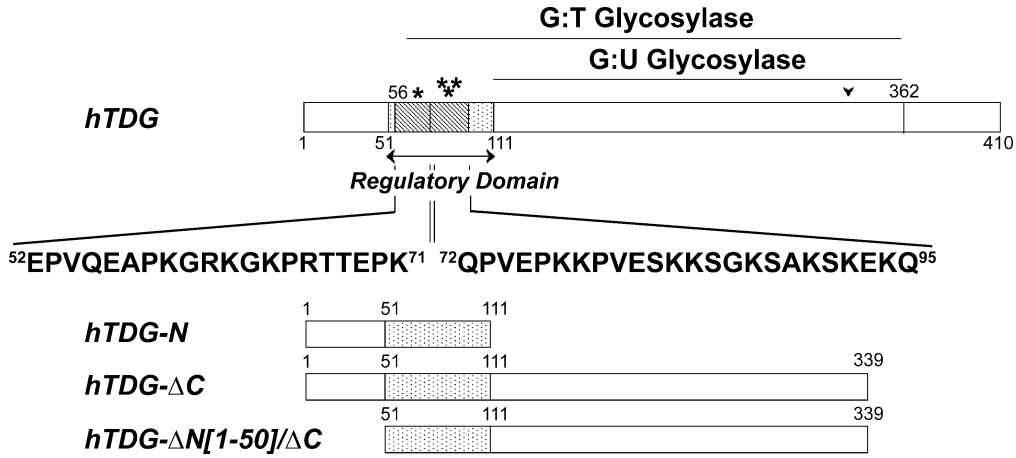

FIGURE 1: Schematic representation of hTDG, hTDG domains, and hTDG peptides involved in this study. The hTDG regulatory domain is indicated by a dotted region in various hTDG constructs. TDG[52−71] and TDG[72−95] peptides are represented by gray-shaded zones encompassing corresponding regions of the regulatory domain, and their primary sequences are indicated below. CBP-mediated acetylation sites are indicated by asterisks, and the sumoylation site is indicated by an arrowhead.

determination. We present here a detailed NMR study of the TDG N-terminus in the context of the entire protein as well as isolated. We have observed that a region encompassing 60 residues immediately preceding the catalytic domain and specifically required for G•T repair undergoes a dramatic dynamic change in the context of the entire protein. Even though the isolated N-terminus exhibits an overall extended conformation, this same region still adopts a small degree of organization and is able to bind small double-stranded DNA molecules in a sequence-independent manner. In contrast, region 1−50 of the N-terminus adopts a totally disordered structure. This region is neither affected by the absence of the catalytic domain nor interacts with DNA, but its deletion as well as the deletion of the full C-terminus leads, as suggested by previous works, to a significant increase of TDG glycosylase activity.

MATERIALS AND METHODS

Peptide Synthesis. Peptides corresponding to residues 52−71 and 72−95 of hTDG protein (see Figure 1 for a description of the primary sequences) were synthesized by solid-phase chemistry (on Rink amide resin; novabiochem) with HOBt/TBTU activation for coupling of protected amino acids, followed by TFA cleavage. Peptide mixtures were purified by reverse-phase chromatography on a C18 Zorbax 300-SB (5μm) column equilibrated in a 0.05% aqueous TFA buffer (buffer A). Separation of peptides in the crude mixture was carried out using a 60 min linear gradient of 0−40% acetonitrile in buffer A at a flow rate of 4 mL/min. Homogeneous fractions, as checked by RP-HPLC on a C18 nucleosil column, capillary electrophoresis, and MALDI-TOF mass spectrometry, were pooled and lyophilized to provide peptides of greater than 95% purity.

Expression and Purification of Recombinant TDG and TDG Domains. TDG proteins corresponding to residues 1−410 (full-length TDG), 1−339 (TDG-ΔC), 1−111 (TDG-N), and 51−339 (TDG-ΔN[1−50]/ΔC) (see Figure 1 for a schematic description of TDG domains) were overexpressed in the BL21(DE3) strain as GST fusion proteins. The cDNA of TDG was amplified by PCR using oligonucleotides containing a *Bam*HI site at the 5′-end and a *Eco*RI site at the 3′-end and then cloned between the *Bam*HI/*Eco*RI sites of the pGEX-6P-1 plasmid. Bacteria were grown at 37 °C in M9 minimal medium reconstituted with 2g/L glucose, 1 g/L ^{15}N-labeled ammonium chloride, 1 mM $MgSO_4$, MEM vitamin cocktail (Sigma), and 100 mg/L ampicillin (or in LB medium for the production of unlabeled proteins). Protein expression was performed with overnight IPTG induction (0.5 mM final concentration of IPTG) at 20 °C. Then, cells were harvested and resuspended in extraction buffer (PBS, 10% glycerol, 1% Triton X-100, 10 mM EDTA, 2 mM DTT) complemented with a protease inhibitor cocktail (Complete; Roche). Cell lysates were obtained by incubation of 0.25 mg/mL lysozyme with the cell suspension in extraction buffer complemented with RNase and DNase and followed by a brief sonication step. The soluble extract was isolated by centrifugation. GST fusion proteins were purified on glutathione−Sepharose (GE Healthcare). Soluble extracts were incubated for 3 h at 4 °C with 25−100 μL resin/mL of soluble extracts. Unbound proteins were extensively washed away with a GST wash buffer (PBS, 5% glycerol, 1% Triton X-100, 10 mM EDTA), and proteins were eluted by digestion with Prescission protease 25 μg/mL resin (GE Healthcare) in one bead volume of elution buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2% glycerol, 0.1% NP-40, 10 mM EDTA, 5 mM DTT) at 4 °C for 20 h. Then beads were eluted twice with one bead volume of elution buffer. The pooled fractions were concentrated and purified by gel filtration on a preparative Superdex75 column (GE Healthcare) equilibrated in the NMR sample buffer. Homogeneous fractions checked on a 12% SDS−polyacrylamide gel were pooled and concentrated to obtain final protein concentrations of 100 μM. The molecular mass of proteins and isotopic labeling were verified by MALDI-TOF mass spectrometry.

NMR Spectroscopy. NMR experiments were performed at 293 K on a Bruker DMX 600 MHz spectrometer (Bruker, Karlsruhe, Germany) equipped with a cryogenic triple resonance probe head. All ^{1}H spectra were calibrated with 1 mM sodium 3-trimethylsilyl-3,3′,2,2′-d_4-propionate as a reference. Samples containing 2 mM peptide in aqueous buffer (100 mM $NaiPO_4$ pH 6.6, 0.5 mM EDTA, 5% D_2O) were used for resonance assignment. Standard water-flipback NOESY and TOCSY experiments, with 400 and 69 ms mixing times, respectively, were recorded for peptide as-

signment. DQF-COSY experiments were acquired with 8K points in the detection dimension to calculate the $^3J_{H_N}-H\alpha$ coupling constants. $^1H-^{15}N$ HSQC spectra at nitrogen natural abundance were recorded with 256 scans per increment and 256 points in the nitrogen dimension.

Spectra on ^{15}N-labeled proteins were recorded on 100 μM samples in the same buffer as for peptides. Conventional assignment strategies using triple resonance experiments [HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCANNH] have been performed at 600 MHz on uniformly $^{13}C/^{15}N$-labeled proteins to correlate the Cα, Cβ, or CO of a given residue with that of its next neighbor. Spectral windows were 16 ppm for proton and 36.54 ppm for nitrogen, centered at 4.7 and 118 ppm, respectively, and sampled with 1024 and 104 complex points. All spectra were recorded with eight scans per increment. The 33.1 ppm Cα window in the HNCA and HN(CO)CA spectra was centered at 51.7 ppm and was sampled with 96 complex points. Similarly, the 16.6 ppm CO window was centered at 172.5 ppm and was sampled with 48 complex points. Finally, the 66.26 ppm combined Cβ and Cα window in the CBCANH and CBCA(CO)NH spectra was centered at 37.4 ppm and was sampled with 192 complex points. Because of the high degeneracy of carbon resonances in such unfolded proteins, assignments were performed in a pairwise manner previously described (*29–31*). All spectra were processed with 2048 complex points in the proton dimension and 512 in the indirect nitrogen and carbon dimensions. HNHA experiments were acquired on uniformly ^{15}N-labeled proteins at 800 MHz to determine $^3J_{H_N}-H\alpha$ coupling constants with spectral windows of 24 ppm for the detection proton dimension and 8.9 and 36.54 ppm for the indirect proton and nitrogen dimensions and sampled with 2048, 128, and 112 points, respectively.

The chemical shift perturbations of individual resonances were calculated with eq 1, taking into account the relative dispersion of the proton and nitrogen chemical shifts (1 and 20 ppm, respectively).

$$\Delta\delta(\text{ppm}) = \sqrt{[\Delta\delta(^1H)]^2 + 0.05[\Delta\delta(^{15}N)]^2} \quad (1)$$

$^3J_{H_N}-H\alpha$ couplings were derived at $\pm$3 Hz from intensity ratios of $H_N-H\alpha$ cross-peaks (S_{cross}) and corresponding H_N-H_N diagonal peaks (S_{diag}) using eq 2 with a total evolution time 2ζ of 26.1 ms and a scaling factor f of 1 (*32*) for these very flexible protein domains.

$$J^3_{HNH\alpha} = \frac{\arctan\sqrt{-(S_{cross}/S_{diag})}}{2\pi\zeta f} \quad (2)$$

Heteronuclear NOEs were measured with the standard refocused HSQC pulse sequence in the presence or absence of proton decoupling during a 5 s relaxation delay on a 200 μM sample of TDG-N. Hetero-NOE values were derived from the intensity ratios of the cross-peak with and without proton decoupling.

TDG Double-Stranded DNA Binding. Annealing of oligonucleotides was performed by heating 1 mM solutions for 5 min at 100 °C and cooling the mixtures slowly to room temperature to get double-stranded 25-mers containing either a central Watson−Crick G·C pair or G·T/U mispairs. These solutions were lyophilized and dissolved at 80 μM final concentration in a 20 μM solution of ^{15}N-TDG-N in a buffer constituted by 100 mM $NaiPO_4$, pH 6.6, 1 mM DTT, and 1 mM EDTA. The interactions between TDG-RD in the context of the entire protein and a dsDNA substrate were measured on a 20 μM solution of ^{15}N-TDG containing 50 μM dsDNA with a canonical G·C pair in the presence of 5% glycerol.

Glycosylase Activity on G·T/U Mismatches. A modified DNA nicking assay was performed on 25-mer dsDNA containing either G·T or G·U mispairs or a G·C pair as a control. The oligonucleotides were labeled on the primary amine modified 3′-end of the T/U/C strand with the Alexa-Fluor 488 dye (Invitrogen, Molecular Probes). TDG proteins were incubated at 0.5 μM final concentrations with oligonucleotides at 5 μM in 80 μL of nicking buffer (25 mM Hepes−KOH, pH 7.8, 1 mM EDTA, 1 mM DTT) at 37 °C. Twenty microliter aliquots were withdrawn at different incubation times, and oligonucleotides were precipitated in 70% ethanol and 300 mM NaCl (final concentrations) and incubated for 30 min at 50 °C with 0.01 N NaOH solution. Oligonucleotides were then analyzed by denaturing polyacrylamide gel electrophoresis.

RESULTS

NMR Study of the Human TDG. The size of human thymine−DNA glycosylase (hTDG, 410 amino acids, 46 kDa; a schematic of the protein primary sequence is shown in Figure 1) is *a priori* not compatible with structural studies by NMR. A $^1H-^{15}N$ HSQC spectrum of hTDG reveals, however, 100 resonances characterized by relatively narrow lines for a protein of this size (Figure 2A, black resonances). The superposition of the spectrum acquired on the full-length protein with a $^1H-^{15}N$ HSQC spectrum on the isolated N-terminus of hTDG (hTDG-N, amino acids 1−111, Figure 1) shows that the resonances observed with hTDG correspond in part to residues located in hTDG-N (Figure 2A, red resonances). Using a hTDG truncated for the C-terminus (hTDG-ΔC, amino acids 1−339, Figure 1), we can show that the remainder of the observed resonances in the hTDG spectra, not accounted for by the N-terminal amino acids 1−111, corresponds to residues located in the C-terminal part of the protein as they disappear using hTDG-ΔC (Figure 2A, blue resonances). Both N- and C-termini of hTDG, according to the poor dispersion of ^1H_N resonances (around 1 ppm), seem to adopt little to no tertiary structure. In contrast to the N- and C-terminus, the central catalytic domain of hTDG (amino acids 112−326) is completely invisible from the $^1H-^{15}N$ HSQC spectra most likely due to the conserved globular structure it adopts (*14, 15*). Neither the deuteration of nonexchangeable protons of TDG full-length protein nor the use of a $^1H-^{15}N$ TROSY pulse sequence leads to a significant improvement of resonance detection. Only a small refinement of some resonance line widths was observed, leading to detection of few new resonances (data not shown). These additional resonances observable on the $^{15}N-^1H$ TROSY spectrum of the $^{15}N,^2H$-labeled TDG could be those of the RD but shifted as compared to the RD resonances of TDG-N. Unfortunately, their intensities are not sufficient to envisage their assignment using the triple resonance experiments. However, it must be noticed that these new resonances appear in the same limited range as the N- and C-terminal domain chemical shifts.

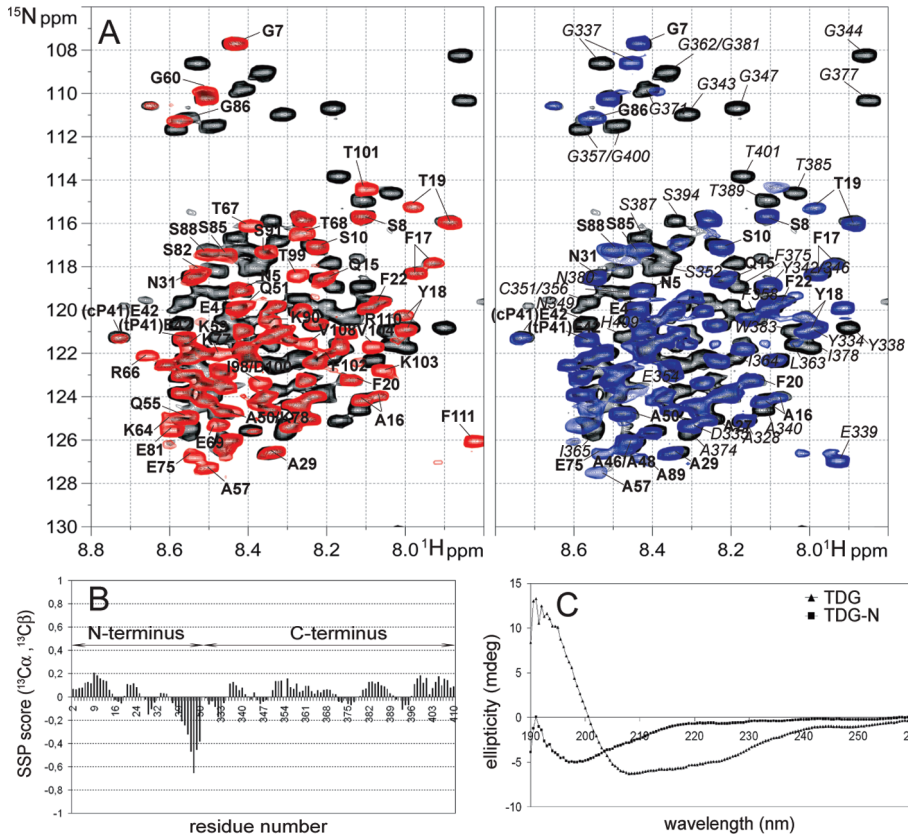

FIGURE 2: (A) ^{1}H−^{15}N HSQC spectra of the full-length hTDG (black), hTDG-ΔC (blue), and hTDG-N (red). Some resonances corresponding to residues of the N-terminus are annotated in bold and residues of the C-terminus in italic. Resonances of the Glu42 residue related to the Pro41 *cis* and *trans* conformers are annotated with "c" and "t" letters, respectively. (B) Secondary structure propensity (SSP) score (*35*) along the primary sequence of TDG calculated with the ^{13}Cα and ^{13}Cβ values. Positive values indicate a propensity to α-structure and negative values a propensity to β-structure. The extreme N-terminus (residues 1−25) would likely have α-propensities, and the region going from residue 40–50 would preferentially have β-propensities. (C) Circular dichroism spectra of the full-length hTDG (▲) and the isolated N-terminus (■).

hTDG N-Terminal Conformation. Using the conventional triple resonance experiments we have partially assigned the resonances of hTDG and the hTDG N-terminus (Figure 2A). In the entire hTDG protein, we have assessed the extended structure of region 1−50 in the N-terminus and the entire C-terminus (residues 328−410) using two independent sets of data. First, the ^{13}Cα and ^{13}Cβ resonances of both regions fall within 1 ppm range around the random coil chemical shifts, indicating few perturbations of their chemical environment when compared to random coil structures (*33, 34*). We have calculated the secondary structure propensity (SSP) score of the TDG N- and C-terminus based on ^{13}Cα and ^{13}Cβ values (*35*). An overall total of 3.9% α-structure and 11.8% β-structure has been found for the 50 first N-terminal residues of the full-length TDG and a total of 5.8% α-structure and 1.7% β-structure for the C-terminus (residues 328−410). The N-terminus exhibits less than 20% α-structure propensities for the 1−30 region and up to 65% β-structure propensities for the 40−50 residues while the C-terminus has propensity mainly for α-structure with, however, overall low scores (lower than 20%) (Figure 2B). Then, the $^{3}J_{H_N}$−Hα coupling constants of residues 1−50 in hTDG-ΔC and in the isolated N-terminus are very similar and comprised in the range of 5−8 Hz, corresponding to disordered structures (see Supporting Information, Figure S1). Together, these results indicate that both regions are largely unfolded.

Concerning the hTDG N-terminal resonances, we observed two different behaviors when comparing the isolated N-terminus to the entire protein. The resonances for amino acids 1−50 of hTDG remain mostly unchanged in both cases (Figure 2A), whereas the resonances for amino acids 51−111 are significantly broadened in the context of hTDG when compared to hTDG-N (the line widths at half-height are comprised between 21 and 25 Hz for hTDG-N resonances and at least twice higher for hTDG), and most of them fail to be detected. Given the high sensitivity of the chemical shifts to their chemical environment, these observations allow to conclude upon a conformational change in the region encompassing amino acids 51−111 when the N-terminal region is isolated from the remainder of the protein. Furthermore, the line width of these resonances is an indicator of interactions with the catalytic domain in the context of the entire protein. Since the structural domain encompassing amino acids 51−111 coincides with the region of hTDG that had previously been shown in functional studies to exert a regulatory effect upon TDG activity, we shall refer to this domain from here on as the regulatory domain of hTDG (hTDG-RD, Figure 1) (*13, 16, 19, 21, 24, 25*). The conformational dynamics observed in hTDG-RD when comparing hTDG-N to hTDG cannot simply be explained by a border effect of the catalytic domain. It has been previously shown that such border effects have a small effective range of two to three amino acids in the case of a simple primary sequence modification in an unfolded polypeptide without any impact on tertiary structure (*29, 31*). By

contrast, the effect of the catalytic domain on the N-terminus extends over 60 amino acids encompassing the entire regulatory domain. Moreover, this effect is characterized by a variation and/or broadening of the observed resonances, indicating a substantial conformational change and different conformational dynamics. It is also worthwhile noticing that some resonances of the regulatory domain (A57, Q55, E75, S85, G86, S88, A89) that are not detectable in hTDG appear in the hTDG-ΔC HSQC even though they are still broad (Figure 2A). In contrast, there is no change in the resonances of residues 1−50 when comparing the entire hTDG and hTDG-ΔC. These data indicate also a slight modification of the regulatory domain dynamics upon C-terminal truncation, raising the possibility of N- and C-terminal cross-talk.

Interestingly, residues 1−50 of TDG-N adopt the same profile of secondary structure propensities as the full-length TDG. The isolated N-terminus exhibits less than 20% α-structure propensities for the residue 1−25 region and up to 70% β-structure for the 40−86 residues encompassing the regulatory domain (see Supporting Information, Figure S2A). Furthermore, the $^{3}J_{H_N}$−Hα coupling constants calculated from HNHA experiments (*32*) on the uniformly ^{15}N-labeled protein fall within the range of 5−8 Hz corresponding to disordered structures (see Supporting Information, Figure S2B). These observations are accompanied by a low proportion of α-helices and β-sheets detected by circular dichroism in the N-terminus when compared to the full-length protein (Figure 2C). Furthermore, we have investigated the dynamical aspect of the TDG-N backbone with the measurement of heteronuclear NOE. The small hetero-NOE values, comprised between −0.3 and 0.2, suggest an overall high degree of flexibility all along the isolated N-terminus. More particularly, the region from residue 44 to residue 95, encompassing the regulatory domain, has negative ^{15}N−{^{1}H} NOE values while the region comprising residues 9−25 is relatively more rigid with positive ^{15}N−{^{1}H} NOE values (see Supporting Information, Figure S3).

Proline Conformational Heterogeneity in hTDG. In both hTDG and hTDG-N ^{1}H−^{15}N HSQC spectra, a doubling of resonances in the region between positions 15 and 20 is observed (Figure 2A). A *cis/trans* conformational heterogeneity of a proline residue at position 21 could explain this phenomenon, and these two sets of resonances could further reflect a slow exchange between both conformations. When analyzing the ^{13}Cα and ^{13}Cβ resonances of the Pro21 residue (63.47 and 31.48 ppm, respectively) on the ^{13}C strips extracted from the Phe22 ^{1}H−^{15}N resonance in the HNCACB spectrum, we found that the signals of higher intensities correspond to a conformer where Pro21 is in the *trans* conformation. The integration of the corresponding signals reveals that the hTDG conformer with the Pro21 in the *trans* conformation accounts for ∼60% of the total protein. This ratio is unchanged whether the isolated N-terminus or the entire protein is analyzed. The proportion of *cis* isomer is thus relatively important compared to what is generally observed in nonstructured peptides or proteins where *cis* conformers represent up to some 23% of the total molecules when an aromatic residue precedes a proline (*36*).

Similar to the above Pro21 *cis*/*trans* conformers we have also observed in both the isolated N-terminus and the entire hTDG two signals for the glutamate at position 42 which follows a proline. This Pro41 also is present in *cis* (corresponding values: 62.74 ppm for the ^{13}Cα and 34.15 ppm for the ^{13}Cβ) and *trans* (^{13}Cα, 62.93 ppm; ^{13}Cβ, 32.08 ppm) conformations with the *cis* conformer accounting for 20% of the total Pro41 population (*37*).

Protein−Protein Interactions between the Regulatory and Catalytic Domains of TDG. Because of the disappearance or broadening of the regulatory domain resonances, a molecular contact with the catalytic domain exists, resulting in broadened resonances for this region and mimicking the NMR behavior of the TDG catalytic core. Furthermore, these protein−protein interactions are confined to the regulatory domain as the initial 50 amino acids of the TDG N-terminus remain very flexible and conserve identical resonances with narrow lines. In order to prove an interaction between TDG-RD and TDG-CAT, we investigated the capacity of the isolated TDG-N to compete with TDG-RD for the core of the protein. We found that the regulatory domain of TDG undergoes little conformational change in the presence of a 50-fold excess of hTDG-N. Indeed, some resonances of the hTDG-RD in the ^{15}N-labeled full-length protein are detectable but are still less intense than those of the TDG 1−50 region (data not shown). Remarkably, they are found to be identical to those of the isolated N-terminus indicating that (i) the isolated N-terminus is able to only partially displace interactions between the catalytic and the regulatory domains and (ii) the displaced TDG-RD adopts the same conformation as in the isolated TDG N-terminus.

Given that the TDG-RD contains a high proportion of basic residues (i.e., 16 lysine and 5 arginine for a total of 60 residues, that is one-third of the TDG-RD residues), we have searched negative patches at the surface of TDG-CAT that could serve as potential binding sites for the positively charged TDG-RD. On the basis of the three-dimensional structure of TDG-CAT conjugated to SUMO-1 (*14*), we have calculated the electrostatic potential surface of TDG-CAT in a search for a negative patch. A surface enriched in acidic residues (E194, D126, D177, D178, E170) located at the opposite side of the TDG active site would be accessible to the TDG-RD and could potentially mediate electrostatic interactions with TDG-RD (see Supporting Information, Figure S4).

Residual Structure of the hTDG N-Terminus. We have measured, by calculating the secondary structure propensity with ^{13}Cα and ^{13}Cβ values (*35*), an overall content of 2.9% α-structure and 11.7% β-structure within the isolated N-terminus of TDG (see Supporting Information, Figure S2A). Given the propension of TDG-N to adopt β-structure, we decided to compare the corresponding resonances with those of two small peptides produced by solid-phase chemistry encompassing the 52−71 and 72−95 regions of the regulatory domain (Figure 1) which represent a large part of the region having a marked β-structure propensity as determined by the SSP score (see Supporting Information, Figure S2A). The peptide resonances were assigned following the conventional strategy of sequential assignment using the two-dimensional homonuclear experiments TOCSY and NOESY, the DQF-COSY sequence for calculating the $^{3}J_{H_N}$−Hα coupling constants, and the acquisition of ^{1}H−^{15}N HSQC spectra at natural abundance. For these short peptides of about 20 amino acids the absence of structure was easily confirmable using the NOESY spectra where only the H_N−Hα correlations were observable. Furthermore, the

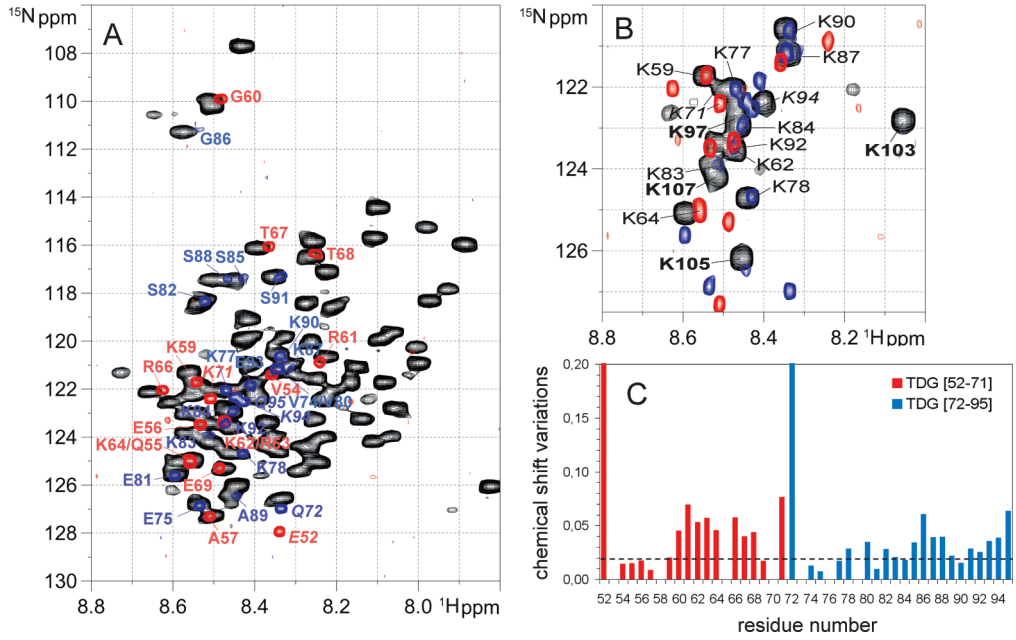

FIGURE 3: Comparison of ^{1}H−^{15}N HSQC spectra of the hTDG-N (black), uniformly ^{15}N-labeled (A) and selectively ^{15}N-Lys-labeled (B) with TDG[52−71] (red) and TDG[72−95] (blue) peptides. (C) Graphical representation of the chemical shift variations between the hTDG N-terminus and the TDG[52−71] (red) or TDG[72−95] (blue) peptide.

coupling constants $^{3}J_{HN}$−Hα measured by DQF-COSY (with 8K points in the detection) are comprised between 6 and 8 Hz, indicating a disordered structure (see Supporting Information, Figure S2B). The comparison of the hTDG-N ^{1}H−^{15}N HSQC spectrum with those obtained for the two peptides hTDG[52−71] and hTDG[72−95] shows small variations in the chemical shifts (Figure 3A,C). The resonances of the lysine residues that lie within an overcrowded region of hTDG-N HSQC were confirmed with the acquisition of a ^{1}H−^{15}N HSQC spectrum on a selectively ^{15}N-Lys-labeled protein (Figure 3B). Strong chemical shift perturbations are observed for the first residues in each peptide primary sequence (Figure 3C). These variations reveal a small border effect consistent with the absence of the flanking residues when dealing with peptides as compared to the entire hTDG-N protein. For other residues, weaker but significant perturbations are observed and indicate that the regulatory domain does not adopt a totally random structure in the context of the isolated TDG N-terminus. It must be noticed, however, that such a residual structure concerning the regions encompassed by residues 60−68 and 78−93 failed to be detected with the chemical shift index (*34*) and $^{3}J_{HN}$−Hα couplings. Both regions have strong β-structure propensity as indicated by their SSP score (*35*) that could explain such a residual structure observed for the isolated TDG N-terminus. Together, these data show that the regulatory domain has a potential to adopt a β-structure in the full TDG context.

In the peptides, the relative proportions of *cis* conformers for prolyl residues match with those of hTDG-N. In the hTDG[52−71] peptide, the *cis* conformers of prolines at the N- and C-termini were found in relatively high quantities (8% for Pro53 and 6% for Pro70) when comparing to those prolines in the middle of the sequence (4% for Pro58 and 3% for Pro65). In the hTDG[72−95] peptide, the Pro73 *cis* conformation is relatively low (2%), and the Pro76 and Pro79 *cis* conformers are undetectable. These *cis* populations are compatible with those generally found in small disordered peptides (*38*).

To further corroborate the hTDG-RD residual structure, we also investigated the conformational stability of the TDG N-terminus in the presence of glycerol. It has been shown that such a polyhydric alcohol used for many years as a protein stabilizer may have a denaturing effect targeting primarily hydrophobic residues in proteins. The recently elucidated mechanism of these effects indicates that polyhydric alcohols do not interact directly with proteins but rather alter the hydration layer around polypeptide chains (*39*). At 5% final concentration of glycerol we observe conformational effects only on resonances of the regulatory domain, the remaining 50 residues of hTDG-N being unaffected albeit containing a higher content of hydrophobic residues (Figure 4B). This effect does not seem to be fortuitous as the regulatory domain is the region contacting the catalytic domain and also displaying a dramatic structural change in the presence of the latter. Surprisingly, with 5% glycerol, the residues of the regulatory domain adopt the same conformation as those of both TDG[52−71] and TDG[72−95] peptides described above as their resonances perfectly match (Figure 4A). This loss of the residual structure of the hTDG-RD is reversible: removing glycerol from the sample by dialysis restores the chemical shifts of the regulatory domain (data not shown). These data demonstrate that hTDG-N retains a residual structure in its regulatory domain located between residues 51 and 111 even when isolated from the remainder of the protein. Similarly, we have investigated the effect of glycerol on the RD conformation in the context of the entire TDG. No significant modifications of RD resonances have been detected, but subtle perturbations of the TDG-RD conformation cannot be excluded (Figure 5D, black spectrum). We suppose that the RD/CAT interactions stabilize the RD structure sufficiently to counteract the effect of glycerol at this concentration.

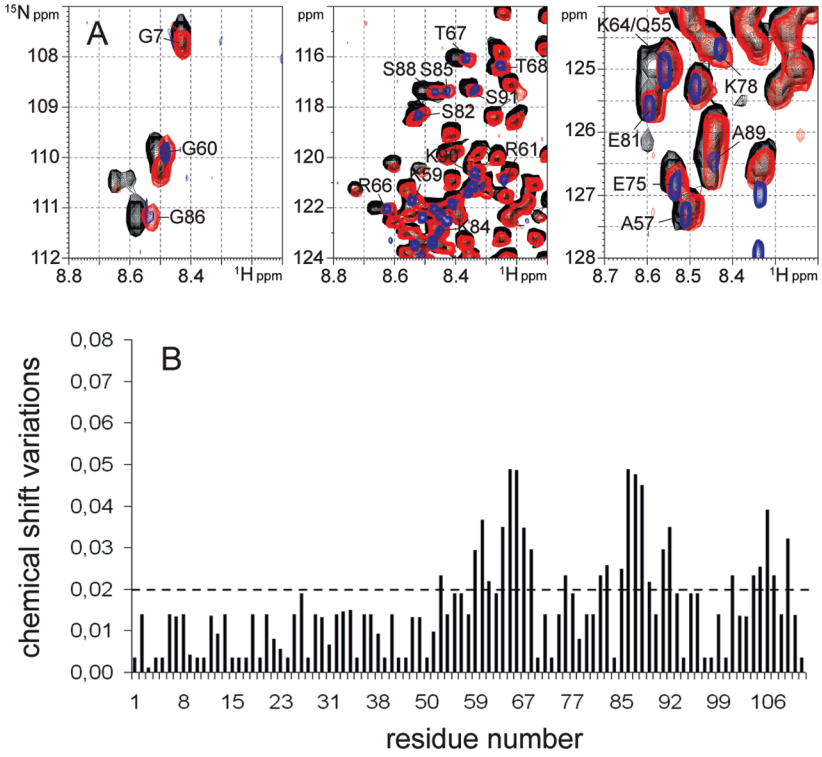

FIGURE 4: Upper panel: Comparison of $^{1}H-^{15}N$ HSQC spectra of the hTDG N-terminus (black), in presence of glycerol 5% (red), and the isolated hTDG peptides (blue). Lower panel: Graphical representation of the chemical shift variations of hTDG-N resonances induced by 5% glycerol.

The TDG-RD Has a Potential To Interact with DNA. In an effort to explain the requirement of the N-terminus of TDG for G•T mismatch processing activity, as well as the evolutionary adaptation of G•U processing enzymes to a broader substrate specificity, it had been suggested that the N-terminus of TDG could directly and specifically interact with DNA (*13, 19, 20*). This interaction was proposed to strengthen the overall DNA binding affinity of TDG in order to allow processing of less favorable substrates at the expense of the enzymatic turnover (*19, 20*). In order to determine whether the N-terminal domain of TDG retains a DNA binding activity outside of the remaining protein, we have investigated the interactions of hTDG-N with a homoduplex DNA or heteroduplexes containing G•T or G•U mispairs. We have found that in the presence of a 4-fold excess of 25-mer dsDNA the resonances corresponding to the hTDG-RD are perturbed while the resonances of the 1−50 N-terminal region remain unchanged (Figure 5A−C). Moreover, we did not observe any significant differences in the hTDG-RD chemical shift perturbations when comparing the three DNA substrates (Figure 5A−C). These chemical shift variations of TDG-RD clearly indicate that this region of TDG has a potential to directly interact with DNA *in vitro* in a sequence-independent manner and with a relatively low affinity even in the absence of TDG-CAT. Furthermore, as one would expect, the regulatory domain of TDG when isolated from its protein context is not able to distinguish between nonmismatched and mismatched dsDNA, nor exerts a mismatch-selective binding activity.

In the context of the full TDG protein, a 2.5-fold excess of dsDNA substrate containing a canonical G•C pair induces a dramatic conformational change of the TDG-RD, leading to the appearance of new resonances corresponding to the RD domain in the TDG $^{15}N-^{1}H$ HSQC spectrum, while the 1−50 region of the N-terminus and the full C-terminus do not exhibit any chemical shift perturbations (Figure 5D,E). Remarkably, these additional resonances perfectly match with those of the isolated TDG-RD interacting with the same substrate even though they are broader (Figure 5D,E). These data indicate a partial dissociation of the TDG-RD from the catalytic domain induced by DNA and indiscriminate with respect to the presence or absence of a G•T/U mismatch. Under these conditions, TDG-RD appears as at least two conformers in slow exchange: a "closed" form, bound to the catalytic domain, undetectable on the $^{15}N-^{1}H$ HSQC spectrum and an "open" form in which the molecular contacts with the catalytic domain are disrupted. In the latter form, the TDG-RD adopts the same conformation as in the isolated N-terminus, preserving its residual structure and its ability to bind DNA in a nonselective manner.

The N- and C-Termini Inhibit TDG Glycosylase Activity. It has been known for some time that truncation of the entire N-terminus of TDG results in increased G•U processing activity while it strongly reduces the G•T repair (*13, 19*). Furthermore, the C-terminus of TDG has no effect on G•U and G•T glycosylase reactions (*19*). In light of our observation that the TDG-RD interacts with DNA, we sought to investigate the different contributions of the extreme N-terminus, the RD, and the C-terminus on TDG G•T and G•U processing activity. Using a DNA nicking assay, we have observed strong modifications of the G•T and G•U glycosylase activity of TDG when removing the 1−50 region of the N-terminus and the full C-terminus. In the case of the G•T glycosylase activity, we have determined an increase of 25% and 55% for TDG-ΔC and TDG-ΔN[1−50]/ΔC, respectively. In the case of the G•U activity, the increase is of 40% and 100%, respectively (Figure 6A,B). The G•U activities of TDG-ΔC and TDG-ΔN[1−50]/ΔC are charac-

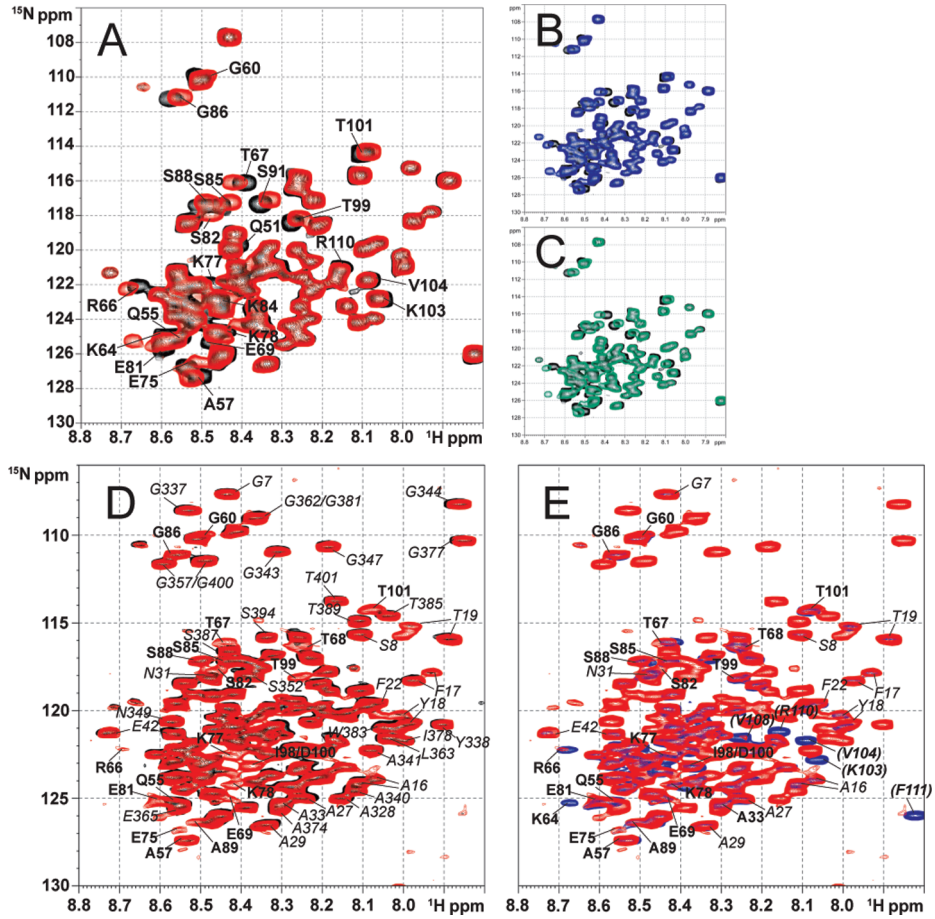

FIGURE 5: *In vitro* DNA binding to the TDG N-terminus. (A−C) Interactions between TDG-N and 25-mer dsDNA. Comparison of ^{15}N−^{1}H HSQC spectra of ^{15}N-TDG-N alone (black) and in the presence of a 4-fold excess of DNA containing (A) a G•T (red) or (C) a G•U (green) mispair or (B) a canonical G•C pair (blue). Resonances of TDG-N affected by the interactions with DNA are annotated. (D, E) Conformational modification of the RD domain within full-length TDG upon DNA binding. (D) ^{15}N−^{1}H HSQC spectra of ^{15}N-labeled TDG in the presence of 5% glycerol without substrate (black) or with a 2.5-fold excess of a dsDNA substrate containing a G•C pair (red). (E) Comparison of ^{15}N−^{1}H HSQC spectra of ^{15}N-labeled TDG in the presence of 5% glycerol with a 2.5-fold excess of a dsDNA substrate containing a G•C pair (red) and ^{15}N-labeled TDG-N in the presence of a 4-fold excess of G•C-containing dsDNA (blue). The TDG-RD resonances are indicated in bold characters and those of the N[1−50]- and C-terminal regions, that are not affected by DNA, are annotated in italic. The RD residues of the isolated N-terminus submitted to a "border effect" due to the absence of the catalytic domain are indicated in parentheses.

terized by turnover rates of 0.0112 and 0.0180 min^{-1}, respectively, while that of the full-length enzyme is of 0.0056 min^{-1}, and the turnover rates related to G•T activities are significantly lower (Figure 6C) as is in accordance with the model of an evolutionarily acquired G•T processing capacity through additional stabilization of TDG/DNA complexes by the TDG-RD (*13*). Moreover, the glycosylase kinetics of TDG-ΔN[1−50]/ΔC on the G•U substrate (Figure 6B) is similar to those described previously for the SUMO-modified TDG or the N-terminally truncated protein (*19, 26*). Interestingly, in this case, the plateau characteristic of product inhibition of enzymatic activity is not observed (Figure 6B).

The RD/CAT Intramolecular Interaction Is Not Essential for G•T Repair and Limits the Enzymatic Turnover. We next investigated whether the glycosylase activity of TDG is dependent on the TDG-RD conformation or not. Given that the N-terminus inhibits the G•U processing while allowing the G•T repair (*13, 19*) and glycerol in presence of DNA modifies the RD conformational equilibrium, more particularly the RD/CAT intramolecular interactions, a modification of TDG glycosylase activity and/or kinetics upon glycerol addition could be expected mimicking the full N-terminus truncation. Hence, we have performed the glycosylase kinetics in the presence of 5% glycerol with TDG full-length and TDG-ΔN[1−50]/ΔC proteins (Figure 6). A significant increase in G•U and G•T activities is detected upon glycerol addition as well as a modification of the kinetic profiles. Interestingly, the impact of glycerol on G•U activity of the entire TDG is of the same order on the enzymatic turnover than the N[1−50]- and C-terminal truncations while only a slight increase on the G•T repair was measured with no modification of the enzymatic turnover. The same difference between the G•U and G•T activities was measured for TDG-ΔN[1−50]/ΔC, however, with equal turnover rates (Figure 6C). Surprisingly, in the case of the G•U activity, both effects (i.e., the constitutive N[1−50]-/C-terminal truncation and the glycerol-induced conformational perturbation) are additional after 1 h incubation (Figure 6B) while they are synergistic in the case of the G•T activity (Figure 6A). This discrepancy between G•U and G•T repair indicates that the RD/CAT intramolecular interaction (i) is the target of the glycerol-induced perturbations and (ii) is not essential for G•T repair but rather has an inhibitory effect on both G•T and G•U repair.

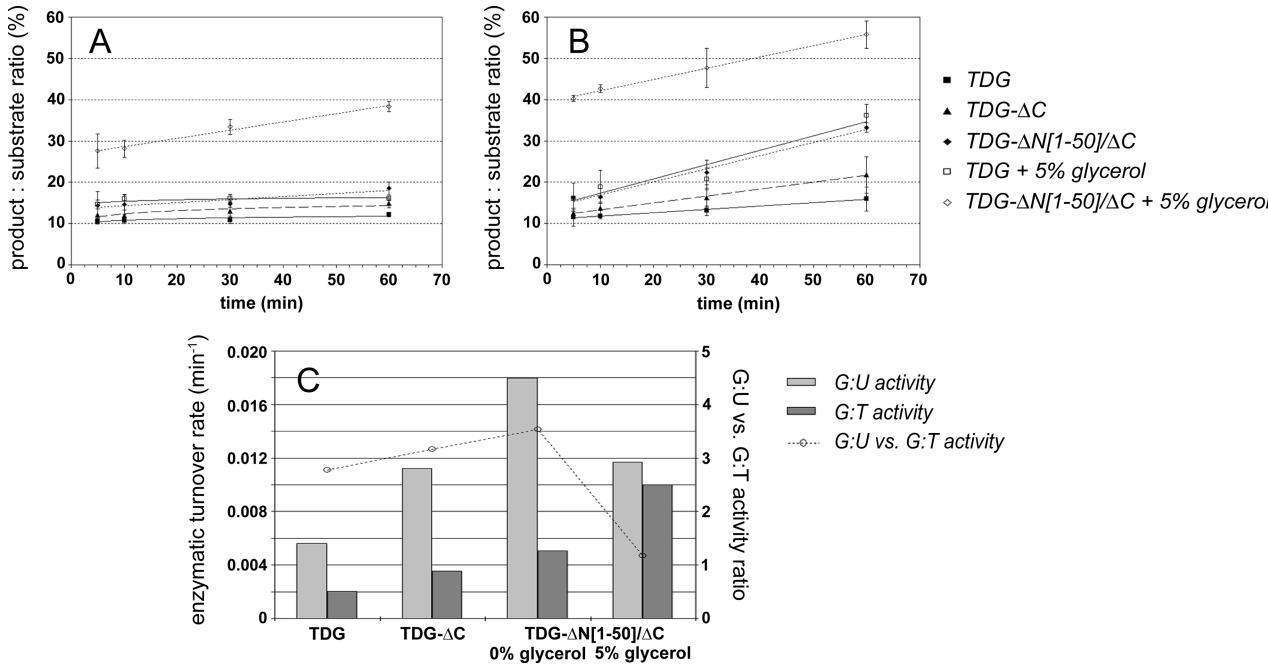

FIGURE 6: *In vitro* TDG glycosylase activity in base excision repair. Activity of 10% molar ratio of full-length TDG (full squares), TDG-ΔC (full triangles), TDG-ΔN[1−50]/ΔC (full diamonds), or TDG (open squares) and TDG-ΔN[1−50]/ΔC (open diamonds) in the presence of 5% glycerol on G•T (A) and G•U (B) containing 25-mer dsDNA. (C) Graphical representation of the enzymatic turnover rates of the three enzymes on G•T (dark gray bars) and G•U (light gray bars) substrates and ratio of G•U over G•T activity (circles and dotted line).

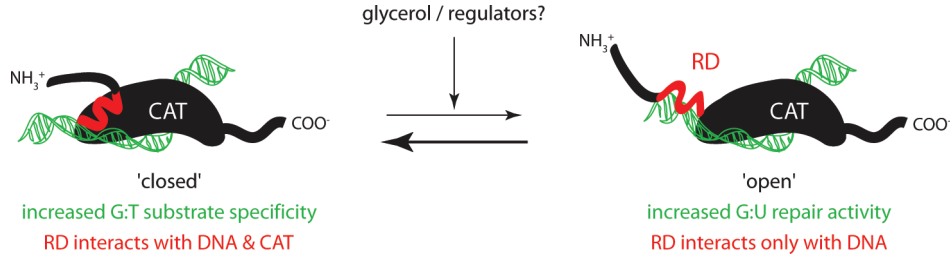

FIGURE 7: Schematic illustration summarizing the different conformers and their activity. The regulatory domain (RD) is shown in red.

DISCUSSION

Using NMR spectroscopy we have shown here that the N-terminus of TDG, previously considered to have little structural organization (*40*), indeed adopts a detectable and distinct conformation in the context of the entire protein with the potential to regulate TDG function. The structural organization of the TDG N-terminus is limited to a region comprised between amino acids 51 and 111 and remains detectable even in the absence of the remainder of the protein. This region is known to be essential for G•T but not G•U DNA repair activity (*13*) and had previously been implicated not only in the regulation of TDG activities in BER but also in transcription and acts as a domain mediating numerous protein−protein interactions (*21–24*). Given that the functional activity of this part of TDG is matched with a distinct structural identity, we refer to the hTDG 51−111 region as the regulatory domain (TDG-RD). In contrast, the extreme N-terminal region from amino acids 1−50 preserves its conformation in both the absence or presence of the catalytic domain (TDG-CAT) and C-terminus of TDG with the $^{1}H-^{15}N$ resonances being strictly identical (Figure 2A). Two proline *cis/trans* isomerizations with a relatively high level of *cis* content are detected in this part of the molecule at positions Pro21 and Pro41, however, with unknown significance. Again, the isomerization equilibrium of both prolines is not modified by the presence or absence of the catalytic domain.

Previously, two distinct hypotheses concerning the function of the RD in establishing G•T repair specificity did exist: (i) where the RD would directly interact with DNA to stabilize TDG on DNA substrates, thereby allowing G•T processing activity, and (ii) an allosteric mechanism with a direct interaction and complementation of the CAT active site by the RD (*13, 19, 26, 40*). Our demonstration of an interaction between the CAT and the RD, even in the absence of DNA substrate, seemed to favor *a priori* the latter hypothesis. Moreover, given the fact that a large excess of TDG-N (50-fold) displaces only partially the interactions between the regulatory and catalytic domains, it is likely that these interactions are intramolecular and relatively strong. Interestingly, using a model of the electrostatic potential surface of the TDG-CAT domain, we were able to identify a negative patch on the CAT surface just adjacent to the N-terminal junction which could provide for the interaction surface with the RD (see Supporting Information, Figure S4). The regulatory domain could thus add supplementary substrate recognition surfaces, notably with the additional methyl group of the thymine, that allow TDG to adapt

substrates containing a G•T mismatch. Such interactions had been postulated and proposed to allow processing of less favorable substrates at the expense of the enzymatic turnover (*13, 19*), which is in accordance also with our enzymatic turnover rate measurements (Figure 6C). On the other hand, we have detected direct interactions of the isolated RD with double-stranded DNA carrying or not a G•T/U mismatch (Figure 5). Similarly, the RD in its protein context was able to bind a DNA substrate containing a canonical G•C pair (Figure 5D). Moreover, this DNA substrate is able to partially compete with the RD/CAT interaction, resulting in a conformational exchange between a "closed" form, where the RD contacts the catalytic domain of TDG, and an "open" form, identical to the isolated RD exhibiting an overall extended conformation with little residual structure (Figure 5E), which is still able to bind DNA in a nonselective manner.

It has been shown that the full N-terminus, while adding G•T processivity (*13*), significantly lowers the processing of G•U substrates (*19*). Our results indicate that the regulation of DNA binding as well as G•T/U recognition and repair by TDG involves two different mechanisms mediated by the N-terminus. First, our data show that the 1−50 region, in combination with the full C-terminal domain (residues 340−410), does not interact directly with DNA but rather destabilizes the TDG/DNA interactions, hence interfering with the enzymatic activity (Figures 5 and 6). The high flexibility and mobility of these two regions conferred by their extended and disordered nature likely induce steric hindrance that hampers the molecular contacts between TDG and DNA, leading to a lower activity for the full-length protein as compared with TDG-ΔN[1−50]/ΔC. Remarkably, this effect has a more pronounced consequence on the repair of G•U substrates. Second, we have shown that the regulatory domain has a nonselective DNA binding activity when the molecular contacts with the catalytic domain are disrupted. Since the enzymatic reaction on G•T and G•U substrates leads to identical products (i.e., AP sites) and the TDG-RD binding affinity of the "open" conformer is of the same order for both substrates (Figure 5), the difference between both processing activities of TDG-ΔN[1−50]/ΔC indicates different recognition modes that are likely mediated by the TDG-RD in its "closed" conformation. Indeed, an overall increase of glycosylase activities and enzymatic turnover was observed upon glycerol addition similarly to the deletion of the full N-terminus (*19*), with the exception that the glycosylase activity on G•T mismatches is not abolished, indicating more likely a conformational effect on the RD/CAT interaction. This upregulation of both activities suggests that the RD/CAT intramolecular interaction is not essential for the G•U and, more interestingly, for the G•T processing but rather has an inhibitory effect probably by improving the efficiency of DNA binding with, as a result, a limitation of the turnover rate. The discrepancy between G•U and G•T repair efficiency is increased upon N[1−50]/C-terminus deletion and/or glycerol addition, indicating that the RD has, as expected, a stronger impact on G•T processing. The schematic representation shown in Figure 7 illustrates and summarizes these findings. In conclusion, the RD enhances the DNA binding affinity of TDG to allow G•T repair at the expense of the G•U activity and enzymatic turnover.

In this context it is also most interesting to note that SUMO binding is influenced by the N-terminal domain (*27, 28*) and sumoylation of TDG impairs G•T processivity while increasing G•U repair activity (*26*). This strongly suggests that SUMO binding and sumoylation are determined by the regulatory domain and have consequences for its autonomous DNA binding capacity or its ability to complement the active center for G•T substrate recognition or both. Whereas at present we cannot determine here the cause and effect relationship, it seems likely that sumoylation directly or indirectly regulates TDG binding to DNA and also APE1 recruitment (*26*) in order to have a selective effect on G•T repair. These hypotheses are significantly corroborated by the observation that CBP interaction and CBP-mediated acetylation of TDG-RD are inhibited by sumoylation of TDG (*27*) and even more so by the fact that in turn acetylation of TDG prevents APE1 recruitment (*24*). The effects of CBP binding and acetylation of TDG hence are mutually exclusive with those for SUMO binding and sumoylation.

Moreover, the intricate mechanism of G•U versus G•T mismatch repair specificity and the role of the TDG-RD in its regulation underline the importance of fine-tuning TDG activity by subtle structural dynamics. The fact that G•T mismatches in DNA arise almost exclusively at sites of epigenetic regulation by cytosine methylation, in conjunction with ample evidence of TDG's role in transcription (*21–24*), as well as the recent observation of a direct interaction between TDG and Dnmt3a (*25*), supports the idea of a regulated link between epigenetic regulation and transcription via the intermediary of TDG and its G•T repair activity, and thus the need to specifically regulate G•T versus G•U processing activity, through the regulation of the RD conformation and its inter- and intramolecular interactions.

ACKNOWLEDGMENT

The authors thank Dr. Guy Lippens and Isabelle Landrieu for insightful discussions.

SUPPORTING INFORMATION AVAILABLE

Figures S1−S4 as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Aravind, L., and Koonin, E. V. (2000) The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol. 1*, (RESEARCH0007).
2. Huffman, J. L., Sundheim, O., and Tainer, J. A. (2005) DNA base damage recognition and removal: new twists and grooves. *Mutat. Res. 577*, 55–76.
3. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995) Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell 80*, 869–878.
4. Krokan, H. E., Drablos, F., and Slupphaug, G. (2002) Uracil in DNA-occurrence, consequences and repair. *Oncogene 21*, 8935–8948.
5. Ehrlich, M., Zhang, X. Y., and Inamdar, N. M. (1990) Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine residues with cytosine residues. *Mutat. Res. 238*, 277–286.
6. Waters, T. R., and Swann, P. F. (2000) Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutat. Res. 462*, 137–147.
7. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) Crystal structure of a G•T/U

mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell 92*, 117–129.
8. Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J. 18*, 6599–6609.
9. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., and Jiricny, J. (1996) Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem. 271*, 12767–12774.
10. Neddermann, P., and Jiricny, J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J. Biol. Chem. 268*, 21218–21224.
11. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature 401*, 301–304.
12. Mol, C. D., Arvai, A. S., Begley, T. J., Cunningham, R. P., and Tainer, J. A. (2002) Structure and activity of a thermostable thymine-DNA glycosylase: evidence for base twisting to remove mismatched normal DNA bases. *J. Mol. Biol. 315*, 373–384.
13. Gallinari, P., and Jiricny, J. (1996) A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature 383*, 735–738.
14. Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2005) Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature 435*, 979–982.
15. Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., and Shirakawa, M. (2006) Crystal structure of SUMO-3-modified thymine-DNA glycosylase. *J. Mol. Biol. 359*, 137–147.
16. Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000) Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem. 275*, 33449–33456.
17. Pearl, L. H. (2000) Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res. 460*, 165–181.
18. Wu, P., Qiu, C., Sohail, A., Zhang, X., Bhagwat, A. S., and Cheng, X. (2003) Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4. *J. Biol. Chem. 278*, 5285–5291.
19. Steinacher, R., and Schar, P. (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol. 15*, 616–623.
20. Ulrich, H. D. (2005) SUMO modification: wrestling with protein conformation. *Curr. Biol. 15*, R257–R259.
21. Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998) Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem. 273*, 20728–20736.
22. Chen, D., Lucey, M. J., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Losson, R., Buluwela, L., Coombes, R. C., Chambon, P., Schar, P., and Ali, S. (2003) T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J. Biol. Chem. 278*, 38586–38592.
23. Lucey, M. J., Chen, D., Lopez-Garcia, J., Hart, S. M., Phoenix, F., Al-Jehani, R., Alao, J. P., White, R., Kindle, K. B., Losson, R., Chambon, P., Parker, M. G., Schar, P., Heery, D. M., Buluwela, L., and Ali, S. (2005) T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. *Nucleic Acids Res. 33*, 6393–6404.
24. Tini, M., Benecke, A., Um, S. J., Torchia, J., Evans, R. M., and Chambon, P. (2002) Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell 9*, 265–277.
25. Li, Y.-Q., Zhou, P.-Z., Zheng, X.-D., Walsh, C. P., and Xu, G.-L. (2007) Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res. 35*, 390–400.
26. Hardeland, U., Steinacher, R., Jiricny, J., and Schar, P. (2002) Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J. 21*, 1456–1464.
27. Mohan, R. D., Rao, A., Gagliardi, J., and Tini, M. (2007) SUMO-1 dependent allosteric regulation of Thymine DNA Glycosylase alters subnuclear localization and CBP/p300 recruitment. *Mol. Cell. Biol. 27*, 229–243.
28. Takahashi, H., Hatakeyama, S., Saitoh, H., and Nakayama, K. I. (2005) Noncovalent SUMO-1 binding activity of thymine DNA glycosylase (TDG) is required for its SUMO-1 modification and colocalization with the promyelocytic leukemia protein. *J. Biol. Chem. 280*, 5611–5621.
29. Lippens, G., Sillen, A., Smet, C., Wieruszeski, J. M., Leroy, A., Buee, L., and Landrieu, I. (2006) Studying the natively unfolded neuronal Tau protein by solution NMR spectroscopy. *Protein Pept. Lett. 13*, 235–246.
30. Lippens, G., Wieruszeski, J. M., Leroy, A., Smet, C., Sillen, A., Buee, L., and Landrieu, I. (2004) Proline-directed random-coil chemical shift values as a tool for the NMR assignment of the tau phosphorylation sites. *ChemBioChem 5*, 73–78.
31. Smet, C., Leroy, A., Sillen, A., Wieruszeski, J. M., Landrieu, I., and Lippens, G. (2004) Accepting its random coil nature allows a partial NMR assignment of the neuronal tau protein. *ChemBioChem 5*, 1639–1646.
32. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HN-Hα) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc. 115*, 7772–7777.
33. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR 5*, 67–81.
34. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*, 1647–1651.
35. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci. 15*, 2795–2804.
36. Dyson, H. J., Bolinger, L., Feher, V. A., Osterhout, J. J., Jr., Yao, J., and Wright, P. E. (1998) Sequence requirements for stabilization of a peptide reverse turn in water solution. Proline is not essential for stability. *Protein Sci. 255*, 462–471.
37. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR 18*, 43–48.
38. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol. 279*, 449–460.
39. Gekko, K., and Timasheff, S. N. (1981) Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry 20*, 4667–4676.
40. Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007) The enigmatic thymine DNA glycosylase. *DNA Repair 6*, 489–504.

BI7022283