Published in final edited form as:

J Am Chem Soc. 2020 February 05; 142(5): 2110–2114. doi:10.1021/jacs.9b09808.

Nanopore sequencing of an expanded genetic alphabet reveals high-fidelity replication of a predominantly hydrophobic unnatural base pair

Michael P. Ledbetter[†], Jonathan M. Craig[‡], Rebekah J. Karadeema[†], Matthew T. Noakes[‡], Hwanhee C. Kim[‡], Sarah J. Abell[‡], Jesse R. Huang[‡], Brooke A. Anderson[†], Ramanarayanan Krishnamurthy[†], Jens H. Gundlach[‡], Floyd E. Romesberg^{†,*}

[†]Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037 United States

[‡]Department of Physics, University of Washington, Seattle, Washington 98195 United States

Abstract

Unnatural base pairs (UBPs) have been developed and used for a variety of *in vitro* applications, as well as for the engineering of semi-synthetic organisms (SSOs) that store and retrieve increased information. However, these applications are limited by the availability of methods to rapidly and accurately determine the sequence of unnatural DNA. Here, we report the development and application of the MspA nanopore to sequence DNA containing the dTPT3-dNaM UBP. Analysis of two sequence contexts reveals that DNA containing the UBP is replicated with an efficiency and fidelity similar to that of natural DNA and sufficient for use as the basis of an SSO that produces proteins with non-canonical amino acids.

Graphical Abstract

*Corresponding Author: fromesberg@synthorx.com.

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website.

Methods, Tables S1 and S2, Figures S1–S9, plasmid map of pSYN-sfGFP-T2F(gg), Supporting References (PDF)

Table S3 – Summary of current pattern events from nanopore analysis of PCR and *in vivo* replicated UBP-containing DNA (XLSX)

The authors declare the following competing financial interests: a patent application has been filed based on the use of UBPs in SSOs and F.E.R. has a financial interest (shares) in is employed by Synthorx, Inc., a company that has commercial interests in the UBP.

We have created semi-synthetic organisms (SSOs) that store and retrieve the information made available by a family of unnatural base pairs (UBPs), typified by **dTPT3-dNaM** (Figure 1A). The SSOs replicate the UBPs contained within plasmid DNA¹⁻³ or chromosome DNA⁵ and then retrieve that genetic information via transcription and translation to produce proteins with site-specifically incorporated non-canonical amino acids (ncAAs).^{3,6} UBPs are also being developed by the Hirao and Benner groups to expand the chemical space available to functional nucleic acids.⁷⁻⁸ In addition, there are many other potential applications, ranging from information storage⁹⁻¹⁰ to the development of novel materials.¹¹⁻¹² However, these applications would all be facilitated by improved sequencing methods.

Methods to detect UBPs within DNA include protection from nuclease cleavage,¹³ induced strand cleavage,¹⁴ or labeling one of the unnatural nucleotides with a fluorophore or an affinity tag.¹⁵⁻¹⁶ Signal termination during Sanger sequencing has been employed to sequence DNA, either when the cognate unnatural triphosphate is not present¹⁷ or when it is replaced with its dideoxynucleoside analog.¹⁸ Condition-dependent forced loss of a UBP has been used to indirectly sequence DNA containing UBPs,⁷⁻⁸ but this method is labor intensive and possibly dependent on sequence context. Moreover, with all of the developed methods, determination of which unnatural nucleotide is present in a given strand is more challenging. This may be a particular concern with predominantly hydrophobic UBPs, such as those developed in the Romesberg and the Hirao groups, due to the possibility of unnatural nucleotide self-pairing (e.g. **dTPT3-dTPT3** or **dNaM-dNaM**) and subsequent inversion of UBP strand context (*i.e.* 5' **AXC** to **AYC**, **X=dNaM** and **Y=dTPT3**). In fact, it has been suggested that rampant inversion of the **dTPT3-dNaM** UBP makes it unsuitable for expansion of the genetic alphabet.¹⁹

Next generation sequencing (NGS) has revolutionized the biological sciences. The most common NGS methods rely on sequencing-by-synthesis, but their application for the sequencing of DNA containing UBPs would require both modification of the unnatural nucleotides and significant modification of existing instrumentation. In contrast, nanopore sequencing is a potentially transformative, emerging NGS technique that obviates the need for nucleotide or instrumentation modification. The nanopore sequencing method employed in this work functions by recording an ionic current through the protein nanopore MspA as the helicase Hel308 controls the transit of DNA through the pore (Figures 1B and S1).²⁰⁻²² Hel308 moves DNA through MspA with two steps per nucleotide, such that the translocation of a single nucleotide through the pore is seen as two distinct ion-current segments²³ (Figure 1C). Because the ion-current signal is generated by the physical motion of ions past the nucleobases, nanopore sequencing is sensitive to not only the structure of the natural nucleotides but also to nucleotide modifications such as methylation²⁴⁻²⁶ and to the unique structures of unnatural nucleotides. This has been demonstrated through direct detection of the **dNaM:d5SICS** pair with the MspA nanopore,²⁷ and marker-based detection of **dNaM** with the α -Hemolysin nanopore.²⁸ Here we use a recently developed MspA/Hel308 nanopore sequencing system that features a time-varying voltage⁴ to assess the replication fidelity of the 6-letter genetic alphabet consisting of the four natural letters, **dTPT3**, and **dNaM**.

To begin the development of a nanopore sequencing method for our expanded genetic alphabet, a reference library of consensus current patterns was generated empirically using chemically synthesized oligonucleotide standards (see Supporting Information and Table S1). 65-mer standards were synthesized and measured with the MspA/He1308 nanopore that possessed a centrally located **AXC** or **GXC** sequence or a mutational derivative: **dNaM** deleted or replaced with **dTPT3**, a natural nucleotide, or a tetrahydrofuran abasic site analog. Each sequence's consensus current pattern was generated by averaging the patterns of 40 unique measurements and was found to be unique from the mutationally related sequences (Figures 1C, S3, and S4). A single-nucleotide polymorphism (SNP) calling algorithm was developed from this reference library to determine the identity of the central nucleotide in the UBP site. The algorithm correctly matches single-molecule current patterns with SNP identity with >90% accuracy for most sequences (see Figure S5 and Supporting Information).

Given the importance of PCR as a tool for both constructing and analyzing DNA containing UBPs, we first sequenced DNAs produced from an unnatural PCR (uPCR, i.e. PCR supplemented with **dTPT3TP** and **dNaMTP** and with templates amplified by One *Taq*, a mixture of *Taq* and DeepVent polymerases; see Supporting Information) templated by the same **AAC**, **AXC**, or **GXC** 65-mer oligonucleotides used to develop the sequence calling algorithm. Templates were amplified via uPCR. Amplification levels were estimated by PAGE analysis of control uPCRs (Supporting Information and Figure S8), and the amplicons were analyzed by streptavidin-shift PCR and nanopore sequencing after 13 doublings (Figure 2, Table S3, and Supporting Information). The mis-calling rate of the SNP calling algorithm limits our ability to detect rare mutations, generally SNPs that comprise less than 2% of a given sample. Thus, in order to improve our ability to detect rare mutation events during uPCR, uPCR products were diluted and reamplified sequentially four times for a total of 65 DNA doublings and analyzed as above.

Fidelity was determined as the percent retention of the central A or **X** within the **AAC**, **AXC**, or **GXC** sequence, normalized by the number of DNA doublings (Supporting Information). The fidelity of the **AAC** template amplified via standard PCR was $99.62 \pm 0.05\%$ per doubling (average $\pm$ s.d) after 13 doublings and $99.84 \pm 0.04\%$ after 65 doublings. When amplified by uPCR, the fidelity was slightly lower ($99.20 \pm 0.13\%$ after 13 doublings and $99.65 \pm 0.07\%$ after 65 doublings), due to low level misinsertion of **dTPT3TP**. For the **AXC** template amplified by uPCR, the fidelities were $99.45 \pm 0.09\%$ after 13 doublings and $99.39 \pm 0.16\%$ after 65 doublings. Finally, for the **GXC** template amplified by uPCR, the fidelities were $99.30 \pm 0.53\%$ after 13 doublings and $99.54 \pm 0.17\%$ after 65 doublings. While these results reveal that uPCR produces DNA with a fidelity that is slightly reduced relative to natural PCR, the fidelity with which the UBP is replicated via uPCR is comparable to that of a natural base pair. Additionally, the observation that the fidelity with which the **AXC** and **GXC** templates were replicated remained constant with increased amplification demonstrates that natural DNA is not significantly more efficiently replicated than DNA containing the UBP.

For the **AAC** template amplified by uPCR, deletion of dA and mutation to **dTPT3** (dA $\rightarrow$ **dTPT3**) were the most frequent mutations at 13 doublings (constituting $1.9 \pm 0.2\%$ and

1.9±1.7% of the amplicons, respectively) and dA → dC (4.4±2.6%) and dA → dTPT3 (8.9±8.1%) at 65 doublings. For the AXC template, dNaM → dTPT3 (2.8±1.2%) and dNaM → dA (2.2±1.1%) were the most frequent mutations at 13 doublings, and dNaM → dTPT3 (17.2±3.4%) and dNaM → dC (3.9±2.4%) at 65 doublings. For the GXC template, dNaM → dTPT3 (2.7±3.2%) and dNaM → dA (3.1±0.7%) were the most frequent mutations at 13 doublings, and dNaM → dTPT3 (12.1±4.0%) and dNaM → dA (8.1±2.0%) at 65 doublings. While this data suggests that mutational introduction of dTPT3 occurs opposite both natural and unnatural nucleotides, the frequency of these events is not higher than mutational introduction of a natural nucleotide at 13 doublings. After 65 doublings, mutational introduction of a natural nucleotide remained minor, however, mutational introduction of dTPT3 increased significantly. We attribute this increase in mutational introduction of dTPT3 during later cycles of the PCR results to misinsertion and termination of synthesis during early cycles, followed by eventual reinitiation and completion of strand extension during a later cycle. However, this increase in mutational introduction of dTPT3 is only significant after more PCR amplification than is required for most practical applications.

To examine replication in the SSO, the AAC, AXC, and GXC oligonucleotides were cloned into a derivative of the pINF plasmid that we have used previously to encode and express proteins with ncAAs (see Supporting Information).³ SSOs (strain ML2⁵) were transformed with one of the three different pINF plasmids and six individual colonies for each sequence context were screened by streptavidin shift PCR to confirm that they had successfully received and initiated replication of an intact pINF plasmid. Three high retention SSO clones for each sequence context were passaged in liquid media for a total of ~75 cell doublings. Retention during the entire experiment was monitored by streptavidin shift PCR. At around 50 doublings, one of the three AXC replicates suffered pronounced loss of the UBP (see Supporting Information); contamination was surmised to be a likely cause, and this replicate is not included in the average metric reported below.

Given the difficulty of accurately measuring plasmid copy number, we estimated the *in vivo* amplification from the number of cell doublings that occurred in liquid media (see Supporting Information for details). This is expected to be accurate if the total pINF copy number remains constant, which is likely based on evidence that the presence of the UBP does not reduce the rate of DNA replication (as demonstrated above with uPCR, and also confirmed with replication in the SSO (see below)). In addition, as obtaining sufficient quantities of plasmid DNA directly from the SSO would have necessitated growth in prohibitively large cultures, plasmids replicated during defined amounts of growth of the SSO were recovered and subjected to 20 cycles of uPCR amplification prior to nanopore sequencing (Figure 3 and Table S3). Correspondingly, fidelity was calculated based on the total DNA amplification level, which includes the plasmid amplification during cell growth and amplicon doublings during PCR (Supporting Information).

Plasmids were isolated and analyzed after 21 SSO doublings, which corresponds to our workflow for expressing ncAA-containing proteins, and after the full 75 doublings to gauge fidelity over longer periods of growth. The fidelity of the *in vivo* replicated AAC sequence with uPCR analysis was 99.69±0.09% per doubling after 21 doublings and 99.88±0.03%

after 75 doublings. For the unnatural sequences, **AXC** fidelity was $99.66 \pm 0.18\%$ and $99.71 \pm 0.15\%$ and **GXC** fidelity was $99.66 \pm 0.13\%$ and $99.86 \pm 0.04\%$ after 21 and 75 doublings, respectively. The observation that these fidelities are similar, or actually slightly higher than with uPCR alone indicates that the fidelity of replication in the *in vivo* environment of the SSO is higher than during uPCR. Moreover, because plasmid replication during colony formation was not included in the estimated amplification and because the observed mutations accumulated during both the *in vivo* and *in vitro* replication, these fidelities represent lower estimates of the true *in vivo* fidelity. Additionally, the fact that the observed fidelities appeared to increase as a function of growth reveals that plasmids that lost the UBP, and thus were fully natural, were not replicated more efficiently than plasmids that had retained the UBP.

Mutations detected after *in vivo* propagation and *in vitro* analysis were more variable than those observed from uPCR alone. For the AAC sequence, the most frequent mutations were dA→dT ($2.6 \pm 1.0\%$ of products) and dA→dTPT3 ($2.8 \pm 1.1\%$) at 21 cell doublings, and dA→dC ($2.0 \pm 0.8\%$) and dA→dTPT3 ($3.7 \pm 1.4\%$) at 75 cell doublings. For the **AXC** sequence, the most frequent mutations were dNaM→dT ($5.8 \pm 4.6\%$) and dNaM→dTPT3 ($3.9 \pm 0.9\%$) at 21 cell doublings, and dNaM→dT ($12.2 \pm 8.9\%$) and dNaM→dTPT3 ($5.5 \pm 1.0\%$) at 75 cell doublings. For the **GXC** sequence, the most frequent mutations were dNaM→dA ($2.5 \pm 2.6\%$) and dNaM→dTPT3 ($3.5 \pm 2.4\%$) at 21 cell doublings; and dNaM→deletion ($2.3 \pm 0.9\%$), dNaM→dT ($4.7 \pm 4.6\%$), and dNaM→dTPT3 ($2.3 \pm 0.8\%$) at 75 cell doublings. Importantly, the data demonstrate that UBP strand inversion (dNaM→dTPT3) occurs infrequently.

The developed nanopore sequencing method has enabled the first thorough analysis of the fidelity with which an expanded genetic alphabet is replicated. Empirical expansion of this method's current pattern reference library with additional UBPs and sequence contexts would enable unrestricted sequencing of any UBP in any sequence context of interest, including both those developed by other groups as well as those formed as part of our ongoing effort to optimize dNaM-dTPT3 UBP. This method showed that, while some of this pairs high relative *in vitro* replication fidelity is due to a reduced fidelity of natural replication due to dTPT3 misinsertion, this only makes a significant contribution during extensive PCR amplification. Moreover, dTPT3 misinsertion does not appear to occur as frequently in the SSO, where the six-letter alphabet appears to be replicated with higher fidelity than *in vitro*. The increased fidelity likely results from the action of accessory proteins that are only available *in vivo* and highlights the importance of evaluating the UBPs in the environment where they are intended to function. Importantly, neither mutation to natural nucleotides nor strand inversion via self-pairing occur at problematic rates during *in vitro* or *in vivo* replication. Since the fidelity of unnatural protein synthesis appears to be determined by the fidelity of unnatural DNA replication,^{3,6} this method will immediately facilitate improved use and further development of the SSO.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health (GM118178 to F.E.R., GM128376 to R.J.K. and HG005115 to J.H.G.). M.P.L. was supported by a National Science Foundation Graduate Research Fellowship (NSF/DGE-1346837). B.A.A. and R.K. were supported by the NSF and the NASA Astrobiology Program under the NSF Center for Chemical Evolution (CHE-1504217).

REFERENCES

- (1). Feldman AW; Romesberg FE, In Vivo Structure-Activity Relationships and Optimization of an Unnatural Base Pair for Replication in a Semi-Synthetic Organism. *J. Am. Chem. Soc* 2017, 139, 11427–11433. [PubMed: 28796508]
- (2). Malyshev DA; Dhami K; Lavergne T; Chen T; Dai N; Foster JM; Correa IR Jr.; Romesberg FE, A Semi-synthetic Organism with an Expanded Genetic Alphabet. *Nature* 2014, 509, 385–388. [PubMed: 24805238]
- (3). Zhang Y; Ptacin JL; Fischer EC; Aerni HR; Caffaro CE; San Jose K; Feldman AW; Turner CR; Romesberg FE, A Semi-Synthetic Organism that Stores and Retrieves Increased Genetic Information. *Nature* 2017, 551, 644–647. [PubMed: 29189780]
- (4). Noakes MT; Brinkerhoff H; Laszlo AH; Derrington IM; Langford KW; Mount JW; Bowman JL; Baker KS; Doering KM; Tickman BI; Gundlach JH, Increasing the Accuracy of Nanopore DNA Sequencing Using a Time-Varying Cross Membrane Voltage. *Nat. Biotechnol* 2019, 37, 651–656. [PubMed: 31011178]
- (5). Ledbetter MP; Karadeema RJ; Romesberg FE, Reprogramming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc* 2018, 140, 758–765. [PubMed: 29309130]
- (6). Dien VT; Holcomb M; Feldman AW; Fischer EC; Dwyer TJ; Romesberg FE, Progress Toward a Semi-Synthetic Organism with an Unrestricted Expanded Genetic Alphabet. *J Am Chem Soc* 2018, 140, 16115–16123. [PubMed: 30418780]
- (7). Kimoto M; Yamashige R; Matsunaga K; Yokoyama S; Hirao I, Generation of High-Affinity DNA Aptamers Using an Expanded Genetic Alphabet. *Nat. Biotechnol* 2013, 31, 453–457. [PubMed: 23563318]
- (8). Sefah K; Yang Z; Bradley KM; Hoshika S; Jimenez E; Zhang L; Zhu G; Shanker S; Yu F; Turek D; Tan W; Benner SA, In Vitro Selection with Artificial Expanded Genetic Information Systems. *Proc. Natl. Acad. Sci. USA* 2014, 111, 1449–1454. [PubMed: 24379378]
- (9). Shipman SL; Nivala J; Macklis JD; Church GM, CRISPR-Cas Encoding of a Digital Movie into the Genomes of a Population of Living Bacteria. *Nature* 2017, 547, 345–349. [PubMed: 28700573]
- (10). Bancroft C; Bowler T; Bloom B; Clelland CT, Long-Term Storage of Information in DNA. *Science* 2001, 293, 1763–1765.
- (11). Roh YH; Ruiz RC; Peng S; Lee JB; Luo D, Engineering DNA-based Functional Materials. *Chem. Soc. Rev* 2011, 40, 5730–5744. [PubMed: 21858293]
- (12). Sacca B; Niemeyer CM, Functionalization of DNA Nanostructures with Proteins. *Chem. Soc. Rev* 2011, 40, 5910–5921. [PubMed: 21975573]
- (13). Chen F; Yang Z; Yan M; Alvarado JB; Wang G; Benner SA, Recognition of an Expanded Genetic Alphabet by Type-II Restriction Endonucleases and their Application to Analyze Polymerase Fidelity. *Nucleic Acids Res.* 2011, 39, 3949–3961. [PubMed: 21245035]
- (14). Karalkar NB; Leal NA; Kim MS; Bradley KM; Benner SA, Synthesis and Enzymology of 2'-Deoxy-7-deazaisoguanosine Triphosphate and Its Complement: A Second Generation Pair in an Artificially Expanded Genetic Information System. *ACS Synth. Biol* 2016, 5, 672–678. [PubMed: 26914388]
- (15). Okamoto I; Miyatake Y; Kimoto M; Hirao I, High Fidelity, Efficiency and Functionalization of Ds-Px Unnatural Base Pairs in PCR Amplification for a Genetic Alphabet Expansion System. *ACS Synth. Biol* 2016, 5, 1220–1230. [PubMed: 26814421]

- (16). Seo YJ; Malyshev DA; Laverne T; Ordoukhanian P; Romesberg FE, Site-specific Labeling of DNA and RNA using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc* 2011, 133, 19878–19888. [PubMed: 21981600]
- (17). Malyshev DA; Seo YJ; Ordoukhanian P; Romesberg FE, PCR with an Expanded Genetic Alphabet. *J. Am. Chem. Soc* 2009, 131, 14620–14621. [PubMed: 19788296]
- (18). Hirao I; Kimoto M; Mitsui T; Fujiwara T; Kawai R; Sato A; Harada Y; Yokoyama S, An Unnatural Hydrophobic Base Pair System: Site-Specific Incorporation of Nucleotide Analogs into DNA and RNA. *Nat. Methods* 2006, 3, 729–735. [PubMed: 16929319]
- (19). Hettinger TP, Helix Instability and Self-Pairing Prevent Unnatural Base Pairs from Expanding the Genetic Alphabet. *Proc. Natl. Acad. Sci. USA* 2017, 114, E6476–E6477. [PubMed: 28768813]
- (20). Manrao EA; Derrington IM; Laszlo AH; Langford KW; Hopper MK; Gillgren N; Pavlenok M; Niederweis M; Gundlach JH, Reading DNA at Single-Nucleotide Resolution with a Mutant MspA Nanopore and Phi29 DNA polymerase. *Nat. Biotechnol* 2012, 30, 349–353. [PubMed: 22446694]
- (21). Kasianowicz JJ; Brandin E; Branton D; Deamer DW, Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Proc. Natl. Acad. Sci. USA* 1996, 93, 13770–13773. [PubMed: 8943010]
- (22). Laszlo AH; Derrington IM; Ross BC; Brinkerhoff H; Adey A; Nova IC; Craig JM; Langford KW; Samson JM; Daza R; Doering K; Shendure J; Gundlach JH, Decoding Long Nanopore Sequencing Reads of Natural DNA. *Nat. Biotechnol* 2014, 32, 829–833. [PubMed: 24964173]
- (23). Derrington IM; Craig JM; Stava E; Laszlo AH; Ross BC; Brinkerhoff H; Nova IC; Doering K; Tickman BI; Ronaghi M; Mandell JG; Gunderson KL; Gundlach JH, Subangstrom Single-Molecule Measurements of Motor Proteins using a Nanopore. *Nat Biotechnol* 2015, 33, 1073–1075. [PubMed: 26414351]
- (24). Laszlo AH; Derrington IM; Brinkerhoff H; Langford KW; Nova IC; Samson JM; Bartlett JJ; Pavlenok M; Gundlach JH, Detection and Mapping of 5-Methylcytosine and 5-Hydroxymethylcytosine with Nanopore MspA. *Proc. Natl. Acad. Sci. USA* 2013, 110, 18904–18909. [PubMed: 24167255]
- (25). Schreiber J; Wescoe ZL; Abu-Shumays R; Vivian JT; Baatar B; Karplus K; Akeson M, Error Rates for Nanopore Discrimination among Cytosine, Methylcytosine, and Hydroxymethylcytosine along Individual DNA Strands. *Proc. Natl. Acad. Sci. USA* 2013, 110, 18910–18915. [PubMed: 24167260]
- (26). Wescoe ZL; Schreiber J; Akeson M, Nanopores Discriminate among Five C5-Cytosine Variants in DNA. *J. Am. Chem. Soc* 2014, 136, 16582–16587. [PubMed: 25347819]
- (27). Craig JM; Laszlo AH; Derrington IM; Ross BC; Brinkerhoff H; Nova IC; Doering K; Tickman BI; Svet MT; Gundlach JH, Direct Detection of Unnatural DNA Nucleotides dNaM and d5SICS using the MspA Nanopore. *PLoS One* 2015, 10, e0143253. [PubMed: 26588074]
- (28). Riedl J; Ding Y; Fleming AM; Burrows CJ, Identification of DNA Lesions using a Third Base Pair for Amplification and nanopore Sequencing. *Nat. Commun* 2015, 6, 8807. [PubMed: 26542210]

Figure 1. dNaM and dTPT3 produce unique current patterns. (A) The dNaM-dTPT3 UBP. (B) The MspA/He308 nanopore system. The MspA porin (gold) is imbedded in a lipid bilayer (purple). The He308 helicase (green) draws ssDNA through the porin as a variable voltage⁴ is applied across the membrane and current is measured. (C) Consensus current patterns from 40 control measurements of the indicated sequences. Shaded areas represent ± 1 S.D. (Figure S2), demonstrating that single-molecule measurements of dNaM and dTPT3 can be identified with high confidence. Typical SNP calling accuracies using variable-voltage sequencing are >90% for the 8 hypotheses (the 6-letter alphabet plus resulting abasic site and single nucleotide deletion, Figure S5).

Figure 2. Mutation spectrum of AAC, AXC, and GXC. Left axis and dots show % Retention_A, as measured by streptavidin-shift PCR, with white dots representing measured retention for each experimental replicate. Black dots represent average retention with ± 2 S.D. error bars, $n=3$. Right axis and stacked bar plots show the average percent abundance of each possible single nucleotide polymorphism (% of SNPs (Table S3)), as measured by nanopore sequencing.

Figure 3:

Mutation spectrum of AAC, AXC, and GXC *in vivo* and in uPCR. Left axis and dots show % Retention_B, as measured by streptavidin-shift PCR, with white dots representing measured retention for each experimental replicate. Red dots highlight the AXC sample that experienced catastrophic UBP loss and was excluded from reported average retention and % SNP abundance. Black dots represent average retention with ± 2 S.D. error bars, $n=3$ (except AXC, $n=2$). Right axis and stacked bar plots show the average percent abundance of each possible single nucleotide polymorphism (% of SNPs (Table S3)), as measured by nanopore sequencing.