9872–9880 *Nucleic Acids Research, 2016, Vol. 44, No. 20* *Published online 9 September 2016*
doi: 10.1093/nar/gkw799

Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method

Yuru Wang and Peter A. Beal*

Department of Chemistry, University of California, One Shields Ave, Davis, CA 95616, USA

Received August 2, 2016; Revised August 30, 2016; Accepted August 31, 2016

ABSTRACT

Adenosine deamination is one of the most prevalent post-transcriptional modifications in mRNA. In humans, ADAR1 and ADAR2 catalyze this modification and their malfunction correlates with disease. Recently our laboratory reported crystal structures of the human ADAR2 deaminase domain bound to duplex RNA revealing a protein loop that binds the RNA on the 5′ side of the modification site. This 5′ binding loop appears to be one contributor to substrate specificity differences between ADAR family members. In this study, we endeavored to reveal detailed structure–activity relationships in this loop to advance our understanding of RNA recognition by ADAR2. To achieve this goal, we established a high-throughput mutagenesis approach which allows rapid screening of ADAR variants in single yeast cells and provides quantitative evaluation for enzymatic activity. Using this approach, we determined the importance of specific amino acids at 19 different positions in the ADAR2 5′ binding loop and revealed six residues that provide essential structural elements supporting the fold of the loop and key RNA-binding functional groups. This work provided new insight into RNA recognition by ADAR2 and established a new tool for defining structure–function relationships in ADAR reactions.

INTRODUCTION

Modified nucleosides in mRNA have received considerable attention recently because of new discoveries related to their abundance (1,2), mechanism of incorporation and removal (1–3) and biological function (3–5). A commonly occurring modified nucleoside in human mRNA is inosine (I), the deamination product of adenosine (A) (6). Because hydrolytic deamination of A changes its base-pairing properties (i.e. I base pairs like guanosine (G)), the A to I reaction can change the meaning of codons in mRNA. Since coding properties of the RNA can be altered by this reaction, adenosine deamination is a type of RNA editing. Two different enzymes carry out A to I editing in humans, ADAR1 and ADAR2 (7,8). ADARs 1 and 2 are expressed in most tissues, whereas a related protein referred to as ADAR3 is expressed exclusively in the brain (9,10). RNA editing malfunction can lead to human disease (11–15). For instance, mutations in the human ADAR1 gene are now known to cause the inherited skin disorder dyschromatosis symmetrica hereditaria and the autoimmune disease Aicardi–Goutieres syndrome (12–15).

ADARs specifically edit certain adenosines over others and ADAR1 and ADAR2 have overlapping yet distinct editing specificity (16–20). For example, in the human glutamate receptor subunit B (GluR B) mRNA, the Q/R site is specifically edited by ADAR2 while the hotspot 1 site is specifically edited by ADAR1. The R/G site in the GluR B mRNA is edited by both ADAR1 and ADAR2 (19,20). Another human substrate, hNEIL1 RNA, is specifically edited by ADAR1 at the K242R recoding site (21).

The ADAR proteins are modular with clearly definable deaminase and RNA-binding domains (22). Double stranded RNA-binding domains (dsRBDs) present in ADARs facilitate interaction with substrate RNAs with duplex secondary structure (23). dsRBD recognition of duplex RNA is well understood (24–26). However, less well characterized is the RNA recognition surface of the ADAR deaminase domains. Recently, we reported crystal structures of the human ADAR2 deaminase domain bound to substrate RNAs that revealed RNA-binding loops near the enzyme active site (27). These studies showed the loop in hADAR2 spanning amino acids 454–479 contacts the RNA ∼10 bp 5′ from the editing site, mainly by interacting with the phosphodiester backbone (Figure 1A and B). Certain residues in this loop are highly conserved in ADAR2 enzymes from different organisms indicating an important role for the loop in ADAR2 function (Figure 1A). Interestingly, hADAR1 and hADAR3 sequences deviate substantially from the hADAR2 sequence in this region suggesting that functional differences among the ADARs may arise, at least in part, from differences in this loop (27). Given its significance in ADAR2 and the larger ADAR family, we endeavored to determine the importance of specific amino acids at different positions in the loop for an efficient RNA

*To whom correspondence should be addressed. Tel: +1 530 752 4132; Fax: +1 530 752 8995; Email: pabeal@ucdavis.edu

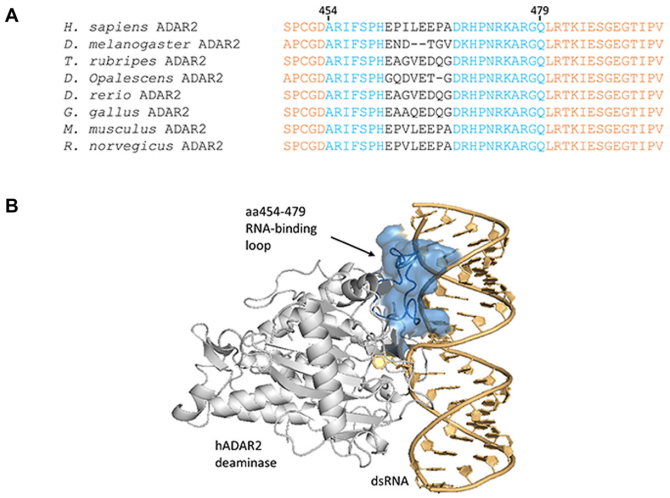


Figure 1. 5′ RNA binding loop of hADAR2. (**A**) Alignment of sequences of the ADAR2 5′ binding loop from different organisms. The conserved residues in the RNA binding loop are colored in blue with residues in the variable region black. (**B**) Crystal structure of hADAR2 deaminase domain bound to dsRNA substrate (27). The 5′ RNA binding loop is highlighted in blue.

editing reaction. To achieve this goal, we developed a new yeast-enhanced green fluorescent protein (yeGFP) reporter that allows one to monitor ADAR activity in single yeast cells. We then used this reporter, in combination with saturation mutagenesis (Sat), fluorescence-activated cell sorting (FACS) and next-generation sequencing (Seq), to carry out a high-throughput mutagenesis study of the hADAR2 5′ binding loop. These efforts provided insight into the substrate recognition by the RNA editing ADAR enzymes and established the Sat-FACS-Seq approach as a useful new tool in defining structure–function relationships for ADAR reactions.

MATERIALS AND METHODS

General methods

Components for yeast culture media were purchased from BTS (biological grade peptone, yeast extracts, biological grade agar), BD (yeast nitrogen base w/o amino acid), Sigma-Aldrich (glucose, glycerol, lactate, galactose, raffinose, amino acids, yeast synthetic dropout supplement medium without uracil, yeast synthetic dropout medium without tryptophan). Phusion hot start DNA polymerase was purchased from Thermo Scientific for high fidelity DNA amplification. Restriction enzymes, T4 ligase and other cloning enzymes were obtained from New England Biolabs. Ligation products were transformed into XL 10 Gold *Escherichia coli* cells (Agilent Technologies). The sequences of plasmids or plasmid libraries were confirmed by Sanger sequencing. Individual plasmids were transformed into yeast using standard lithium acetate methods. Plasmid libraries were transformed into yeast using a high efficiency lithium acetate protocol (28). *E. coli* cells were grown in LB media with ampicillin selection at 37°C with shaking at 300 rpm. Yeast cells were grown in yeast extract peptone dextrose media (YPD) or appropriate dropout media at 30°C with shaking at 270–300 rpm. Plasmid extraction from *E. coli* was carried out with QIAprep Spin Miniprep kit (Qiagen). Plasmid extraction from yeast cells was carried out with glass beads from Sigma-Aldrich and QIAprep Spin Miniprep kit (Qiagen) following manufacturer's protocol. Oligonucleotides were synthesized by Integrated DNA Technology.

yeGFP reporter construction

Plasmid pGADT7-ADH700-yCherry-pTEF1(ashbya)-yeGFP-DHFR was purchased from Addgene. From the plasmid, the region encoding yeGFP was amplified by polymerase chain reaction (PCR) using forward primer BDF2-yeG-1, which contains BDF2 substrate sequence, and reverse primer yeGFP-reverse, which contains XhoI restriction site. A second PCR was then performed to introduce HindIII restriction site at 5′ of the previous PCR product using forward primer BDF2-yeG-2 and the same reverse primer. The sequence was then inserted into pYES3/CT (Invitrogen) by standard cloning procedures. This plasmid has a *GAL1* promotor, a *TRP1* auxotrophic marker for selection of transformed yeast and an ampicillin resistance marker for selection of transformed *E. coli*. Full sequence of the gene in the plasmid is found in Supplementary Figure S1.

Yeast culture

Saccharomyces cerevisiae INVSc1 strain (Invitrogen) was used in this study. The yeast strain was sequentially transformed with the yeGFP reporter plasmid with *TRP1* selection followed by hADAR-D expression plasmid (29) or libraries with *URA3* selection. For the individual hADAR-D plasmid, after transformation, yeast cells were plated onto complete media (CM) – ura – trp + 2% glucose plate for growth. Individual colonies were then harvested and used to inoculate 5 ml CM – ura – trp + 2% glucose media for overnight growth. The resulting culture (0.1 ml) was used to inoculate 20 ml CM – ura – trp + 3% glycerol + 2% lactate media for another growth until the OD_{600} of the culture reaches 1–2 (∼36 h). Galactose was then added to a final concentration of 3% for induction. For transforming hADAR-D plasmid libraries, an upgraded high efficiency lithium protocol was used (28). The transformation mixture was used to inoculate 10 ml CM – ura – trp + 2% glucose media. After 24–48 h of growth (until OD_{600} = 7–8; time taken depends on the transformation scale and efficiency), 0.25 ml glucose culture was used to inoculate 25 ml CM – ura – trp + 3% glycerol + 2% lactate media. After an additional 36 h of growth (OD_{600} = 1–2), galactose was added to a final concentration of 3% for induction. After induction, cells were collected, washed twice with phosphate buffered saline (PBS), re-suspended with PBS and analyzed or stored at −70 as cell pellets for future analysis.

Fluorescence detection and FACScan

Cells collected after induction were washed twice and re-suspended with PBS to appropriate density for microscopy imaging, fluorescence measurement and FACS analysis.

Cell visualization was performed on Leica DMI3000 B Manual Inverted Microscope under the setting for GFP. Images were analyzed with ImageJ. When detecting fluorescence with a multi-well plate reader, the same amount of cells (10^7-10^8) re-suspended in 100 μl PBS was measured in an Optiplate-96 black, black opaque 96-well microplate (PerkinElmer) using a CLARIOstar plate reader (BMG labtech), with excitation at 482/16 nm and emission at 520/10 nm. FACScan analysis was performed with Becton Dickinson FACScan at UC Davis flow cytometry shared resource laboratory.

Saturation mutagenesis for ADAR2 loop

DNA oligonucleotides used for carrying out saturation mutagenesis contained degenerate codons (NNS) and a silent mutation (positional barcode) immediately 5′ of the NNS codon. Quick change II XL site-directed mutagenesis kit (Agilent) was used to generate mutant libraries at desired sites in hADAR2-D E488Q (in YEpTOP2PGAL1). The resulting PCR products were transformed into XL 10 Gold *E. coli* cells following manufacturer's protocol. For each library, at least 300 *E. coli* colonies were pooled to ensure coverage at least 10 times greater than the library diversity.

Plasmid samples from each saturation mutagenesis library in equal amounts were pooled to make the start library. The yeGFP reporter plasmid and the start library were sequentially transformed in yeast INVSc1 strain. Three parallel transformations were carried out, each with 2×10^8 cells and 3 μg plasmid DNA from the start library. Cells (1/200) from each transformation were plated onto a selection plate to determine the transformation efficiency, while the remaining cells were combined and used to inoculate 10 ml CM – ura – trp + 2% glucose media for growth. Catalytically inactive hADAR2-D mutant E396A was used as negative control and processed in the same way. Cells after 20 h of induction were collected for cell sorting analysis.

Cell sorting

Cells expressing hADAR2-D E396A or the hADAR2-D loop library were diluted in PBS to 20 000 cells/μl. Cells were sorted using Beckman Coulter Astrios EQ cell sorter at UC Davis flow cytometry shared resource laboratory. GFP excitation was at 488 nm and emission at 529/28 nm. Cells expressing hADAR2-D mutants were collected in eppendorf tubes corresponding to each gate (R1–R5). More than 150 000 cells were collected for each gate. Collected cell samples were each used to inoculate 2 ml CM – ura – trp + 2% glucose media in a BD falcon polypropylene round-bottom tube and cultured for approximately 24 h, after which the 2 ml cultures were used to inoculate 11 ml fresh CM – ura – trp + 2% glucose media for another 24 h growth. Cells were then pelleted and plasmids were isolated from each cell pellet. Contour plots creation and data analysis were performed using FlowJo.

Library preparation and Illumina sequencing

The sequence in hADAR2 spanning amino acids 445-489 (135 nt) was amplified from the plasmid libraries collected using two-step PCR. The first PCR used primers containing from 5′ to 3′ an overhang adapter sequence and amplicon-specific sequence. The PCR products were gel purified and underwent a second PCR which uses primers containing from 5′ to 3′ a leader sequence for Illumina sequencing, an 8 bp barcode (specific to each sample) and a sequence overlapping with the adapter sequence in the primers used in the first PCR. The resulting PCR products were gel purified and concentrations were determined by using a bioanalyzer. The PCR products were pooled in equal amounts and sequenced in a single run using Illumina Miseq.

Sequencing data processing

Paired-end reads (80-base) from Miseq were first decomplexed according to the sample specific barcodes. Within the data for each sample, reads were trimmed on both sides with a quality cutoff of 20 and then reads whose average quality score after trimming was below 30 were discarded. The resulting reads were merged and further selected for reads of 135 nt long. For each position of variation, the abundance of each codon was determined by searching a sequence of 12 nt including the target codon, the positional barcode, plus six more nucleotides in the 5′ direction. The search was performed overall on all the data by entering command lines on the platform of MobaXterm.

The abundance of amino acids in R1 through R5 libraries were weighed against their abundance in the input library to obtain enrichment levels in each gate. Frequencies of amino acids enriched in R5 gate were calculated from their enrichment levels and plotted with Seq2 Logo 2.0 server. Average fluorescence F_{ave} for each amino acid at each position was estimated using the equation:

$$F_{ave} = \frac{\sum_{i=1}^{5} (\text{enrichment level in Ri} \times \text{mean fluorescence of Ri})}{\sum_{i=1}^{5} (\text{enrichment level in Ri})}$$

Protein purification and *invitro* deamination

ADAR protein purification and preparation of RNA substrates have been described previously (29,30). Deamination assays were performed under single turnover conditions with 1 μM enzyme and 10 nM RNA substrate under conditions of 20 mM Tris–HCl pH 7.0, 60 mM KCl, 30 mM NaCl, 1.5 mM ethylenediaminetetraacetic acid, 8% glycerol, 0.003% Nonidet P-40, 0.5 mM DTT, 0.25 mM BME, 160 U/ml RNasin and 1.0 mg/ml yeast $tRNA^{Phe}$ in a final volume of 10 μl. Each reaction solution was incubated at 30°C for 45 min before adding enzyme and the reaction was allowed to proceed for different times at 30°C prior to being stopped with 10 μl 1% sodium dodecyl sulphate and heating at 95°C for 2 min. Deaminated RNA was purified by phenol-chloroform extraction and ethanol precipitation. RT-PCR was performed on the RNA and the PCR products were purified using an agarose gel. The editing levels were then determined by Sanger sequencing of the RT-PCR products. The sequencing data were quantified using Chromas Lite (Technelysium) and ImageJ and were fit to the equation $[P]t = \alpha[1 - e^{(-kobs*t)}]$, where [P]t is the fraction of deamination product at time t, α is the fitted reaction end point and k_{obs} is the fitted rate constant using KaleidaGraph. Each experiment was carried out in triplicate.

RESULTS

A fluorescent reporter for monitoring A-to-I RNA editing in single yeast cells

We previously described a colorimetric screen using *S. cerevisiae* as the host system useful for establishing structure–function relationships in ADAR reactions (31). The assay relies on a galactosidase reporter system which renders yeast colonies colored when an editing event converts a stop codon into a tryptophan codon within the yeast cells. This assay was recently improved by employing a novel ADAR substrate RNA that reacts efficiently with both hADAR1 and hADAR2 deaminase domains (30,32). Although this assay has been successfully applied to probe different aspects of the ADAR reaction (30,31,33,34), monitoring colony color on growth plates makes screening of large libraries time consuming and labor intensive. In order to facilitate the analysis of larger libraries of ADAR-derived mutants, a different screening assay was desired. Efficient fluorescence assays have been established for application in yeast and, in combination with flow cytometry, allow analysis in single cells with high analytical sensitivity and quantitative detection (35–37). Therefore, we aimed to upgrade our plate-based screening assay to a fluorescence assay. We modified the editing reporter by replacing galactosidase with yeGFP and removing the secretion signal to retain expressed yeGFP inside the yeast cells (Figure 2A–C). The yeGFP reporter plasmid and ADAR plasmid were sequentially transformed into yeast INVSc1 strain with *TRP* and *URA* selection, respectively. After induction, cells were collected and visualized by fluorescence microscopy. Cells bearing the yeGFP reporter produced fluorescence with wild-type human ADAR1 deaminase domain (hADAR1-D) and human ADAR2 deaminase domain (hADAR2-D), indicating editing of the substrate and expression of the fluorescent protein (Figure 2D). As a negative control, no fluorescence was produced with the inactive mutant hADAR1-D E912A. In addition, we quantified fluorescence from cells expressing both the yeGFP reporter and various ADAR proteins using a standard multiwell plate reader with fluorescence detection (Figure 2E). Inactive ADAR mutants produced low fluorescence compared to suspension buffer alone; likely auto fluorescence from the yeast cells. On the other hand, cells expressing wild-type hADAR1-D or hADAR2-D showed approximately 15-fold higher fluorescence compared to cells expressing their corresponding inactive mutants (Figure 2E). To test whether the fluorescence assay is able to distinguish different levels of editing activity, we measured the fluorescence intensity of cells expressing hADAR1-D E1008Q and hADAR2-D E488Q (Figure 2E), which are known to be more active than the corresponding wild-type proteins (30,33,38). In agreement with the previously determined activity differences, both mutants produced approximately 10-fold higher fluorescence compared to their wild-type proteins (Figure 2E). In addition, the fluorescence of cells expressing wild-type hADAR2-D increased with time within the time frame tested (0–38 h) (Figure 2F), while the fluorescence of cells expressing inactive mutant E396A remained unchanged. Next, cells expressing the yeGFP reporter and hADAR2-D or the inactive mutant were analyzed by FACScan (Figure 2G). We found that for cells expressing hADAR2-D, 44% of the population had above-background fluorescence with cells expressing E396A used as a negative control to define background fluorescence levels. These results indicated that individual yeast cells expressing different levels of ADAR activity can be distinguished in this manner.

A Sat-FACS-Seq method for defining structure–activity relationships in ADAR reactions

Kohli *et al*. recently described the Sat-Sel-Seq method for defining structure–activity relationships in an enzymatic reaction (39). The approach combines saturation mutagenesis (Sat), functional selection (Sel) and next-generation sequencing (Seq) to rapidly screen libraries of mutants to probe the role of individual amino acid residues within a putative substrate-binding surface (39). Here we adapted this approach to the study of the 5′ binding loop in the hADAR2 deaminase domain using our yeGFP reporter and FACS to screen for functional mutants (Figures 1 and 3A). The 5′ binding loop of hADAR2 spanning amino acids 454–479 contains 18 conserved residues and eight sequential non-conserved residues in the 'variable region' (Figure 1A). The conservation of residues suggests their importance in protein function. Therefore, we chose to carry out site saturation mutagenesis at codons for each of the 18 conserved residues in the hADAR2 5′ binding loop. To provide a negative control for the assay, we also introduced mutations at one of the non-conserved residues, since these residues are expected to exhibit no preference in amino acid identity (Figure 3A). To maximize sensitivity of the assay, the mutations were introduced into the highly active hADAR2-D E488Q mutant (33). Mutagenic oligonucleotides encompassing degenerate NNS codon were used to introduce mutations at each position, with N representing any natural nucleotide and S representing either guanosine or cytosine. With the NNS degenerate codon, each library includes all 20 amino acids and one stop codon. Since NNS degenerate codon does not cover the entire codon diversity and different amino acids do not correspond to same number of codons, variants in each library are not equally represented, however the codon frequency can be monitored by next-generation sequencing and weighted for analysis. In each oligonucleotide, a silent mutation was placed in the codon immediately 5′ of the randomized codon to serve as an internal barcode marking the position of the original NNS codon (39) (Figure 3A). The benefit of introducing the internal barcode is to facilitate downstream analysis of next-generation sequencing data such that wild-type amino acids from sequences whose varied positions are different from the one under analysis are avoided. To assess the assay reproducibility, a duplicate library was generated for a residue randomly chosen from this loop (H471) using a different 5′ silent mutation barcode to differentiate the two libraries for the same site. A total of 20 saturation mutant libraries were generated and pooled in equal amounts to make the start library. The yeGFP reporter plasmid and the start library were sequentially transformed into *S. cerevisiae* INVSc1 strain. A small fraction of transformation mixture was plated to calculate transformation efficiency and it was

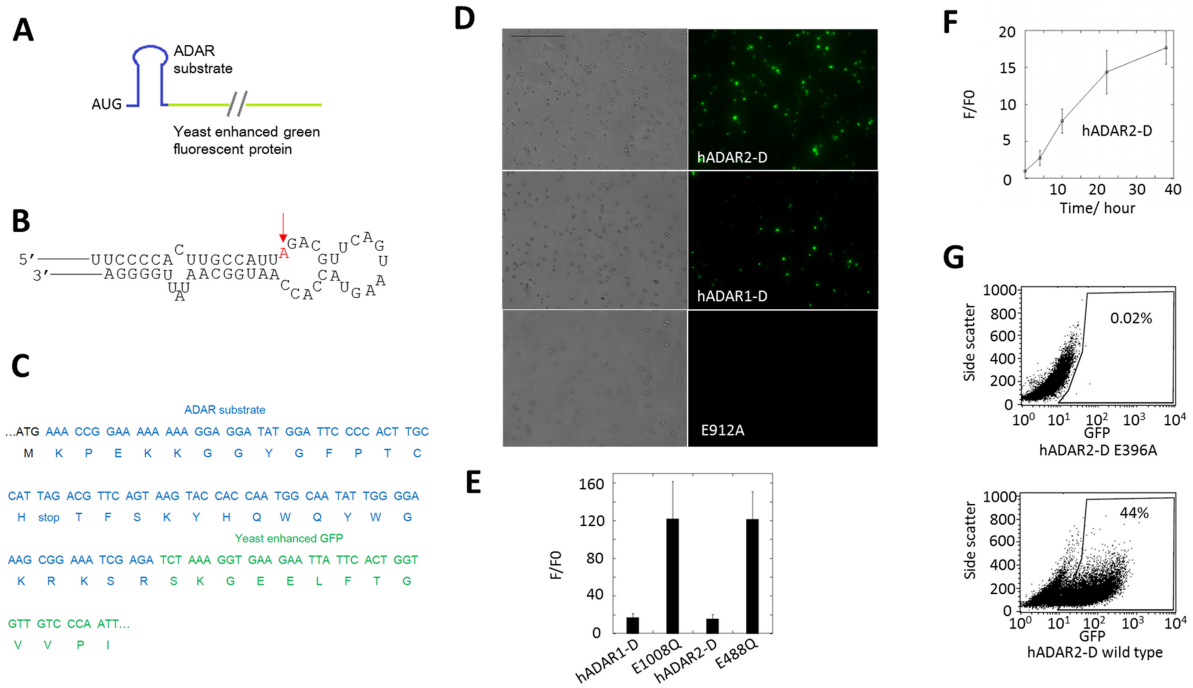
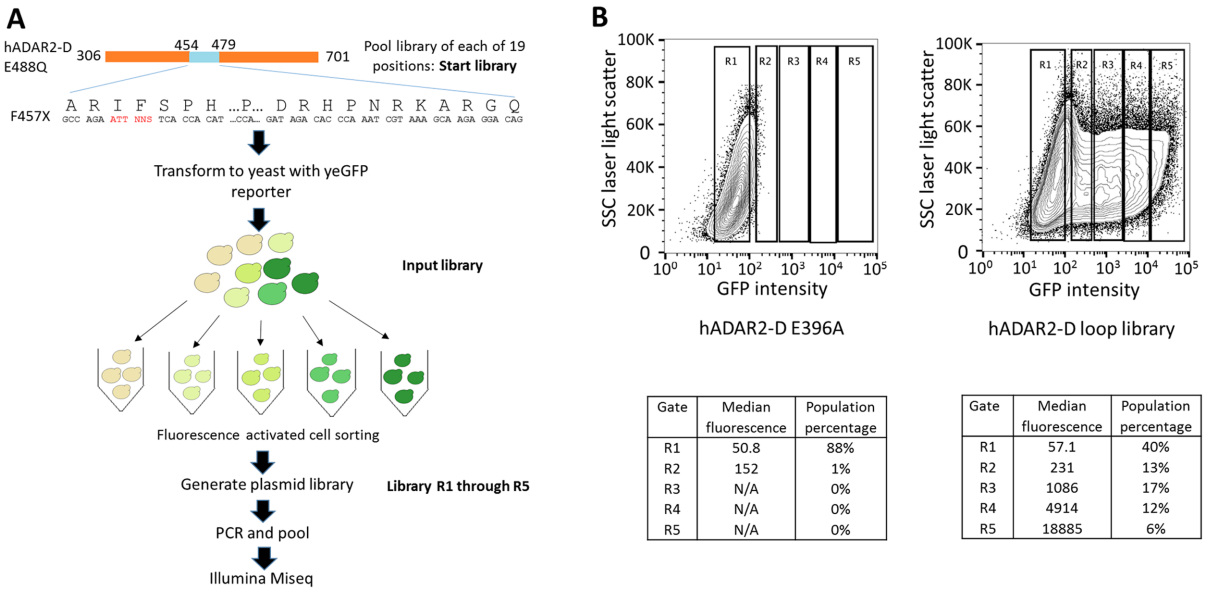


Figure 2. Design and characterization of an RNA editing reporter for single yeast cell analysis. (**A**) Schematic of reporter mRNA. (**B**) BDF2-derived sequence used as substrate with edited A indicated by an arrow (30). (**C**) Full sequence of reporter near editing site with the encoded protein sequence shown. (**D**) Images of cells expressing the yeGFP reporter and ADAR deaminase domains detected by fluorescence microscopy. Scale bar: 100 μm. (**E**) Fluorescence measurements of cell suspensions. F/F0 is the ratio of sample fluorescence divided by negative control (inactive mutant) fluorescence. Error bar indicates SD, $n \geq 3$. (**F**) Change of fluorescence of cells expressing hADAR2-D with time. Error bar indicates SD, $n \geq 3$. (**G**) FACScan analysis of cells expressing inactive E396A mutant and wild-type hADAR2-D.



Gate	Median fluorescence	Population percentage
R1	50.8	88%
R2	152	1%
R3	N/A	0%
R4	N/A	0%
R5	N/A	0%

Gate	Median fluorescence	Population percentage
R1	57.1	40%
R2	231	13%
R3	1086	17%
R4	4914	12%
R5	18885	6%

Figure 3. (**A**) Sat-FACS-Seq workflow for defining structure–function relationships in the 5′ binding loop of hADAR2. (**B**) Fluorescence activated cell sorting. Top left: Cells expressing the yeGFP reporter and hADAR2-D E396A were used to define background fluorescence and above-background fluorescence categories. Top right: Cells expressing the yeGFP reporter and hADAR2-D loop library were sorted into different categories based on different levels of fluorescence. Bottom: Parameters summarized corresponding to each sorting channel.

shown that the start library size was around 50-fold over-represented. The remaining transformation mixture was directly grown in liquid dropout media and protein expression was induced by adding galactose. Cells were collected after induction and plasmids were isolated from a portion of the cells and designated as the input library.

Cells after induction were analyzed with FACS (Figure 3B). Cells expressing inactive E396A mutant were used to define background fluorescence, and five gates, R1 through R5, were used for separating cells with various fluorescence intensities. R1 contained cells with background fluorescence and R2 through R5 had cells with above-background fluorescence, in an order of increasing fluorescence intensity from R2 to R5. Cells expressing the inactive E396A mutant mostly fell into the R1 gate. In contrast, cells expressing the loop library were found in each gate, with a population peak in R3. The mean fluorescence value of cells sorted in each gate increased by around four-fold from gate to gate. Cells distributed to each gate were collected and cultured, followed by plasmid isolation to obtain R1 through R5 libraries.

To analyze the selection for different amino acids at each position, we used PCR primers with barcode specific to each library generated along the screening process (i.e. start library, input library and R1 through R5 libraries) to amplify the region in hADAR2-D centered around the varied loop (aa445–aa489, 135 nt). The PCR products were pooled in equal amounts and sequenced in a single run using Illumina Miseq. The resulting data were decoded for the sample-specific barcode and the internal positional barcode to obtain the abundance of each amino acid for each position in each sample (Supplementary Table S1). The abundance of variants in R1 through R5 were weighed against their abundance in the input library and the resulting data were used to analyze the distributions of the variants in each fluorescence category (Supplementary Table S1).

The cells found in the R5 gate gave the highest fluorescence intensity and mutants that are enriched in this gate are expected to have the highest editing activity. Encouragingly, all wild-type amino acids are highly enriched (R5/input library ratios are around 5–7) in this gate but not in other gates. Meanwhile, within R1, which corresponds to low or no editing activity, the stop codon is highly enriched at each amino acid position (R1/ input library ratios are around 4–5) while it is depleted in other gates (Supplementary Table S1). Also as expected, all 20 amino acids were highly enriched in the R5 gate for position 462, which is in the variable region of the loop based on alignment of ADAR2 sequences from different organisms (Figure 1A). Using the enrichment levels of amino acids across the five gates and the mean fluorescence value of populations falling into each gate, we calculated an overall average fluorescence value (F_{ave}) for each amino acid at each position (Supplementary Table S1). The highest F_{ave} value at each location was normalized to 1 and the other values were adjusted accordingly. Thereby, we quantified the impact of all 20 common amino acids at each of 19 locations in the 5′ binding loop (Figure 4). To confirm that relative F_{ave} values of the variants could be an indication of their relative activities, we purified and characterized two variants, N473S and N473V, whose normalized F_{ave} values are 0.57 and 0.13 respectively, in the *in vitro* deamination assay (Supplementary Figure S2). N473S exhibited a clearly higher deamination rate than N473V under the condition tested, which agreed with their relative F_{ave} values. Importantly, the two saturation mutant libraries at H471 marked by different silent mutation barcodes showed a high degree of concordance based upon the normalized F_{ave} values of amino acids for this position, demonstrating the reproducibility of the assay (Figure 4 and Supplementary Figure S3).

To provide an integrated picture of the selection at each position within the loop, the distribution of amino acids in R5, which is highly correlated with the F_{ave} values, was expressed as a logo plot (Figure 5A). We set a cutoff for R5/input ratio to be 1 to include variants of highest editing activity. The residues under study (except for P462) are all conserved among ADAR2 sequences from different organisms (Figure 1A). However, the screening results showed that not all positions require the wild-type residue for efficient RNA editing activity. Notably, only F457, D469, H471, P472, R474 and R477 were preferred as the wild-type amino acid over all others (Figure 5A). Analysis of crystal structures of hADAR2-D bound to RNA suggests these residues provide essential structural elements supporting the fold of the 5′ binding loop and key RNA-binding functional groups (Figure 5B). For instance, residue R477, in addition to contacting a phosphodiester in the RNA backbone, stabilizes the loop structure by ion pairing with the D469 side chain. In addition, the F457 side chain provides a platform onto which R477 stacks. It appears an aromatic group is essential at this position. The next best amino acid is tyrosine, whose normalized F_{ave} value is 0.17 and all other amino acids showed normalized F_{ave} values <0.1 (Figure 4). The side chain of R474 contacts both the 5′ and 3′ phosphodiesters of a nucleotide on the unedited strand while the H471 side chain hydrogen bonds to the 5′ phosphodiester of a nucleotide found on the edited strand. This same phosphodiester is also hydrogen bonded by N473 through its main chain NH. Interestingly, the 473 position is highly tolerant of mutation consistent with its role of providing a main chain and not a side chain, interaction (Figure 5A). Another essential residue, P472, is located between RNA contact residues H471 and N473. The role of proline at this position is likely to induce the proper main chain conformation to position side chains of H471 and R474 and the main chain of N473 for RNA binding. Finally, while R470 is conserved among ADAR2s, mutation at this position to many different amino acids is tolerated in our activity screen (Figure 5A). This is consistent with the observation that the R470 side chain is pointing away from the RNA in the available structures (27). Another residue of interest in this region is S458. The corresponding residue is altered by an auto-editing reaction in *D. melanogaster* to glycine (40,41). Interestingly, the glycine mutant at this position has a normalized F_{ave} value of only 0.24, indicating low editing activity of this mutant (Figure 4).

DISCUSSION

Here, we developed a yeGFP reporter and FACS-based selection strategy capable of monitoring ADAR-catalyzed RNA editing activity in single yeast cells. We then used this

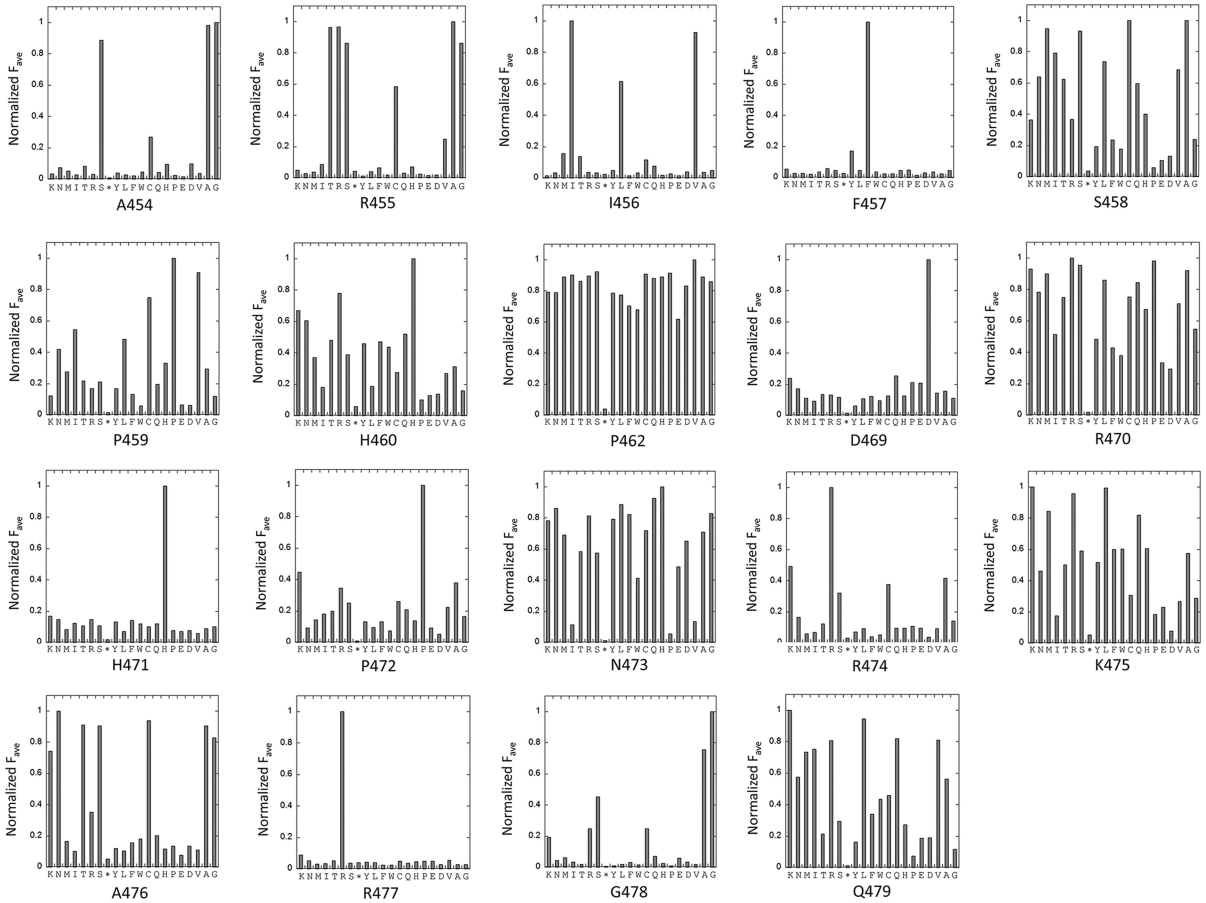


Figure 4. Normalized average fluorescence corresponding to each of the 20 common amino acids at 19 positions in the hADAR2 5′ binding loop. The star means stop codon.

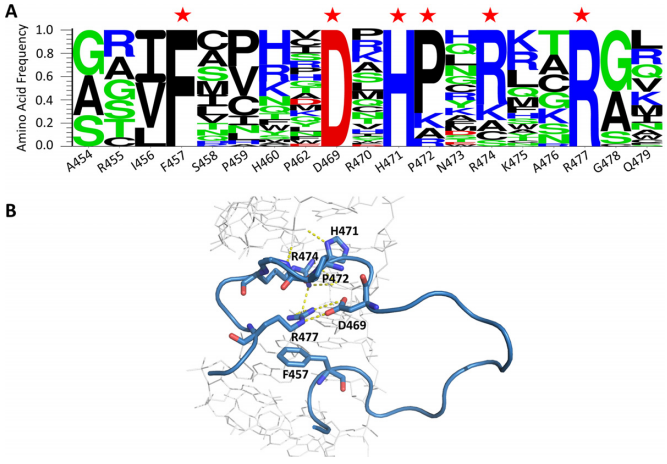


Figure 5. Functional determinants for the 5′ binding loop of hADAR2-D revealed by the screening. (**A**) Logo plot for residues 454–479 summarizing the frequency of each amino acid at each position in the most fluorescent category (R5). Positions where the wild-type residue is highly preferred are labeled with red stars. (**B**) Structural evidence showing a network of interactions among the essential residues and the substrate RNA (27).

system to carry out a high-throughput mutagenesis study (i.e. Sat-FACS-Seq) that combined saturation mutagenesis at multiple codon positions (Sat), FACS to separate yeast cells with varying levels of yeGFP fluorescence (FACS) and next-generation sequencing to identify and quantify the ADAR mutants with varying levels of activity (Seq). This approach was applied to the analysis of a recently identified RNA-binding loop present in the deaminase domain of human ADAR2 and provided new insight into the importance of the individual amino acid residues found in this loop.

The reliability of the Sat-FACS-Seq approach described here was demonstrated in multiple ways. First, wild-type amino acids were highly enriched in the most fluorescence category and have high F_{ave} values while the stop codon at each location was highly enriched in the background fluorescence category and has the lowest F_{ave} values. In addition, at position 462, which is not conserved among ADAR2 enzymes from different organisms, all variants were equally selected in the R5 category. Also, duplicate libraries generated for H471 produced concordant results. It is also worth noting that the screening results for position R455 in this study agreed well with results from a previous screen that used full length hADAR2 and our plate-based colorimetric assay with a different RNA substrate (34). However, the earlier screen only revealed the most favored amino acids at this position while the Sat-FACS-Seq approach provided a comprehensive evaluation for all 20 amino acids at this position (Figure 4).

An important observation in this study is that only 6 of the 18 conserved residues in ADAR2's 5′ binding loop require the wild-type amino acid for an efficient RNA editing reaction. Analysis of recently reported crystal structures of human ADAR2 bound to RNA showed the side chains from these six residues form contacts that stabilize the loop conformation and provide RNA-binding functional groups. The remaining loop residues appear to be conserved for other reasons, perhaps because they are needed to bind proteins involved in ADAR2 regulation or ADAR2 functions not related to adenosine deamination (42–45). It is also possible that these residues are important for binding RNA substrates that differ substantially from the type used in the screen. Indeed, the functional screen described here was carried out with an RNA editing reporter bearing an ADAR substrate RNA similar to those used in structural studies of hADAR2–RNA complexes (27). Since ADAR substrates can vary in sequence and secondary structure found 5′ to the editing site (16,17), it is possible that different substrates use different subsets of the conserved residues in the 5′ binding loop for contacting RNA. Future application of the Sat-FACS-Seq approach described here with editing reporters bearing different ADAR substrate RNAs will be helpful in addressing this intriguing possibility.

The 5′ binding loop in *D. melanogaster* ADAR contains a residue that is altered by an auto-editing reaction, i.e. dADAR deaminates its own mRNA at this codon (40). The edit converts a serine codon to one for glycine. This residue corresponds to S458 in human ADAR2 and our screening results indicate a glycine mutation at this position substantially reduces editing activity with the substrate RNA used in the screen (Figure 1A). Interestingly, studies with flies engineered to express only the edited or unedited form of dADAR show the S to G mutation at this position inhibits editing at some, but not all, known A to I editing sites in this organism (41). This observation is consistent with the idea that some ADAR substrates are highly dependent on interactions with the RNA-binding surfaces of the ADAR catalytic domain while others are less dependent on these interactions and rely more heavily on binding by the dsRBDs (32).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors acknowledge technical assistance from Bridget McLaughlin in the University of California Davis Flow Cytometry Shared Resource Laboratory and Lutz Froenicke in the University of California Davis DNA Technologies Core.

FUNDING

National Institutes of Health (NIH) [GM061115 to P.A.B.]; National Cancer Institute [P30 CA093373]; National Center for Research Resources [C06-RR12088, S10 RR12964, S10 RR 026825]. Funding for open access charge: NIH [GM061115 to P.A.B.].

Conflict of interest statement. None declared.

REFERENCES

1. Carlile,T.M., Rojas-Duran,M.F., Zinshteyn,B., Shin,H., Bartoli,K.M. and Gilbert,W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
2. Li,X., Xiong,X., Wang,K., Wang,L., Shu,X., Ma,S. and Yi,C. (2016) Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nat. Chem. Biol.*, **12**, 311–316.
3. Gilbert,W.V., Bell,T.A. and Schaening,C. (2016) Messenger RNA modifications: form, distribution, and function. *Science*, **352**, 1048–1412.
4. Chen,K., Zhao,B.S. and He,C. (2016) Nucleic acid modifications in regulation of gene expression. *Cell Chem. Biol.*, **23**, 74–85.
5. O'Connell,M.A., Mannion,N.M. and Keegan,L.P. (2015) The epitranscriptome and innate immunity. *PLoS Genet.*, **11**, e1005687.
6. Keegan,L.P., Gallo,A. and O'Connell,M.A. (2001) The many roles of an RNA editor. *Genetics* **2**, 869–878.
7. Bass,B.L., Nishikura,K., Keller,W., Seeburg,P.H., Emeson,R.B., O'Connell,M.A., Samuel,C.E. and Herbert,A. (1997) A standardized nomenclature for adenosine deaminases that act on RNA. *RNA*, **3**, 947–949.
8. Bass,B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.*, **71**, 817–846.
9. Melcher,T., Maas,S., Herb,A., Sprengel,R., Higuchi,M. and Seeburg,P.H. (1996) Communication—RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J. Biol. Chem.*, **271**, 31795–31798.
10. Chen,C., Cho,D., Wang,Q., Lai,F., Carter,K. and Nishikura,K. (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA*, **6**, 755–767.
11. Maas,S., Kawahara,Y., Tamburro,K.M. and Nishikura,K. (2006) A-to-I RNA editing and human disease. *RNA Biol.*, **3**, 1–9.
12. Li,M., Yang,L., Li,C., Jin,C., Lai,M., Zhang,G., Hu,Y., Ji,J. and Yao,Z. (2010) Mutational spectrum of the ADAR1 gene in dyschromatosis symmetrica hereditaria. *Arch. Dermatol. Res.*, **302**, 469–476.
13. Miyamura,Y., Suzuki,T., Kono,M., Inagaki,K., Ito,S., Suzuki,N. and Tomita,Y. (2003) Mutations of the RNA-specific adenosine deaminase gene (*DSRAD*) are involved in dyschromatosis symmetrica hereditaria. *Am. J. Hum. Genet.*, **73**, 693–699.
14. Rice,G.I., Kasher,P.R., Forte,G.M., Mannion,N.M., Greenwood,S.M., Szynkiewicz,M., Dickerson,J.E., Bhaskar,S.S., Zampini,M., Briggs,T.A. *et al.* (2012) Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nat. Genet.*, **44**, 1243–1248.
15. Zhang,X.J., He,P.P., Li,M., He,C.D., Yan,K.L., Cui,Y., Yang,S., Zhang,K.Y., Gao,M., Chen,J.J. *et al.* (2004) Seven novel mutations of the ADAR gene in Chinese families and sporadic patients with dyschromatosis symmetrica hereditaria (DSH). *Hum. Mutat.*, **23**, 629–630.
16. Eggington,J.M., Greene,T. and Bass,B.L. (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.*, **2**, 319–327.
17. Lehmann,K.A. and Bass,B.L. (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry*, **39**, 12875–12884.
18. Maas,S., Melcher,T., Herb,A., Seeburg,P.H., Keller,W., Krause,S., Higuchi,M. and O'Connell,M.A. (1996) Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.*, **271**, 12221–12226.
19. Melcher,T., Maas,S., Herb,A., Sprengel,R., Seeburg,P.H. and Higuchi,M. (1996) A mammalian RNA editing enzyme. *Nature*, **379**, 460–464.
20. Wong,S., Sato,S. and Lazinski,D.W. (2001) Substrate recognition by ADAR1 and ADAR2. *RNA*, **7**, 846–858.
21. Yeo,J., Goodman,R.A., Schirle,N.T., David,S.S. and Beal,P.A. (2010) RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20715–20719.

22. Goodman,R.A., Macbeth,M.R. and Beal,P.A. (2012) ADAR proteins: structure and catalytic mechanism. *Curr. Top. Microbiol. Immunol.*, **353**, 1–33.
23. Barraud,P. and Allain,F.H. (2012) ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.*, **353**, 35–60.
24. Bass,B.L., Hurst,S.R. and Singer,J.D. (1994) Binding properties of newly identified Xenopus proteins containing dsRNA binding motifs. *Curr. Biol.*, **4**, 301–314.
25. Doyle,M. and Jantsch,M.F. (2002) New and old roles of the double-stranded RNA-binding domain. *J. Struct. Biol.*, **140**, 147–153.
26. Ryter,J.M. and Schultz,S.C. (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA binding domain complexed with dsRNA. *EMBO J.*, **17**, 7505–7513.
27. Matthews,M.M., Thomas,J.M., Zheng,Y., Tran,K., Phelps,K.J., Scott,A.I., Havel,J., Fisher,A.J. and Beal,P.A. (2016) Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.*, **23**, 426–433.
28. Gietz,R.D. and Schiestl,R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.*, **2**, 31–34.
29. Macbeth,M.R. and Bass,B.L. (2007) Large-scale overexpression and purification of ADARs from *Saccharomyces cerevisiae* for biophysical and biochemical studies. *Methods Enzymol.*, **424**, 319–331.
30. Wang,Y., Havel,J. and Beal,P.A. (2015) A phenotypic screen for functional mutants of human adenosine deaminase acting on RNA 1. *ACS Chem. Biol.*, **10**, 2512–2519.
31. Pokharel,S. and Beal,P.A. (2006) High-throughput screening for functional adenosine to inosine RNA editing systems. *ACS Chem. Biol.*, **1**, 761–765.
32. Eifler,T., Pokharel,S. and Beal,P.A. (2013) RNA-Seq analysis identifies a novel set of editing substrates for human ADAR2 present in Saccharomyces cerevisiae. *Biochemistry*, **52**, 7857–7869.
33. Kuttan,A. and Bass,B.L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3295–E3304.
34. Pokharel,S., Jayalath,P., Maydanovych,O., Goodman,R.A., Wang,S.C., Tantillo,D.J. and Beal,P.A. (2009) Matching active site and substrate structures for an RNA editing reaction. *J. Am. Chem. Soc.*, **131**, 11882–11891.
35. Cormack,B. (1998) Green fluorescent protein as a reporter of transcription and protein localization in fungi. *Curr. Opin. Microbiol.*, **1**, 406–410.
36. Michener,J.K. and Smolke,C.D. (2012) High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng.*, **14**, 306–316.
37. Shiroishi,M. and Kobayashi,T. (2015) Screening of stable G-protein-coupled receptor variants in Saccharomyces cerevisiae. *Methods Mol. Biol.*, **1261**, 159–170.
38. Phelps,K.J., Tran,K., Eifler,T., Erickson,A.I., Fisher,A.J. and Beal,P.A. (2015) Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res.*, **43**, 1123–1132.
39. Gajula,K.S., Huwe,P.J., Mo,C.Y., Crawford,D.J., Stivers,J.T., Radhakrishnan,R. and Kohli,R.M. (2014) High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. *Nucleic Acids Res.*, **42**, 9964–9975.
40. Palladino,M.J., Keegan,L.P., O'Connell,M.A. and Reenan,R.A. (2000) dADAR,a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA*, **6**, 1004–1018.
41. Savva,Y.A., Jepson,J.E., Sahin,A., Sugden,A.U., Dorsky,J.S., Alpert,L., Lawrence,C. and Reenan,R.A. (2012) Auto-regulatory RNA editing fine-tunes mRNA re-coding and complex behaviour in *Drosophila*. *Nat. Commun.*, **3**, 790–799.
42. Deffit,S.N. and Hundley,H.A. (2016) To edit or not to edit: regulation of ADAR editing specificity and efficiency. *Wiley Interdiscip. Rev. RNA*, **7**, 113–127.
43. Garncarz,W., Tariq,A., Handl,C., Pusch,O. and Jantsch,M.F. (2013) A high throughput screen to identify enhancers of ADAR-mediated RNA-editing. *RNA Biol.*, **10**, 192–204.
44. Marcucci,R., Brindle,J., Paro,S., Casadio,A., Hempel,S., Morrice,N., Bisso,A., Keegan,L.P., Del Sal,G. and O'Connell,M.A. (2011) Pin1 and WWP2 regulate GluR2 Q/R site RNA editing by ADAR2 with opposing effects. *EMBO J.*, **30**, 4211–4222.
45. Tariq,A., Garncarz,W., Handl,C., Balik,A., Pusch,O. and Jantsch,M.F. (2013) RNA-interacting proteins act as site-specific repressors of ADAR2-mediated RNA editing and fluctuate upon neuronal stimulation. *Nucleic Acids Res.*, **41**, 2581–2593.