5602–5613 *Nucleic Acids Research, 2017, Vol. 45, No. 9* *Published online 22 March 2017*
doi: 10.1093/nar/gkx196

NMR-based method of small changes reveals how DNA mutator APOBEC3A interacts with its single-stranded DNA substrate

Stefan Harjes, Geoffrey B. Jameson, Vyacheslav V. Filichev, Patrick J. B. Edwards and Elena Harjes*

Institute of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand

Received December 23, 2016; Revised February 21, 2017; Editorial Decision March 10, 2017; Accepted March 14, 2017

ABSTRACT

APOBEC3 proteins are double-edged swords. They deaminate cytosine to uracil in single-stranded DNA and provide protection, as part of our innate immune system, against viruses and retrotransposons, but they are also involved in cancer evolution and development of drug resistance. We report a solution-state model of APOBEC3A interaction with its single-stranded DNA substrate obtained with the 'method of small changes'. This method compares pairwise the 2D ^{15}N–^{1}H NMR spectra of APOBEC3A bearing a deactivating mutation E72A in the presence of 36 slightly different DNA substrates. From changes in chemical shifts of peptide N–H moieties, the positions of each nucleotide relative to the protein can be identified. This provided distance restraints for molecular-dynamic simulations to derive a 3-D molecular model of the APOBEC3A–ssDNA complex. The model reveals that loops 1 and 7 of APOBEC3A move to accommodate substrate binding, indicating an important role for protein-DNA dynamics. Overall, our method may prove useful to study other DNA–protein complexes where crystallographic techniques or full NMR structure calculations are hindered by weak binding or other problems. Subsequent to submission, an APOBEC3A structure with a bound DNA oligomer was published and coordinates released, which has provided an unbiased validation of the 'method of small changes'.

INTRODUCTION

Members of the family of human APOBEC3 (A3A−H) proteins attack retro- (RNA) viruses and other pathogens by hyper-mutating their single-stranded DNA (ssDNA), produced upon entry of the retrovirus into the cell. A3 proteins mutate cytidine to uridine in ssDNA, thereby destroying genetic information (1–3). *In vitro* A3 proteins are only active towards ssDNA and do not mutate, at least to a measurable extent, double-stranded DNA (dsDNA) or RNA (3). This specificity allows A3 proteins to target a unique but essential viral life-cycle intermediate, when the viral RNA genome is reverse transcribed and temporarily exists as single-stranded DNA. Normally, by attacking only ssDNA, potentially deadly deamination of human genetic information, which is mostly present in the form of dsDNA, is averted. Each A3 protein has its own ssDNA sequence preference and counteracts a distinct set of pathogens. A3A, for example, is capable of clearing plasmids and other foreign DNA from human cells (4), as well as inhibiting retro elements (5) and viruses (6,7). A3A can also clear dsDNA from cells, but the *in vitro* deamination of dsDNA could not be shown (4).

Moreover, the single-domain proteins A3A and A3H as well as a double-domain protein A3B, have been shown to be present in the nucleus of human cells and to be associated with mutagenesis and carcinogenesis (8–15). A3A has been shown to induce double-stranded DNA breaks and is proposed to be involved in skin cancer (10,11). A high level of A3B in human breast cancer is reportedly a chronic source of mutations and contributes to rapid tumor growth and evolution of resistance to anti-cancer drugs (16). There is evidence that A3B is also involved in bladder, cervix, lung, head and neck carcinogenesis (9).

Although related structurally and functionally to cytidine deaminases (17), which are well-characterised in ligand-free and ligand-bound forms (18–24), A3 proteins do not deaminate individual cytosine or cytidine moieties at significant rates, and conversely cytidine deaminases do not accept as substrates oligomeric DNA. Using high-resolution 3D NMR and X-ray crystallographic techniques, multiple A3 structures have now been solved (25–35), but at the time of submission, there was only one crystal structure of an A3–substrate complex, where just one base from the DNA substrate was resolved (36). We report here a novel approach, which we call the 'method of small changes', to derive an

*To whom correspondence should be addressed. Tel: +64 6 356 90 99; Fax: +64 6 355 7947; Email: E.Harjes@massey.ac.nz

NMR-based model for the interaction of A3A with its ssDNA substrate. In contrast to standard titrations that monitor the chemical-shift perturbations in NMR spectra of the protein upon addition of ssDNA, we compared small changes in chemical shifts of A3A, due to adding slightly different ssDNA molecules (adding or removing a single nucleotide, changing just one base in an oligomeric DNA and altering one phosphate group) (Figure 1). This method allows us to derive an experimentally based model, which, on optimisation by molecular dynamics, reveals new features of A3A–ssDNA interactions. Subsequent to submission of this paper, the X-ray structure at 3.15 Å resolution of an A3A complex with a 15-mer single-stranded DNA was published and coordinates released (37), allowing an unbiased comparison of that model with our NMR-MD model.

MATERIALS AND METHODS

Oligonucleotides and other materials

Desalted DNA oligonucleotides were purchased (Integrated DNA Technologies) at 1 μmol synthesis scale and dissolved in one of the buffers described below to give 10 mM solutions. The modifications ordered are outlined in Figure 1 and a table listing all oligonucleotides is included in the Supplementary Material. Chemicals were purchased from Sigma or as indicated. A zinc determination kit (Abcam) was used to select only zinc-containing A3A–E72A in chromatographic separations.

Protein expression and purification

Human APOBEC3A (uniprot code P31941) was cloned as the inactive E72A mutant with a His_6 C-terminal fusion tag into an expression vector (pETite, Lucigen) and expressed in *Escherichia coli* BL21 DE3 cells (Hi-Control, Lucigen). The expression medium was supplemented with 100 μM Zn^{2+}; for ^{13}C and ^{15}N labeling M9 media (38) were used. Cells were grown in 5 l shake flasks at 37°C and shortly before induction cooled to 20°C. Protein expression was induced with 0.25 mM IPTG and protein was expressed at 20°C overnight. Cells were harvested and resuspended in the following buffer: 25 mM sodium phosphate, pH 7, 500 mM sodium chloride, 5 mM β-mercaptoethanol and 0.2 mM Na_2-EDTA or optimized buffer (25 mM sodium phosphate, 500 mM sodium chloride, 5 mM β-mercaptoethanol and 0.2 mM Na_2-EDTA at pH 6.0 plus 300 mM choline acetate and 0.1 mM LysoFos Choline 12 (Anatrace), see protein buffer optimisation). Cells were disrupted by sonication or french press, with about twice the yields for the latter method. Lysate was cleared by centrifugation (25 min, 4°C, 38 000 *g*) and the supernatant was collected. The protein was purified by metal affinity chromatography. After loading the supernatant, the column was washed three times with 10 column volumes of the above buffer containing 50 mM imidazole followed by one wash with 200 mM imidazole and eluted with 1 M imidazole. The eluted protein was loaded on a gel filtration column (Highload 16/600 Superdex 75 pg) and the peak shortly before 100 ml elution volume was collected. The protein was concentrated to about 100 μM using Centricons (Vivaspin 20, 10 kDa MWCO) and aliquots were frozen at –70°C.

Thermal shift assay

A commercial SYPRO Orange stock solution (Sigma) was used as dye for the assay. A master mix of dye at a volumetric dilution of 1:55.6 in buffer described above with a protein concentration between 5 and 50 μM was prepared and centrifuged for 1 min (1000 *g*) and supernatant was used. The components or oligonucleotides to be tested were dispensed into 96 multiwell qPCR plates and diluted in the above master mix (1:10). Samples were placed in a qPCR machine Lightcycler II, Roche and slowly (1°C per 15 s) heated from 30 to 80°C. Fluorescence emission spectra were recorded with several different combinations of excitation and emission filters with identical results (465–580, 483–610 and 483–568 nm). The first derivative was calculated from the raw data and plotted *versus* temperature. The maxima of these curves were recorded as melting points.

Protein buffer optimisation

The Hampton screens 'Solubility and Stability Screen 2' and the 'Detergent Screen HT' were used to search for stabilising components. As recommended by the supplier, each of the 192 components was used at a volumetric ratio of 1:10 (see above). The solubility screen was tested using the above thermal shift assay. Increased melting points were used as indicator of increased protein stability. Detergents cannot normally be used in thermal shift assays, because they strongly increase the baseline fluorescence of the samples at room temperature. Thus, the 96 detergents were screened using ultra centrifugation. Each detergent was diluted 10 times in protein in base buffer (above). The samples were incubated at room temperature for 30 min and then centrifuged at 435 630 *g* for 20 min (Sorvall MTX150, rotor: S100-AT3). The protein concentration in the supernatant was determined using Bradford assay. Increased protein concentration was taken as indicator for increased protein solubility and stability. The optimised buffer is 25 mM sodium phosphate, 500 mM sodium chloride, 5 mM β-mercaptoethanol and 0.2 mM Na_2-EDTA at pH 6 plus 300 mM choline acetate and 0.1 mM LysoFos Choline 12 (Anatrace). Choline acetate concentration was reached by using the individual components (choline chloride and sodium acetate). The sodium chloride content of the buffer was adjusted accordingly to keep the pH on 6.0. The additive, although stabilising, reduced the quality of the HSQC spectra slightly.

NMR spectroscopy

All spectra were collected with a 700-MHz Bruker Avance NMR spectrometer with 3-channel inverse-detect cryoprobe using high-salt optimised 5 mm NMR tubes (Shigemi) at 25°C. ^{15}N HSQC spectra were used to monitor chemical shift perturbations. 3D HNCA spectra, both with and without ssDNA, were used to adjust the assignment of the previously reported solution structure of A3A (30) to our conditions. Spectra were processed with Topspin 2.1.8 (Bruker) and analysed with CARA (http://www.cara.nmr.ch). Protein was usually measured in optimised buffer pH 6 (above) at a concentration of 100 μM. The final concentration of oligodeoxynucleotides was 500 μM.

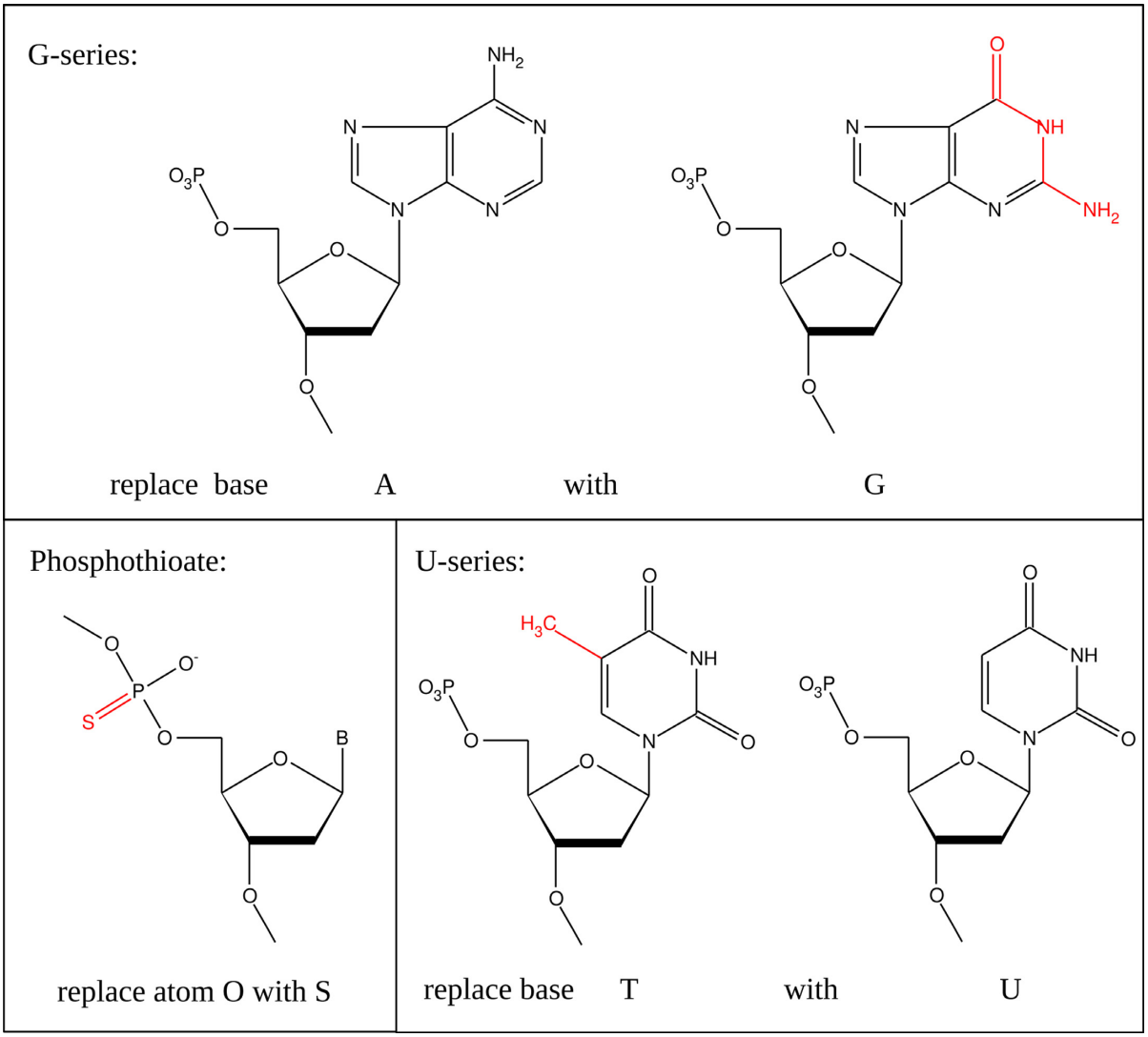

Figure 1. Scheme of chemical modifications used in the 'method of small changes' (the list of oligonucleotides is shown in Supplementary Table S1).

Isothermal titration calorimetry

ITC experiments were conducted at 25°C using a MicroCal ITC200 (now Malvern Instruments) isothermal titration calorimeter. 130 μM A3A–E72A was titrated in optimised buffer at pH 6 (above). DNA oligonucleotides at 1.6 mM were added in 18 steps at 2.0 μl each (plus a first addition with reduced volume of 0.4 μl to prevent dilution of the DNA in the syringe due to the long wait before the start of the experiment).

Model calculation

A model for the A3A–ssDNA complex was calculated with YASARA (39) using molecular dynamic simulations in the presence of experimentally derived restraints. The simulations were run with explicit water (TIP3P model, 3% (w/v) sodium chloride, pH 6), using the YASARA2 force field (with long-range Coulomb forces and 8.0 Å cut-off (40). Restraints were derived from chemical shift perturbations observed when comparing titrations of A3A with slightly different ssDNA oligomers. We progressed from comparing ssDNA oligomers with one nucleotide length difference to smaller and smaller differences (Figure 1). All protein residues, where changes in positions of resonances in overlaid pairs of NMR spectra were noticeable to the eye, and where assignment was unambiguous, defined peptide-nucleotide restraints to be included in molecular-dynamics simulations (Supplementary Tables S2–S4). Simulations were computed with different distances assigned to the restraints, with the best (that is, converging simulations with different initial starting velocities) results at 4 Å. The distance was allowed to vary from the minimal (van der Waals) distance of the two atoms to 6 Å. The chemical titration data were filtered, such that physically impossible interactions, arising from chemical shifts of completely buried residues, or residues on the opposite side of the protein, were removed. In total, 84 experimentally derived distance restraints between the substrate and the protein were included. Each restraint was set as a distance between any atoms of the amino acids for which the peptide

NH proton changed to any atoms of the corresponding ssDNA oligonucleotide that induced the change in chemical shift. The restraints covered all nucleotides of the substrate dA_5TTCA_5 except the most 3′ dA, which does not appear to be interacting with the protein.

In order to build the model of the ssDNA–A3A complex we used the crystal structure of A3A (pdb code: 4XXO) and extended it by including information about the missing residues (seven N-terminal, three C-terminal residues and N42, G43, T44) from the solution structure of A3A (pdb code: 2M65). The substrate was modeled in three parts. First the 5′ end of the substrate was placed close to the protein model, according to the distance restraints for this part (Supplementary Table S5). Then the RNA complex of TadA (pdb code: 2B3J) (41) served as model for the dTCA triplet. TadA was aligned with A3A. The dTCA triplet of the dA_5TTCA_5 substrate was superimposed onto the three corresponding bases of the TadA RNA ligand. Finally the 3′ end of the substrate was positioned. Here, only the restraints to loop 3 (Supplementary Table S5) were considered for placement. To relax the starting structure, the triplet dTCA and the protein (secondary structure only) were fixed in place and the newly placed nucleotides of the simulated substrate were allowed to relax.

From this starting structure several simulations under different conditions were run. In order to model the substrate-bound state, the simulations were split into three steps: (i) a strongly fixed and (ii) a moderately fixed step each of 5 ns, and (iii) after 10 ns all restraints were removed and motions of the complex were simulated for at least a further 40 ns. During the strongly fixed step, the two dCA nucleotides of the dTCA triplet were fixed, as well as the zinc and the zinc-bound water in the active site. In addition to the experimentally derived distance restraints, temporary distance restraints were added to enhance stacking of the first six and the last four nucleotides. These additional restraints prevented the helical structure of the substrate from being disrupted by the strong forces originating from the experimental restraints. Additionally, in order to allow the protein to adapt to the bound substrate more quickly, several amino acids were mutated to alanine (see below). Then, during the moderately fixed step, only the three atoms: zinc, zinc-bound water and C4 of 2′-deoxycytidine were fixed. Finally in the third step all restraints were removed.

In all simulations, it was ensured that the zinc coordination is tetrahedral. The coordinating histidine is protonated on one nitrogen and the zinc is coordinated to the (non-protonated) imine nitrogen atom. Thus, the histidine was modelled with two resonance bonds of order 1.5 between ND1 and CE1 and between CE1 and NE2. Also it was checked that no false isomers or *cis* peptides were formed.

The simulations mostly differed in the number and type of amino acids which were mutated to alanine during the first step of simulation. The model adapted very quickly, once we mutated the following list of amino acids: N23, N24, H29, K60, N61, L62, L63, F66, Y67, R69, W104, R128, Y130 and Y132. After the first (strongly fixed) simulation step, these residues were mutated back to their original sequence and their sidechain orientation was optimized (improving physical realism, stereochemistry, and side-chain accuracy in homology modelling).

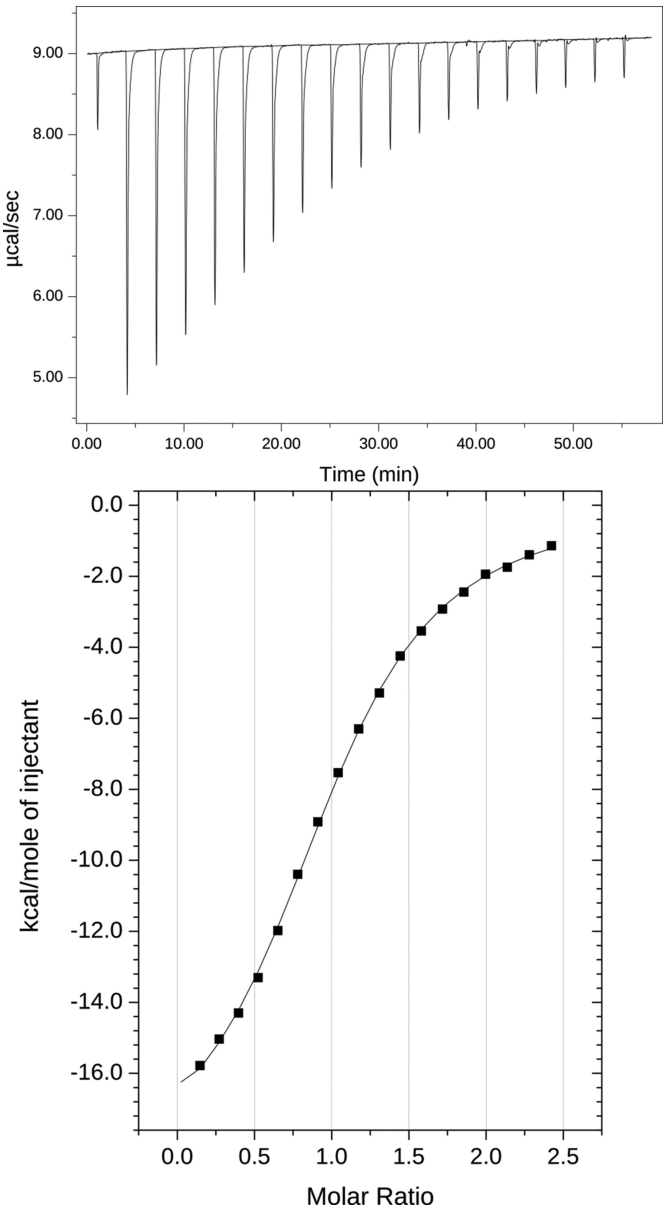

Figure 2. Binding of substrate to A3A–E72A: isothermal titration calorimetry (ITC) of A3A–E72A titrated with ssDNA is shown on the top. The bottom panel shows the integrated data on the *y*-axis plotted *versus* the molar ratio of ssDNA to protein on the *x*-axis. The fit of these data to a one-to-one stoichiometric model is shown as solid line.

RESULTS

Thermodynamics of DNA binding to A3A–E72A

To explore the binding between A3A and ssDNA, we performed isothermal titration calorimetry (ITC) experiments. In this experiment ssDNA as the A_8TTCA_9 oligomer was added to A3A–E72A, and thermodynamic parameters of binding were determined (Figure 2).

The dissociation constant K_d of 24 ± 2 μM at 25°C is roughly consistent with previously reported data (9.1 ± 2.5 μM (42) and 57 ± 11 μM (30) obtained with differing experimental techniques at different pH (42)). The binding is energetically favoured with a change in enthalpy upon bind-

ing of $\Delta H = -19.0 \pm 0.6$ kcal mol^{-1}, and is accompanied by a negative change in entropy $\Delta S = -44$ cal mol^{-1} K^{-1}. The ITC data fitted very well with a one-to-one stoichiometry.

For the shorter sequence A_2TTCA_4, very similar values to the sequence A_8TTCA_9 were obtained (K_d of 17 ± 6 μM, $\Delta H = -18 \pm 3$ kcal mol^{-1} and $\Delta S = -39$ cal mol^{-1} K^{-1}). These ITC data provide validation that the thermal shift assay described below offers a facile method for assaying DNA binding to A3A–E72A.

Thermal shift assay to study A3A–ssDNA binding

ITC is a laborious method, where a lot of protein and DNA are consumed. Up to now there was no report of a quick and easy assay to test DNA binding of A3 proteins. We developed a thermal shift assay for an initial evaluation of DNA binding to A3 proteins.

Thermal shift assays are broadly used to test ligand binding to proteins. In general, the enthalpy change during complex formation is negative and as a consequence the melting of the complex requires more energy than the apo enzyme (43). This is the physical basis of thermal shift assays. The thermal unfolding of a protein is measured in the presence of a fluorescent dye, whose fluorescence is quenched in aqueous buffers at the beginning of the experiment. Upon thermal denaturation (melting) of a natively folded protein, the hydrophobic core of the protein becomes exposed to the environment and the dye interacts with the hydrophobic amino acids, which results in increased fluorescence. The transition or melting temperature is generally higher with a ligand bound to the protein and, in general, stronger binding corresponds to a higher melting temperature. This assay is fast and amenable to high-throughput screening. To investigate the binding between A3A and ssDNA, we have expressed the inactivated mutant A3A–E72A, because the binding between active protein and substrate cannot be investigated directly, as part of the substrate is inevitably converted to the product. As shown in Figure 3, the assay yields different melting temperatures for the free enzyme, the product-bound enzyme and the substrate-bound enzyme of 58.8, 60.4 and 63.1°C, respectively, where the product is 5′-dATTTUATTT and the substrate is 5′-dATTTCATTT. Thus, with a thermal shift assay, A3A–E72A binding to substrate and product (Figure 3, Supplementary Figure S1) can be distinguished.

The thermal shift assay was then used to optimize the buffer for protein stability at 25°C and pH 6.0. We screened several hundred components and observed that ionic liquids increase the melting point of the A3A–E72A protein, in particular choline acetate (300 mM). Unfortunately, the thermal shift assay is not compatible with high concentrations of additives. Thus, we screened detergents for increased protein solubility with a different assay (see Materials and Methods). We found several detergents which increase the solubility of A3A preparations. We mostly worked with the synthetic detergent LysoFos 12 Choline, as it was solubilising at low concentration of 0.1 mM. Additives increase protein solubility (more than two times) and stability (by 3–4°C).

The thermal shift assay was further used to determine the minimal binding sequence by shortening a dA_5TTCA_5 oligonucleotide on one side of the recognition triplet dTTC (Figure 4). Changes in melting are only observed when the 3′-dA_5 including dA at +2 and +1 positions are removed. We use the generally adapted numbering scheme, where the target 2′-deoxycytidine has the number 0, the nucleotide directly 5′ of it, has the number –1 and the nucleotide directly 3′ of the target 2′-deoxycytidine has the number +1. The melting temperature of the complex is not influenced by removal of 5′ nucleotides. The data indicate a minimal substrate of dTTCAA. Also under the experimental conditions (optimized low pH buffer) there is no longer a significant increase of melting temperature in the presence of product (Figure 4).

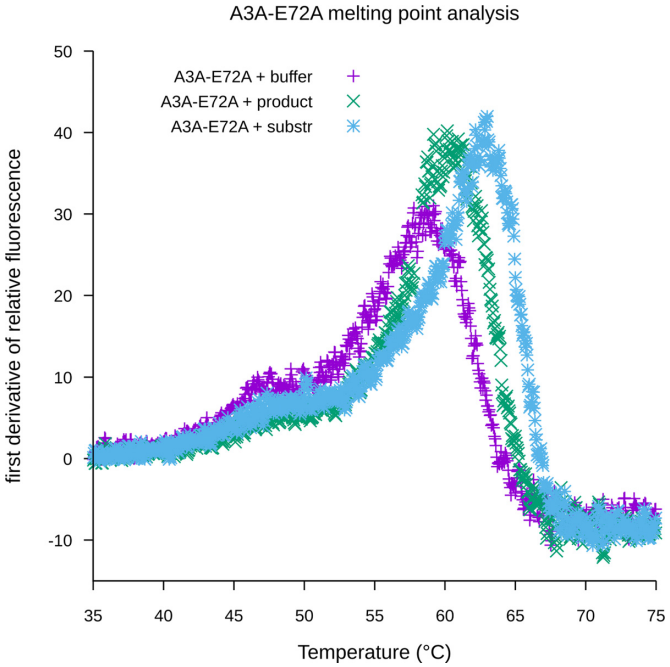

Figure 3. Thermal shift assay showing the first derivative of the relative fluorescence intensity data *versus* temperature for A3A–E72A with substrate 5′-dATTTCATTT (blue), product 5′-dATTTCATTT (green) and buffer (magenta, at 25 mM sodium phosphate, 500 mM sodium chloride, 5 mM β-mercaptoethanol and 0.2 mM Na_2-EDTA at pH 6 plus 300 mM choline acetate and 0.1 mM LysoFos Choline 12). The raw data was smoothed by taking the average of 5 points and then differentiated by subtracting the fluorescent values of point *n-5* from those of point *n* and dividing by the temperature difference of the two points.

Method of small changes

In order to study ssDNA binding to A3A–E72A, heteronuclear two dimensional ^{15}N-HSQC NMR spectra of the free enzyme were compared to spectra of the protein in the presence of substrate (Figure 5A). In optimised buffer all residues are in slow exchange relative to the NMR time scale for ssDNA-bound and unbound species. This means that adding more ssDNA does not change the position of a peak but affects only the signal intensity. The overlay of the two spectra shows that the chemical environment of ∼100 HN protons, corresponding to about half of the protein's 199 + His_6 amino acids, change upon ssDNA binding, as can be judged by looking at Figure 5A. These data

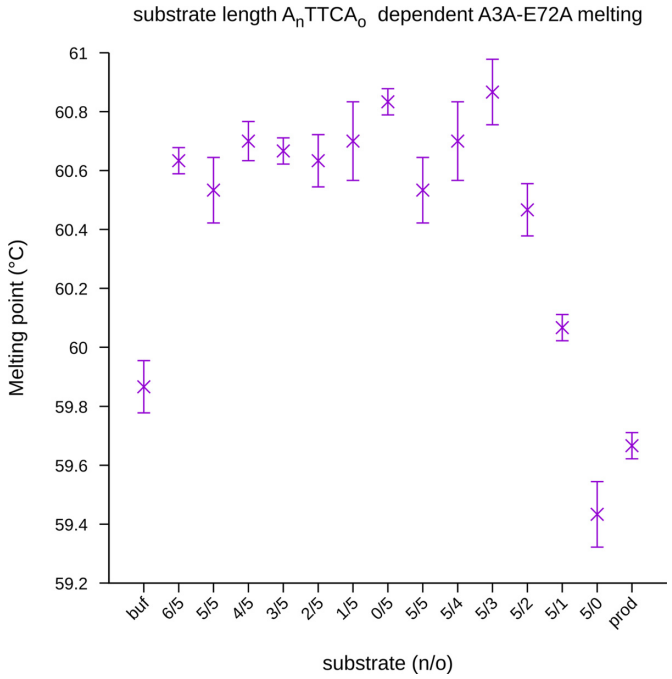

Figure 4. Strength of ssDNA binding with respect to oligonucleotide length. The results of the thermal shift assay showing the melting point for different lengths of oligomeric DNA, A_nTTCA_o. The first number (*n*) of the oligonucleotide denotes the number of dA nucleotides on the 5′-side of the oligonucleotide; the second number (*o*) denotes the number of dA nucleotides 3′ from TTC (see deletion series in the list of oligonucleotides). All experiments were conducted at 25 mM sodium phosphate, 500 mM sodium chloride, 5 mM β-mercaptoethanol and 0.2 mM Na_2-EDTA at pH 6 plus 300 mM choline acetate and 0.1 mM LysoFos Choline 12. Each point is the average of a triplicate measurement where the error bars represent the RMSD.

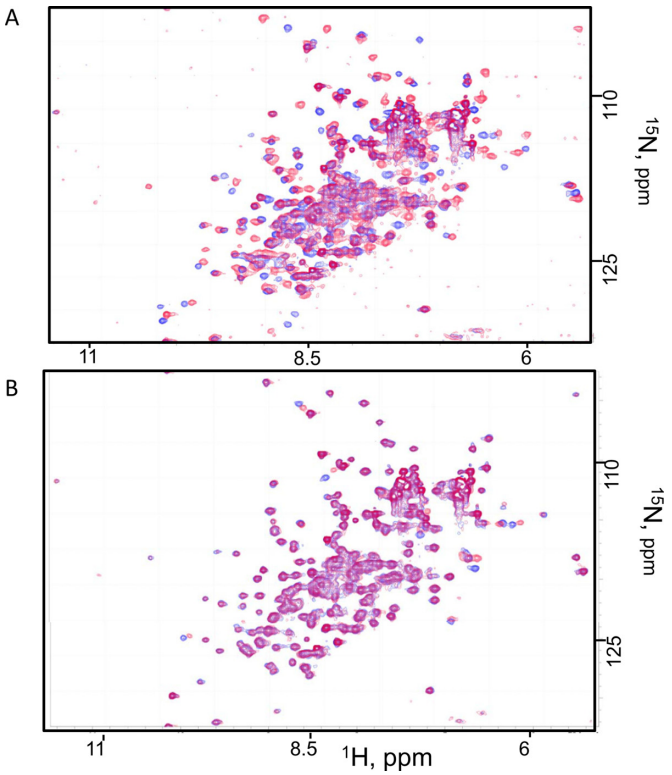

Figure 5. Small changes allow better focus. The difference between the free protein and the substrate-bound protein changes the chemical environment of more than half of the proteins amino acids, as shown in the overlay of the two corresponding ^{15}N-HSQC spectra in panel (**A**) (free A3A–E72A protein in blue, A3A–E72A with ssDNA sequence dATTTCATTT at 130:500 μM in red). In contrast, the overlay of ^{15}N-HSQC spectra of A3A–E72A with dTTCAT (blue) and dTTTCAT (red) (**B**), shows only a few changes, which can thus be assigned to the nucleotide dT(−3). (protein at 130 uM and both ligands at 500 uM).

are too unspecific to allow conclusions to be made about the position and orientation of the bound ssDNA. This and the differences in binding of substrates of different lengths to A3A–E72A as seen in the thermal shift assay, led us to develop the method of small changes. First, we titrated A3A–E72A with ssDNAs of different lengths and monitored changes in the chemical shifts upon 'removal' of a single nucleotide. When comparing two spectra containing substrate (for example dTTCAT *versus* dTTTCAT, Figure 5B), there are far fewer changes between the two spectra and thus one can determine where thymidine in position –3 changes the chemical environment of HN protons of the protein. This method enables the prediction of the position of the 5′-part of the ssDNA substrate (Figure 6). Here and elsewhere, 5′ part and 3′ part refer to the positions relative to target cytosine. The 5′ nucleotides interact with a surface which leads to and contains the catalytic site and is consistent with biochemical data reported in the literature. Several sources report that the 5′-part preceding the target C interacts with loop 1 and loop 7 of A3A–E72A (31,44–48). Comparing ^{15}N-HSQC spectra of ssDNA A3A–E72A complexes that differ by one nucleotide in the 3′ part of the substrate does not allow a simple binding surface prediction. Instead the data indicate that some nucleotides on the 3′-region interact with amino acids that were already assigned to interactions with the 5′ part of the substrate.

In order to further clarify the binding we decreased the number of atoms changed between the pair of substrates whose spectra were overlaid and compared (Figure 1). Instead of removing whole nucleotides, each adenine in the dA_5TTCA_5 substrate was replaced one-by-one by a guanine nucleobase, which has similar spatial requirements but inverted hydrogen-bonding donor–acceptor characteristics. This replacement leads to a series of ten comparisons, in each of which one of the ten 2′-deoxyadenosines is substituted by a 2′-deoxyguanosine. Similarly the thymidines in the substrate can be replaced by 2′-deoxyuridines narrowing the focus to only one methyl group in the whole substrate. In a similar fashion, the interaction of the DNA backbone with A3A–E72A can be probed by substituting single phosphate oxygen by sulfur (phosphothioate modification as a mixture of diastereomers, Figure 1). Nevertheless, the above observation, that changes in the 3′-part result in chemical shift perturbations of residues already implicated in interactions with the 5′-part, was confirmed (Figure 7). This is a strong indication, then, that the 3′-part of the DNA is folded back on itself close to the 5′-part, where the two parts (nucleotides close to target cytosine) can simultaneously influence several residues of the protein. Obviously, a double-stranded RNA adopting a hairpin turn is a prime exam-

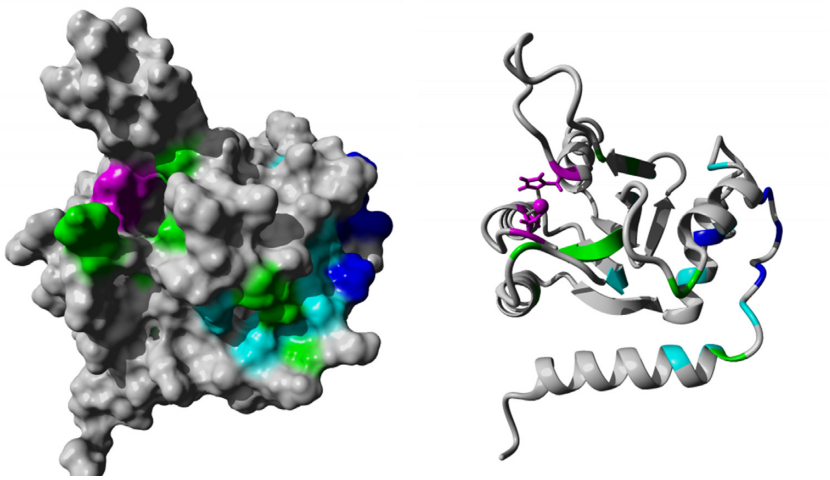

Figure 6. Mapping of the 5′-part of the substrate to the protein's surface yields a continuous patch leading towards the catalytic site. The chemical shift changes of A3A–E72A corresponding to interactions with 5′-part of the oligonucleotide have been coloured: dA(−5) blue, dA(−4) cyan, dA(−3) green on the surface of A3A (pdb code: 4XXO). Zinc, its coordinating water, His70, Cys101 and Cys106 are shown in magenta. Residue dA(−3) affects two regions on the surface of A3A. One of them involves loop 7, which is moved towards the second region in the DNA-containing model (Supplementary Figures S4 and S9).

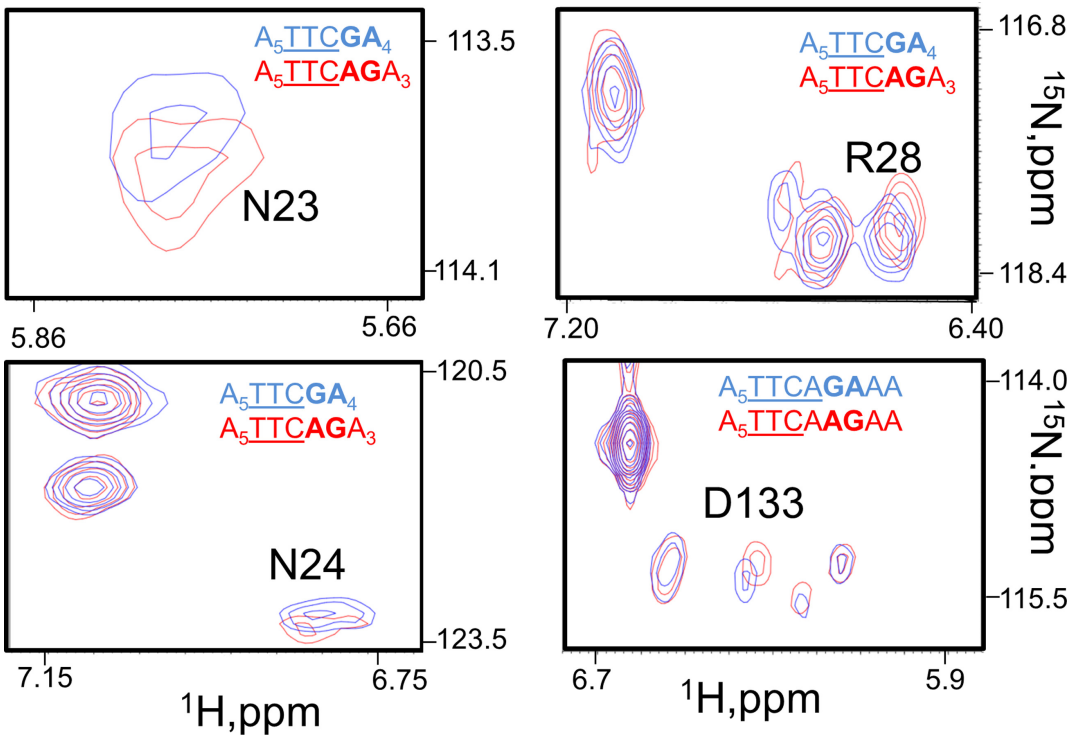

Figure 7. Changes in the 3′-part of the substrate lead to changes in chemical shifts of A3A–72A parts already mapped to 5′-nucleotides. Chemical shift changes of amino acid HN signals in loop 1 and loop 7 in ^{15}N HSQC spectra of A3A–E72A with DNA due to the small changes in 3′ nucleotides are shown. The changed nucleotides are highlighted bold.

ple of such a structure and has been reported for the tRNA adenosine deaminase (TadA, pdb code 2B3J) (41).

When comparing the binding of substrate *versus* product to the A3A–E72A protein, ∼20 chemical shift perturbations are observed. While this number is larger than those generally recorded for small changes, the number is still much smaller than when the apo protein is compared to either the substrate- or product-bound complexes. Some of these changes involve residues 98, 104 and 105, which are located in or close to the active site. Those signals did not shift when comparing apo-protein and product–protein complex. Instead they only shifted in the presence of substrate, indicating some substrate-specific interaction, most likely the full immersion of dC into the active site (Figure 8).

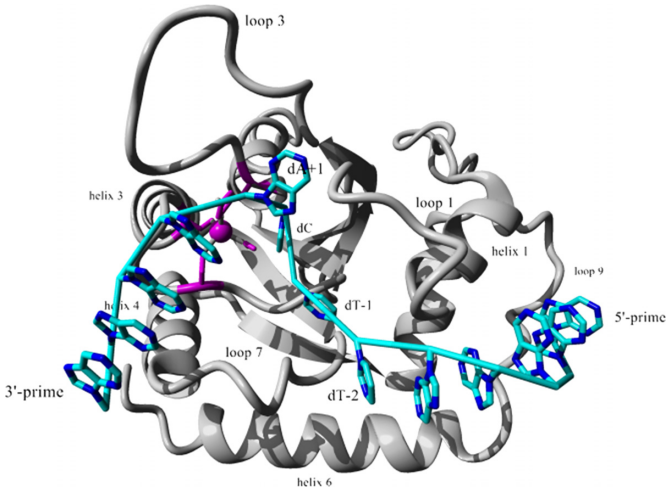

Figure 8. Model of the A3A–E72A–ssDNA complex shown in ribbon representation. The Zn^{2+} site and its ligands: the coordinating water, His70, Cys101 and Cys106 are shown in magenta; the dA_5TTCA_5 oligonucleotide is shown in cyan, with only the aromatic rings of the nucleobases drawn.

Model of ssDNA–A3A interaction

In order to build an atomic-level 3D structure, a starting model for the A3A–ssDNA complex was generated, where the 5′-end of the ssDNA substrate was placed close to the A3A–E72A model (see Materials and Methods) such that the appropriate nucleotides come close to the protein at the experimentally indicated position of this part of the substrate. Four structures of homologous enzymes with nucleotide-like residues bound close to the catalytic site were well aligned with our A3A–E72A model. Three of these—the already mentioned TadA structure (pdb code: 2B3J), the primate APOBEC3G N-terminal domain in complex with ssDNA (pdb code: 5K83) and human cytidine deaminase (pdb code: 1MQ0)—contain at least ribose, which allows prediction of the direction of the 5′-end of the ssDNA substrate (Supplementary Figure S2). In all three cases, the 5′-end of the crystal structure ligands point towards the same direction making it easy to position the ends of the A3A–E72A substrate (Supplementary Figure S2).

When the three active deaminase ligand complexes (2B3J, 1MQ0, 4LD4) are superimposed on A3A, the nucleobases homologous to cytosine are in the same plane and the homologous carbon with attached chemical group to be deaminated is almost at the same position in all three ligands (Supplementary Figure S3).

We used several different starting models. They all had the same above position of the 2′-deoxycytidine, but different conformations of T(−1) and dA(+1). Only models in which the T(−1) adopted the structure of the corresponding nucleotide of the TadA-RNA complex were stable with regard to the 2′-deoxycytidine position in the catalytic cleft. Taken together these observations indicate that the A3A substrate triplet dTCA adopts a similar position as the equivalent residues in the adenosine deaminase (TadA). The starting model for molecular dynamics simulations was completed, placing the 3′-dA_5 close to loop 3.

Once experimental restraints are included, the ssDNA adopts a strongly bent structure, which resembles a question mark or the digit seven, depending on viewing angle (Figure 7). The quality of the model and the convergence of the ensemble are presented in Supplementary Figures S6–S8, and Table S7.

The cytosine to be deaminated is situated at the corner of the seven, deeply embedded into the A3A active site. The 5′-end of the substrate (dA_5T) comes to rest on top of helix 1, almost inserting between helix 1, the N-terminal end of the last helix (helix 6) and the last loop (loop 9) consistent with our chemical shift data (Figure 4, Supplementary Figure S4). However, there is no strong bonding between the 5′-end of the substrate and the protein. Instead, during simulations it was observed that the 5′-end adopted conformations where it moved away from the protein's surface. The long and flexible loop 3 of A3A moves towards the substrate and covers it partially. The 3′-end of the substrate exposes the phosphate groups of the last three residues pointing towards the protein (in particular towards the N-terminal ends of helices 3 and 4). However, the 3′-end is not fixed in place by strong hydrogen bonds or other interactions and thus is quite flexible. The corresponding nucleobases of the 3′-end are arranged towards the 5′-end. The shortest distance between 3′ and 5′-ends is between the proton H2 of the nucleobase dA(−2) and the phosphate oxygens of the dC and is on average ∼3 Å.

At both ends of the substrate, nucleotide bases stack with each other. During simulations, stacking of nucleobases is observed for the six most 5′ nucleotides and the four outermost 3′ dA nucleotides. In some simulations the T(−2) stacking to dA(−3) is also disrupted. The bases of the dTCA triplet at the tight bend are not stacked to other nucleotides. The 2′-deoxycytidine and the T(−1) are buried in the structure and even without restraints remain fixed in position. The 2′-deoxycytidine has three hydrogen bonds which keep it in the catalytic site (Supplementary Table S6). Two of them involve the nucleobase, namely the carbonyl oxygen O2 is bound to the NH proton of Ala 71 and the amino group to be deaminated binds to the carbonyl oxygen of Ser99. Additionally the O3' of the 2′-deoxycytidine ribose forms a hydrogen bond with His29. The thymidine (−1) is also buried quite deeply. It is stacked between Trp98 and Tyr130, and hydrogen bonded between the carbonyl oxygen O2 of the nucleobase with the NH proton of Ser99 (Supplementary Table S6). When the methyl group of T(−2) is removed, by replacing the nucleotide with dU, no chemical shift perturbations are observed, indicating that the methyl group is not pointed toward the protein, but towards the solvent.

When compared to the apo A3A crystal structure, the substrate residue T at position –1 basically occupies the position of the side chains of Trp98 and Tyr130 and Tyr132. Trp98 is held in place by a cation π-bond to Arg128. In the relaxed model, where all restraints are removed, the indole ring of Trp98 has flipped by almost 180°. The nucleobase of T(−1) is then sandwiched by Trp98 on one side and Tyr130 on the other. Interestingly, the residue corresponding to Tyr130 in A3A of the N-terminal domain of primate A3G (Tyr124) (pdb code: 5K83) has been proposed to work as a switch between the substrate-bound and -unbound pro-

tein conformations (36). When looking at the triplet dTTC, it can be observed that all nucleobases are positioned in such a way that the planar aromatic rings are almost in the same plane, as if positioned for sliding of substrate through the 'alley' formed by Trp 98 and Try 130. However, sliding of the substrate through the enzyme in this structure would be prevented by the strong bend at dA(+1), which requires a 180° flip for the nucleotide that exits the catalytic site.

The comparison between the structures of apo protein and its ssDNA complex (Supplementary Figure S6) shows different positions for loops 1, 3 and 7. The binding of the substrate triplet dTTC to the ligand-free protein, especially the T(−1) between Trp98 and Tyr130, requires the movement of Arg128 as positioned in the crystal structure. Interestingly, in the A3A solution structure (pdb code: 2M65) this arginine is distant from Trp98, but has a similar position as in our ssDNA–A3A complex model. Also Asn24 in the crystal structure forms a hydrogen bond with helix 6 (Trp176 and Leu179). This interaction holds loop 1 in place, closing the gap between helix 1 and the N-terminal end of helix 6 and the loop preceding it. This gap is the position where our experimental data indicate the 5′-end of the substrate interacts with the protein. In order for the ssDNA to come close to the protein the hydrogen bonds of Asn24 have to be broken, so that the whole of loop 1 can move towards loop 3, thereby opening the gap to allow ssDNA to bind. Interestingly, about half the structures of the NMR ensemble (2M65) do not have Asn24 forming a hydrogen bond with helix 6.

Dynamic insights into DNA binding and activity

The negative change in entropy upon binding, seen in the above ITC experiment, correlates with the greatly reduced freedom of movement of the bound substrate compared to the free substrate in solution. According to our model, the bound substrate is forced into a kinked conformation, which is very rarely sampled in solution. In addition, the negative entropy change may indicate reduction in flexibility of A3A loops.

The flexibility of a protein can be assessed using a random coil index (RCI), obtained by analysis of chemical shifts (http://randomcoilindex.com/cgi-bin/rci_cgi_current.py) (49).

The RCI analysis of all APOBECs (not just of A3 proteins) for which chemical shifts have been deposited in the BMRB databank revealed a difference in the dynamical pattern between active (that is having deaminase activity) and inactive APOBEC proteins for which deaminase activity could not be detected (Figure 9).

Active proteins have loops with a remarkably similar pattern of flexibility, where the flexibility increases from loop 1 over loop 2 to the most flexible loop 3 (Figure 9). Inactive proteins, such as mouse APOBEC2 (human APOBEC2 shows a similar pattern but is not shown) and the human A3G N-terminal domain show the opposite trend, with loop 1 being the most flexible and the movement steadily decreases for loops 2 and 3, respectively (Figure 9). From our point of view this indicates that concerted motion of loops 1, 2 and 3 may be necessary for deamination. Overall, our data indicate that dynamics are important in the deamination reaction.

DISCUSSION

The quick and robust thermal shift assay was applied to access the binding of ssDNA to A3A–E72A protein. Thermal shift assays are very simple, consume few resources and can be used for high-throughput screening, provided that the fluorescence probe binds differentially to folded and unfolded protein.

A new method for developing models of protein-oligonucleotide interactions was developed, where ^{15}N-HSQC spectra of pairs of subtly different ssDNA–A3A–E72A complexes were compared. The protein titration with RNA sequences differing in length by one nucleotide was used by Wurm *et al*. (50) to determine the position of a 3′-nucleotide, but to the best of our knowledge our work is the first to perform such a comparison systematically to locate an entire oligonucleotide. The smaller the change in the ssDNA ligand, the more informative were the recorded chemical shift perturbations of the protein, allowing proximity of specific nucleotide and amino-acid residues to be precisely deduced. The method uses the chemical shifts comparisons between ^{15}N-HSQC spectra and does not require hunting for specific NOEs and the assignment of specific proton–proton distance restraints, which are frequently difficult to obtain for weak binding. Unexpectedly, small changes in the 3′-part of the dA_5TTCA_5 substrate caused chemical shift perturbations in loops 1 and 7 for which chemical-shift perturbations had also been observed when changes were made to the 5′-end of the ssDNA oligomer. These data support a ssDNA/A3A interaction model, similar to TadA RNA adenosine deaminase, where here the single-stranded nucleic acid is folded back on itself. When the chemical shift perturbations were used as distance restraints in molecular dynamics simulations, a molecular model of ssDNA bound A3A–E72A could be developed.

The ssDNA is clearly positioned in the protein and adopts a structure resembling a '7', where the 5′-end forms the base of the digit, the 2′-deoxycytidine is present at the corner, inserted into the active site and, thereby very accessible to deamination, and the 3′-part folds back toward the 5′-part around the target cytosine. The position of the 5′-part is consistent with previously available biochemical data, where loop 7 and loop1 are involved in recognition of −1 and −2 residues (31,44–48).

APOBEC3 proteins bind ssDNA in two modes: the sequence-specific enzymatically susceptible active form and a non-specifically bound form. While significant product inhibition has not been reported, both binding modes have rather similar binding constants (25,51). Our NMR-titration data comparing substrate-bound A3A–E72A to product-bound A3A–E72A reveal some chemical shift differences between the two complexes, located at and around the active site. Therefore it is likely, that ssDNA without the target 2′-deoxycytidine in the recognition motive binds similarly to the protein and thus the sequence-specific and non-specific binding modes differ mainly in the binding of dC and T(−1).

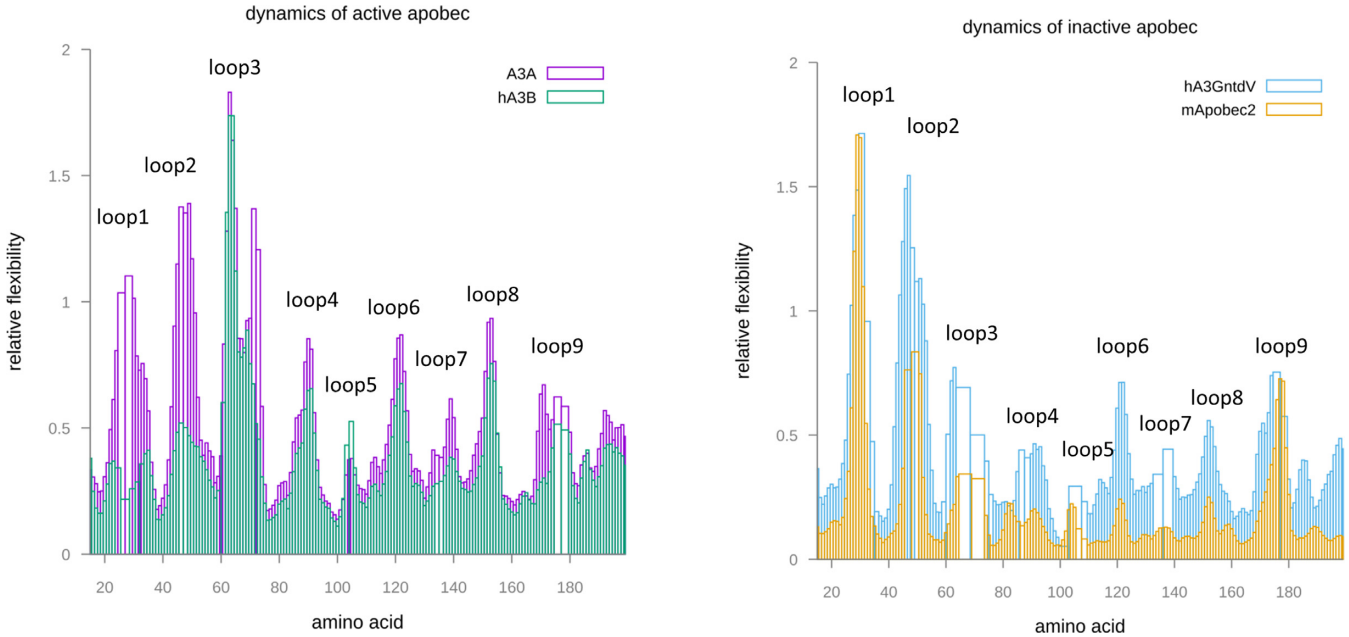

Figure 9. Dynamics of active and inactive deaminases. Comparison of scaled RCI (random coil chemical indices) values as a marker of relative flexibility of active (left, A3A (30) magenta, A3B C-terminal domain (31) green) and inactive (right, A3G N-terminal domain variant (34) blue, mAPOBEC2, pdb code 2RPZ, yellow) proteins, showing different dynamic pattern. Loop numbers are indicated.

Only when the correct hydrogen bonds are formed, can the substrate insert deeper into the enzyme and deamination can occur.

During the review process, the X-ray structures of A3A (5sww at 3.15 Å resolution) and the C-terminal domain of A3B (5td5 at 1.72 Å resolution), both proteins in complex with ssDNA appeared (37) and coordinates were released. For A3A, the central five or six nucleotides, depending on subunit, of the 13-mer were observed; for A3B four of the nucleotides of the pentamer were observed. Since the development of our model was unbiased by knowledge of these structures, this has then provided the highest possible level of validation of our 'method of small changes'. The comparison of our model to the A3A–ssDNA structure shows very similar overall shape (Figure 10). The directions of bases of three nucleotides around the target cytosine are very similar in our model and in the crystal structures (Supplementary Figure S11). Many of the interactions are between the same partners in crystal structures and in our model (Supplementary Table S8). The differences outside of the TCA motive may be due to the different length and composition of the substrates in the model and in the structure. Note, that the 5′- and 3′-ends of the oligonucleotide are flexible in the crystal structures (Supplementary Figure S10) and in our molecular dynamic simulations (Supplementary Figure S7).

Our RCI analysis suggests the hypothesis that concerted movement of loops is related and probably pivotal to deamination. Alternatively, loop 1 which is more flexible in non-active proteins, is located close to the helix 1. Properly positioned helix 1 has been shown to stabilize the active site (27). Whereas large differences (such as apo versus liganded structure) lead to chemical shift changes at sites remote to the actual binding site of the ligand, here this method

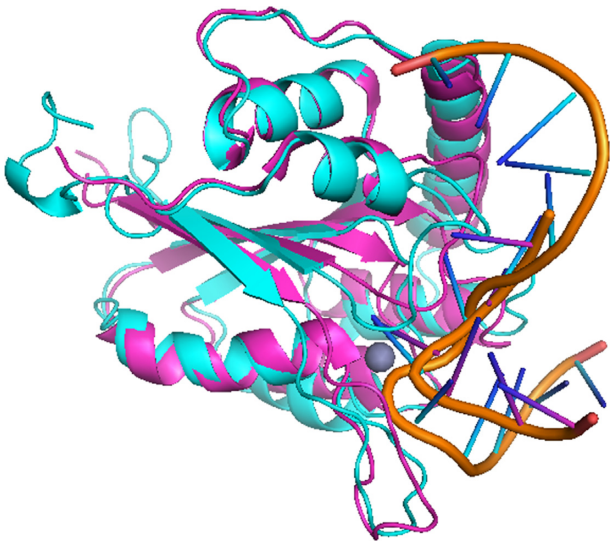

Figure 10. Superposition of the average NMR-MD model (cyan and long oligonucleotide) onto the X-ray structure (5sww; purple and short oligonucleotide). Although loop 3 appears to differ in conformation, the X-ray conformation is approximately sampled in the MD simulations. Nucleotides at positions +1, 0 and −1 superimpose relatively closely. The difference in positions of nucleobases moving away from the active-site nucleobase position 0 toward the 5′ end may be a consequence of the longer oligonucleotide used in NMR experiments and modelling.

of small changes allows sites of interaction occurring between nucleotide and protein to be sensitively and precisely probed. Given the challenges of co-crystallising protein-DNA complexes coupled with the relatively inexpensive cost of buying a suite of oligonucleotides, the method de-

scribed here may prove useful to developing molecular models of solution-state protein–oligonucleotide structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Health Research Council HRC 14/629, Worldwide Cancer Research 16-1197 and Massey University Research fund (MURF 2015, 7003) for funding, the Massey University bio-NMR facility for the NMR time, Reuben Harris and his Cancer Program members for critical discussions and reagents, structural biology seminar members for helpful discussions and Hiroshi Matsuo for synthesizing DNA for protein expression.

FUNDING

Health Research Council of New Zealand [HRC 14/629]; Worldwide Cancer Research [16-1197]; Massey University Research Fund [7003] from Massey University. Funding for open access charge: Worldwide Cancer Research and Institute of Fundamental Sciences, Massey University.
Conflict of interest statement. None declared.

REFERENCES

1. Harris,R.S. and Liddament,M.T. (2004) Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.*, **4**, 868–877.
2. Izumi,T., Shirakawa,K. and Takaori-Kondo,A. (2008) Cytidine deaminases as a weapon against retroviruses and a new target for antiviral therapy. *Mini Rev. Med. Chem.*, **8**, 231–238.
3. Harris,R.S., Bishop,K.N., Sheehy,A.M., Craig,H.M., Petersen-Mahrt,S.K., Watt,I.N., Neuberger,M.S. and Malim,M.H. (2003) DNA deamination mediates innate immunity to retroviral infection. *Cell*, **113**, 803–809.
4. Stenglein,M., Burns,M., Li,M., Lengyel,J. and Harris,R. (2010) APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.*, **17**, 222–229.
5. Bogerd,H.P., Wiegand,H.L., Hulme,A.E., Garcia-Perez,J.L., O'Shea,K.S., Moran,J.V. and Cullen,B.R. (2006) Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 8780–8785.
6. Warren,C.J., Xu,T., Guo,K., Griffin,L.M., Westrich,J.A., Lee,D., Lambert,P.F., Santiago,M.L. and Pyeon,D. (2014) APOBEC3A functions as a restriction factor of human papillomavirus. *J. Virol.*, **89**, 688–702.
7. Chen,H., Lilley,C.E., Yu,Q., Lee,D.V., Chou,J., Narvaiza,I., Landau,N.R. and Weitzman,M.D. (2006) APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.*, **16**, 480–485.
8. Burns,M.B., Lackey,L., Carpenter,M.A., Rathore,A., Land,A.M., Leonard,B., Refsland,E.W., Kotandeniya,D., Tretyakova,N., Nikas,J.B. *et al.* (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*, **494**, 366.
9. Burns,M.B., Temiz,N.A. and Harris,R.S. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.*, **45**, 977–983.
10. Pham,P., Landolph,A., Mendez,C., Li,N. and Goodman,M.F. (2013) A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *J. Biol. Chem.*, **288**, 29294–29304.
11. Mussil,B., Suspene,R., Aynaud,M.M., Gauvrit,A., Vartanian,J.P. and Wain-Hobson,S. (2013) Human APOBEC3A isoforms translocate to the nucleus and induce DNA double strand breaks leading to cell stress and death. *PLoS One*, **8**, e73641.
12. Kuong,K.J. and Loeb,L.A. (2013) APOBEC3B mutagenesis in cancer. *Nat. Genet.*, **45**, 964–965.
13. Lackey,L., Law,E.K., Brown,W.L. and Harris,R.S. (2013) Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination. *Cell Cycle*, **12**, 762–772.
14. Suspène,R., Aynaud,M.M., Guétard,D., Henry,M., Eckhoff,G., Marchio,A., Pineau,P., Dejean,A., Vartanian,J.P. and Wain-Hobson,S. (2011) Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 4858–4863.
15. Starrett,G.J., Luengas,E.M., McCann,J.L., Ebrahimi,D., Temiz,N.A., Love,R.P., Feng,Y., Adolph,M.B., Chelico,L., Law,E.K. *et al.* (2016) The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat. Commun.*, **7**, 12918.
16. Law,E.K., Sieuwerts,A.M., LaPara,K., Leonard,B., Starrett,G.J., Molan,A.M., Temiz,N.A., Vogel,R.I., Meijer-van Gelder,M.E., Sweep,F.C.G.J. *et al.* (2016) The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.*, **2**, e1601737.
17. Huthoff,H. and Malim,M.H. (2005) Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. *Virology*, **334**, 147–153.
18. Chung,S.J., Fromme,J.C. and Verdine,G.L. (2005) Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.*, **48**, 658–660.
19. Betts,L., Xiang,S., Short,S.A., Wolfenden,R. and Carter,C.W. Jr (1994) Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.*, **235**, 635–656.
20. Teh,A.-H., Kimura,M., Yamamoto,M., Tanaka,N., Yamaguchi,I. and Kumasaka,T. (2006) The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse‡. *Biochemistry*, **45**, 7825–7833.
21. Johansson,E., Mejlhede,N., Neuhard,J. and Larsen,S. (2002) Crystal structure of the tetrameric cytidine deaminase from Bacillus subtilis at 2.0 Å resolution. *Biochemistry*, **41**, 2563–2570.
22. Xiang,S., Short,S.A., Wolfenden,R. and Carter,C.W. (1996) Cytidine deaminase complexed to 3-deazacytidine: a 'valence buffer' in zinc enzyme catalysis†. *Biochemistry*, **35**, 1335–1341.
23. Lu,X., Zhang,T., Xu,Z., Liu,S., Zhao,B., Lan,W., Wang,C., Ding,J. and Cao,C. (2014) Crystal structure of DNA cytidine deaminase ABOBEC3G catalytic deamination domain suggests a binding mode of full-length enzyme to single-stranded DNA. *J. Biol. Chem.*, **290**, 4010–4021.
24. Ko,T.-P., Lin,J.-J., Hu,C.-Y., Hsu,Y.-H., Wang,A.H.-J. and Liaw,S.-H. (2003) Crystal structure of yeast cytosine deaminase: insights into enzyme mechanism and evolution. *J. Biol. Chem.*, **278**, 19111–19117.
25. Chen,K.M., Harjes,E., Gross,P.J., Fahmy,A., Lu,Y., Shindo,K., Harris,R.S. and Matsuo,H. (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature*, **452**, 116–119.
26. Furukawa,A., Nagata,T., Matsugami,A., Habu,Y., Sugiyama,R., Hayashi,F., Kobayashi,N., Yokoyama,S., Takaku,H. and Katahira,M. (2009) Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.*, **28**, 440–451.
27. Harjes,E., Gross,P.J., Chen,K.M., Lu,Y., Shindo,K., Nowarski,R., Gross,J.D., Kotler,M., Harris,R.S. and Matsuo,H. (2009) An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J. Mol. Biol.*, **389**, 819–832.
28. Holden,L.G., Prochnow,C., Chang,Y.P., Bransteitter,R., Chelico,L., Sen,U., Stevens,R.C., Goodman,M.F. and Chen,X. (2008) Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*, **456**, 7121–7214.
29. Kitamura,S., Ode,H., Nakashima,M., Imahashi,M., Naganawa,Y., Kurosawa,T., Yokomaku,Y., Yamane,T., Watanabe,N. and Suzuki,A. (2012) The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.*, **19**, 1005–1010.
30. Byeon,I.J., Ahn,J., Mitra,M., Byeon,C.H., Hercik,K., Hritz,J., Charlton,L.M., Levin,J.G. and Gronenborn,A.M. (2013) NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.*, **4**, 1890.
31. Byeon,I.-J.L., Byeon,C.-H., Wu,T., Mitra,M., Singer,D., Levin,J.G. and Gronenborn,A.M. (2016) NMR structure of the APOBEC3B catalytic domain: structural basis for substrate binding and DNA deaminase activity. *Biochemistry*, **55**, 2944–2959.

32. Shi,K., Carpenter,M.A., Kurahashi,K., Harris,R.S. and Aihara,H. (2015) Crystal structure of the DNA deaminase APOBEC3B catalytic domain. *J. Biol. Chem.*, **290**, 28120–28130.
33. Bohn,M.F., Shandilya,S.M., Albin,J.S., Kouno,T., Anderson,B.D., McDougle,R.M., Carpenter,M.A., Rathore,A., Evans,L., Davis,A.N. *et al.* (2013) Crystal structure of the DNA cytosine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. *Structure*, **21**, 1042–1050.
34. Kouno,T., Luengas,E.M., Shigematsu,M., Shandilya,S.M., Zhang,J., Chen,L., Hara,M., Schiffer,C.A., Harris,R.S. and Matsuo,H. (2015) Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. *Nat. Struct. Mol. Biol.*, **22**, 485–491.
35. Shandilya,S.M., Nalam,M.N., Nalivaika,E.A., Gross,P.J., Valesano,J.C., Shindo,K., Li,M., Munson,M., Royer,W.E., Harjes,E. *et al.* (2010) Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. *Structure*, **18**, 28–38.
36. Xiao,X., Li,S.X., Yang,H. and Chen,X.S. (2016) Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat. Commun.*, **7**, 12193.
37. Shi,K., Carpenter,M.A., Banerjee,S., Shaban,N.M., Kurahashi,K., Salamango,D.J., McCann,J.L., Starrett,G.J., Duffy,J.V., Demir,O. *et al.* (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.*, **24**, 131–139.
38. Hill,J.M. (2008) In: Kobe,B, Guss,M and Huber,T (eds.) *Structural Proteomics: High-Throughput Methods*. Humana Press, Totowa, pp. 437–446.
39. Krieger,E., Koraimann,G. and Vriend,G. (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Struct. Funct. Bioinformatics*, **47**, 393–402.
40. 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**, 114–122.
41. Losey,H.C., Ruthenburg,A.J. and Verdine,G.L. (2006) Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.*, **13**, 153–159.
42. Love,R.P., Xu,H. and Chelico,L. (2012) Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *J. Biol. Chem.*, **287**, 30812–30822.
43. Semisotnov,G.V., Rodionova,N.A., Razgulyaev,O.I., Uversky,V.N., Gripas,A.F. and Gilmanshin,R.I. (1991) Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, **31**, 119–128.
44. Logue,E.C., Bloch,N., Dhuey,E., Zhang,R., Cao,P., Herate,C., Chauveau,L., Hubbard,S.R. and Landau,N.R. (2014) A DNA sequence recognition loop on APOBEC3A controls substrate specificity. *PLoS One*, **9**, e97062.
45. Kohli,R.M., Abrams,S.R., Gajula,K.S., Maul,R.W., Gearhart,P.J. and Stivers,J.T. (2009) A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J. Biol. Chem.*, **284**, 22898–22904.
46. Wang,M., Rada,C. and Neuberger,M.S. (2010) Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. *J. Exp. Med.*, **207**, 141–153.
47. Rathore,A., Carpenter,M.A., Demir,Ö., Ikeda,T., Li,M., Shaban,N.M., Law,E.K., Anokhin,D., Brown,W.L., Amaro,R.E. *et al.* (2013) The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution. *J. Mol. Biol.*, **425**, 4442–4454.
48. Carpenter,M.A., Rajagurubandara,E., Wijesinghe,P. and Bhagwat,A.S. (2010) Determinants of sequence-specificity within human AID and APOBEC3G. *DNA Repair*, **9**, 579–587.
49. Berjanskii,M.V. and Wishart,D.S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.*, **127**, 14970–14971.
50. Wurm,J.P., Meyer,B., Bahr,U., Held,M., Frolow,O., Kötter,P., Engels,J.W., Heckel,A., Karas,M., Entian,K.-D. *et al.* (2010) The ribosome assembly factor Nep1 responsible for Bowen–Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res.*, **38**, 2387–2398.
51. Harjes,S., Solomon,W.C., Li,M., Chen,K.M., Harjes,E., Harris,R.S. and Matsuo,H. (2013) Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J. Virol.*, **87**, 7008–7014.