Engineering a Cytidine Aminotransferase for Biocatalytic Production of the Covid-19 Antiviral Molnupiravir

Ashleigh J. Burke[a], William R. Birmingham[a], Ying Zhuo[a], Bruna Zucoloto da Costa[a], Rebecca Crawshaw[a], Thomas W. Thorpe[a], Ian Rowles[a], James Finnigan[b], Carl Young[b], Simon J. Charnock[b], Sarah L. Lovelock[a]*, Nicholas J. Turner[a]* and Anthony P. Green[a]*

[a] Department of Chemistry, University of Manchester, Manchester Institute of Biotechnology, 131 Princess Street, Manchester, M1 7DN, UK

[b] Prozomix Ltd, Building 4, West End Ind. Estate, Haltwhistle, UK, NE49 9HA.

Abstract

The Covid-19 pandemic highlights the urgent need for cost-effective processes to rapidly manufacture antiviral drugs at scale. Here we report a concise biocatalytic process for Molnupiravir, a nucleoside analogue currently under review for emergency use authorization as an orally available treatment for SARS-CoV-2. Key to the success of this process was the development of a cytidine aminotransferase for the production of *N*-hydroxy-cytidine through evolutionary adaption of the hydrolytic enzyme cytidine deaminase. This engineered biocatalyst performs >100,000 turnovers in less than 3 hours, operates at 180 g/L substrate loading and benefits from *in situ* crystallization of the *N*-hydroxy-cytidine product (>90% yield), which can be converted to Molnupiravir by a selective 5'-acylation using Novozym® 435.

Introduction

Molnupiravir **1** is a nucleoside analogue currently in the advanced stages of clinical trials for the treatment of Covid-19[1]. Positive clinical data prompted early termination of recent phase 3 trials in order to seek emergency use authorization to treat adult patients with mild-to-moderate Covid-19 who are at risk of progressing to severity or hospitalization. The structural simplicity of Molnupiravir compared to alternative therapies (*e.g.* Remdesivir), combined with its oral availability and broad-spectrum antiviral activities, make it an attractive candidate for the global treatment of viral infections. In view of the urgent need to develop therapies for treating the Covid-19 pandemic there has been considerable interest in developing synthetic routes to this API. The original synthesis required 10 steps and proceeded with <10% overall yield[2,3]. More streamlined approaches have subsequently been reported[4,5,6,7,8], however these routes rely on chemical methods to install the *N*-hydroxy unit, which require elevated temperatures, prolonged reaction times, and/or additional steps for activation and protection/deprotection chemistry. These factors compromise reaction yields and process productivity, ultimately leading to increased costs.

We sought to develop a more efficient and sustainable route to **1** that could be implemented on scale, where modification of the base to install the *N*-hydroxy unit is achieved catalytically using an engineered enzyme. Since *N*-hydroxy-cytidine **4** has previously been shown to undergo lipase catalyzed acylation at C-5-OH[4,5], this approach would provide an integrated biocatalytic process to Molnupiravir **1** from readily available starting materials. Our desired route requires a cytidine aminotransferase capable of selectively installing the *N*-hydroxy unit from NH_2OH in a bulk water phase (Fig. 1). To meet the demands of a robust industrial biocatalyst, we set ourselves the target of developing an engineered enzyme capable of operating at [S] >100 g/L and achieving >90% conversion

within 12h using low catalyst loadings to drive down the enzyme cost contribution to the overall process. At the outset of the project there was no known enzyme for this transformation. However, we were intrigued by a report which demonstrated that *N*-hydroxy-cytidine **4** could be hydrolysed to uridine **3** by wild-type cytidine deaminase, albeit at a very low rate (*ca.* 2% of activity compared to cytidine **2**)[9]. Cytidine deaminase (CD; EC 3.5.4.5) is a zinc containing hydrolytic enzyme which catalyzes the conversion of cytidine **2** to uridine **3**[10,11,12,13,14]. The enzyme is widely distributed amongst organisms and is involved in the salvage of cytidine for uridine mono-phosphate synthesis. Interestingly wild-type cytidine deaminase from *E. coli* has been used on scale for hydrolysis of the nucleoside analogue 2'-deoxy-3'-thiacytidine and hence it appeared to be an attractive choice as a potential industrial biocatalyst[15]. On the basis of these observations we initiated experiments aimed at converting cytidine **2** to *N*-hydroxy-cytidine **4** catalyzed by CD.

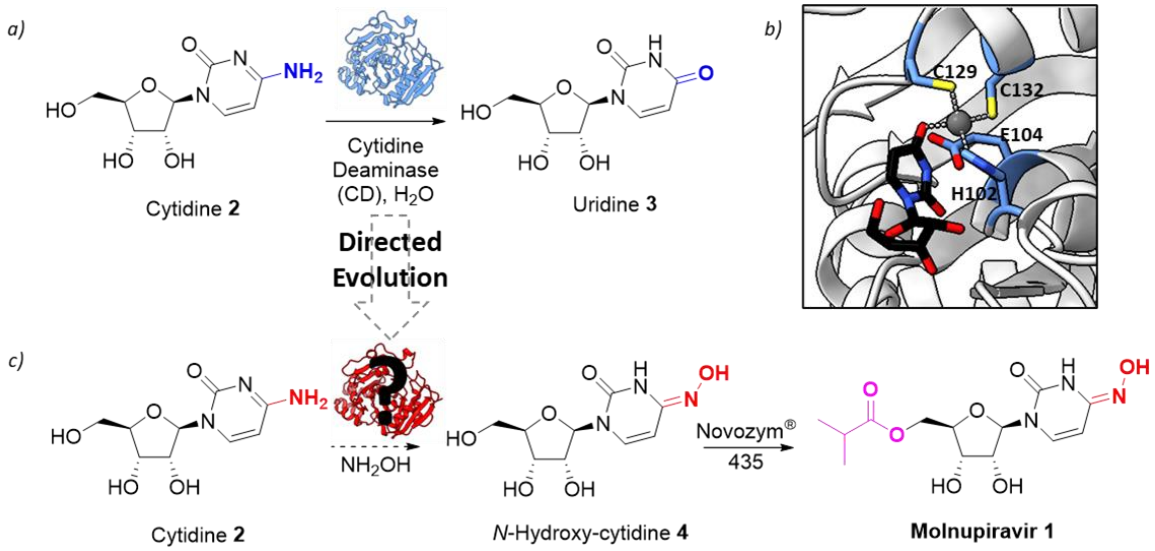


Fig. 1. Proposed biocatalytic route to Molnupiravir. a) Wild-type cytidine deaminase (CD) catalyzes the conversion of cytidine **2** to uridine **3** in the presence of water. b) The active site of CD with uridine bound (PDB code: 1AF2). The Zn^{2+} ion is shown in grey. The Zn^{2+} binding residues His102, Cys129 and Cys132, and catalytic Glu104 are shown as atom-coloured sticks with blue carbons. Uridine ligand is shown as atom-coloured sticks with black carbons. c) Proposed synthetic route to Molnupiravir **1**. Cytidine **2** is converted to *N*-hydroxy-cytidine **4** by an engineered cytidine aminotransferase followed by acylation using Novozym®-435 to give Molnupiravir **1**[4,5].

Results & Discussion

CD from *E. coli* was expressed in BL21(DE3) cells and purified to homogeneity via nickel affinity chromatography. To investigate whether this enzyme could serve as a biocatalyst for the preparation of *N*-hydroxy-cytidine **4**, we initially established direct spectrophotometric assays to monitor the interconversion of cytidine **2**, uridine **3** and *N*-hydroxy-cytidine **4**, based on diagnostic differences between the UV-Vis spectra of these species. Such assays are valuable to allow high-throughput and real-time analysis of biotransformations. In particular, the *N*-hydroxy-cytidine **4** spectrum shows a diagnostic feature at >310 nm that is not present in either cytidine **2** or uridine **3**, thus allowing *N*-hydroxy-cytidine **4** formation/decay to be easily monitored.

We initially employed this assay to investigate the conversion of *N*-hydroxy-cytidine **4** to uridine **3**. Consistent with previous reports, cytidine deaminase promotes *N*-hydroxy-cytidine **4** hydrolysis under ambient conditions, with complete conversion of 100 µM substrate achieved in 60 minutes (0.1% enzyme loading, Fig. S2). Encouraged by these results, we next turned our attention to the more challenging synthesis of *N*-hydroxy-cytidine **4** using either uridine **3** or cytidine **2** as a starting material. Pleasingly, biotransformations carried out in the presence of 1% NH_2OH in water led to the accumulation of a product with spectral features consistent with *N*-hydroxy-cytidine **4**. Similar final concentrations of **4** are formed using either **2** or **3** as a starting material, or in reactions starting from **4**, suggesting that the enzyme establishes an equilibrium distribution of products (Fig. 2b). As anticipated, the ratio of **4**:**3** formed is dependent on NH_2OH concentration and 1:6 and 3:1 mixtures were achieved using 10% and 50% NH_2OH solutions, respectively (Fig. 2c, S3). Significantly, the concentration of product **4** does not accumulate beyond the equilibrium distribution irrespective of the NH_2OH concentration used (Fig. 2a), suggesting that cytidine hydrolysis followed by uridine amination (pathway B) outcompetes direct amine transfer (pathway A) with the wild type enzyme.

These results highlight biotransformations with cytidine deaminase as a promising strategy for the production of *N*-hydroxy-cytidine **4**. Unfortunately, under thermodynamic control significant quantities of uridine by-product are formed, even at elevated NH_2OH concentrations. Furthermore, the use of such high NH_2OH concentrations is undesirable for large scale applications and leads to the formation of impurities over extended reaction times. To address these shortcomings, we elected to engineer CD *via* directed evolution to optimize pathway A whilst minimizing pathway B. Similar strategies have been employed to control the partitioning of transglycosylation/hydrolysis by glycoside hydrolases[16,17]. Accelerating the direct amination of cytidine in this way should allow accumulation of **4** under kinetic control at reduced NH_2OH concentrations.

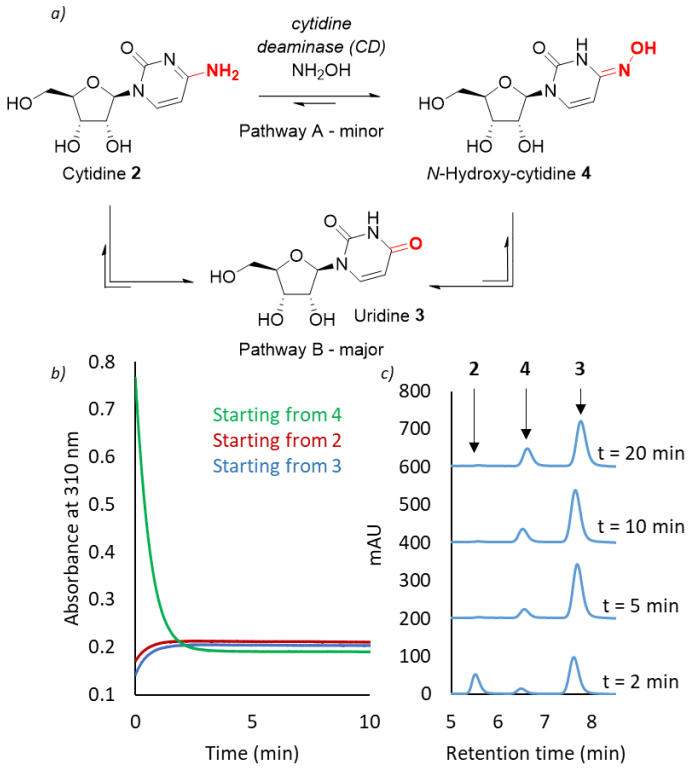


Fig. 2. Characterization of wild-type cytidine deaminase (CD). a) Pathways for conversion of cytidine **2** to *N*-hydroxy-cytidine **4** catalyzed by cytidine deaminase. Pathway A involves direct conversion of **2** to **4** using NH_2OH as a nucleophile. Pathway B involves initial hydrolysis of **2** to uridine **3** which is subsequently transformed to **4** through condensation with NH_2OH. Pathway B is the dominant pathway when using the wild-type enzyme, leading to an equilibrium distribution of **3** and **4** that is dependent on NH_2OH concentration. b) The conversion of cytidine **2** (1mM) and uridine **3** (1mM) to *N*-hydroxy-cytidine **4** by CD (5 µM) in the presence of 1% NH_2OH (pH 7) is monitored by increasing absorbance at 310 nm. Similar final concentrations of *N*-hydroxy-cytidine are formed using either **2** (red) or **3** (blue) as a starting material, or in reactions starting from *N*-hydroxy-cytidine **4** (green). c) The conversion of cytidine **2** (750 mM) by CD (25 µM) to uridine **3** to *N*-hydroxy-cytidine **4** is monitored by HPLC analysis. The time course of the reaction indicates CD operates *via* pathway B.

To this end, iterative rounds of site saturation mutagenesis were performed using NNK degenerate codons, targeting residues within close proximity to the Zn^{2+} cofactor and the substrate binding pocket. Beneficial diversity identified during rounds of evolution was combined by DNA shuffling. The aforementioned spectrophotometric assay was used to evaluate individual variants as crude cell lysates arrayed in 96-well plates, using 50 mM **2** as the substrate and 1% NH_2OH (pH 7) as the reaction medium. Throughout evolution, we identified improved variants with kinetic profiles consistent with rapid initial accumulation of **4** beyond the equilibrium position followed by slower redistribution of products to equilibrium (Fig. 3b).

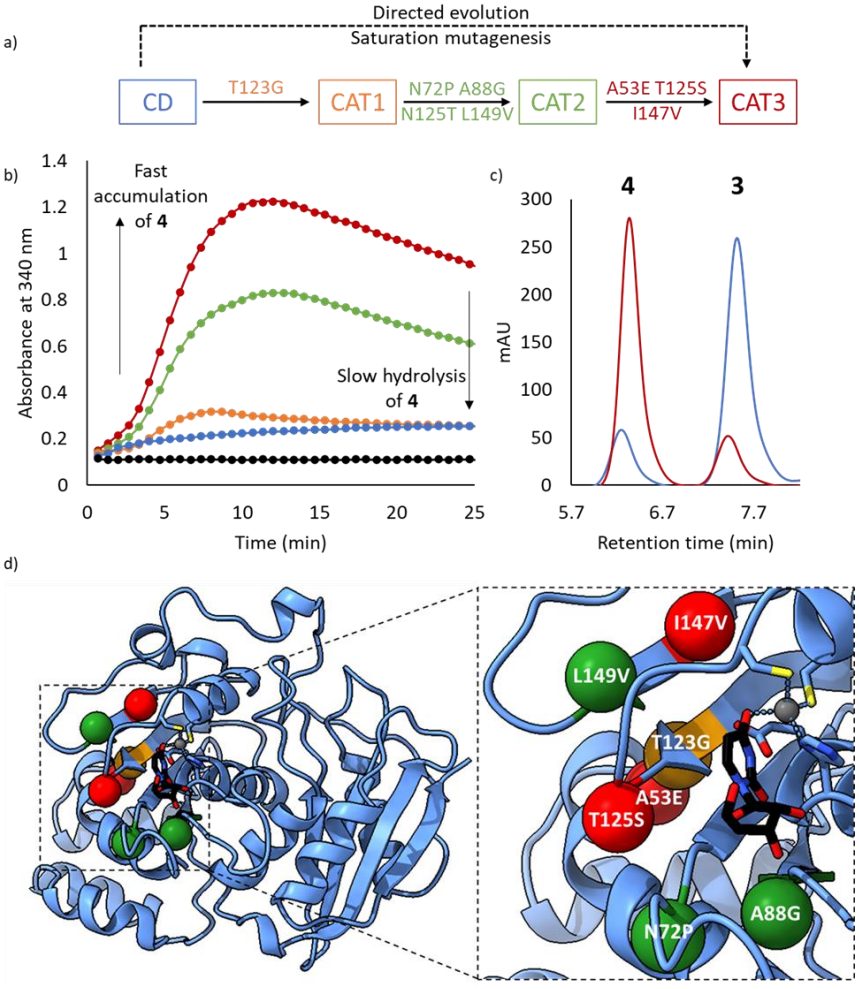


Fig. 3. Directed evolution of a cytidine aminotransferase (CAT). a) Directed evolution of CD to CAT3 showing mutations installed during each round. b) Time course for the formation of *N*-hydroxy-cytidine **4** from cytidine **2** (50 mM) catalyzed by CAT3 (2.5 µM, red), CAT2 (2.5 µM, green), CAT1 (2.5 µM, orange), wild-type CD (2.5 µM, blue) and no enzyme (black) in the presence of 1% NH_2OH, pH 7. The evolved CD variants show initial accumulation of *N*-hydroxy-cytidine **4** beyond the equilibrium position followed by slower redistribution of products to equilibrium. c) HPLC traces showing cytidine **2** (500 mM) conversion to *N*-hydroxy-cytidine **4** and uridine **3** catalyzed by CAT3 (25 µM, red) and wild-type CD (25 µM, blue) in the presence of 10% NH_2OH, pH 7. Retention times of **3** and **4** are identical to authentic standards. d) The active site of CD with uridine bound (PDB code: 1AF2). Mutations installed in round 1 (T123G) are shown as orange spheres, round 2 (N72P A88G L149V) as green spheres, and round 3 (A53E T125S I147V) as red spheres. N125T from round 2 is not included as this was subsequently mutated in round 3. The uridine ligand is shown as atom-coloured sticks with black carbons and the Zn^{2+} cofactor is shown in grey. Zn^{2+} binding residues His102, Cys 129 and Cys132, and catalytic Glu104 are shown as atom-coloured sticks with blue carbons.

The most promising variant to emerge following three rounds of evolution (CAT3) contains 7 mutations clustered around the active site. As intended CAT3 operates as an efficient aminotransferase, promoting the conversion of **2** to **4** even at low NH_2OH concentrations and high substrate loadings. For example, at 500 mM **2** and 10% NH_2OH (pH 7) CAT3 produces an 8:1 ratio of **4**:**3**. In contrast, the wild-type enzyme produces a 1:6 ratio of **4**:**3** under identical conditions at equilibrium.

With a promising biocatalyst in hand, our attention turned to reaction intensification and product isolation. We observed that *N*-hydroxy-cytidine is substantially less soluble than either cytidine or uridine in aqueous solutions, presenting an opportunity for dynamic *in situ* crystallization of **4** to allow facile product isolation and further favour the distribution of **4**:**3**. By increasing the substrate loading (750 mM) and reducing the reaction temperature (4 °C) *in situ* crystallization of **4** was observed. Following optimisation of reaction conditions, **4** was isolated in 91% yield and >98% purity in only 3 hours using 0.001 mol% CAT3.

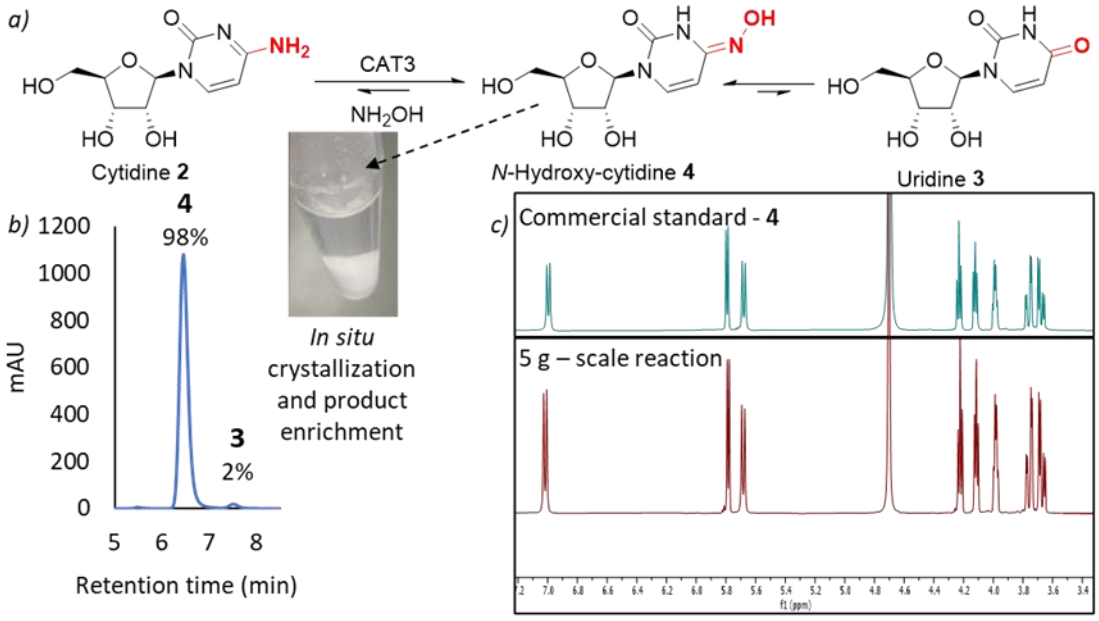


Fig. 4. A biocatalytic process for *N*-hydroxy-cytidine synthesis. a) *In situ* crystallization of *N*-hydroxy-cytidine **4** in reactions catalyzed by CAT3 leads to product enrichment. Reaction conditions: 750 mM **2**, 7.5 µM CAT3, 10% NH_2OH (pH 7), 4°C. b) HPLC trace of the product isolated from the biotransformation described in (a). *N*-hydroxy-cytidine **4** is isolated in 91% yield and >98% purity as confirmed by HPLC analysis and analysis of the ^{1}H NMR spectra. c) Stacked ^{1}H NMR traces of *N*-hydroxy-cytidine **4**, commercial standard (top) and the product isolated from the biotransformation described in (a) (bottom).

Conclusions

In summary, we have developed an efficient biocatalytic synthesis of *N*-hydroxy-cytidine **4**, a key intermediate for the production of Molnupiravir **1**. The process takes advantage of an engineered cytidine aminotransferase and benefits from dynamic product crystallization to provide a scalable and sustainable manufacturing route to an important molecule in the fight against Covid-19. Through rounds of directed evolution we were able to achieve the target metrics initially set, namely to operate at [S] > 100 g/L with 90% conversion in 24h. Remarkably this engineered biocatalyst is able to achieve TON >10^5 which places it within the sphere of well-established industrial biocatalysts of proven utility[18,19,20,21]. Given that selective 5'-acylation of *N*-hydroxy-cytidine can be achieved with lipases[4,5], this work establishes an integrated biocatalytic strategy for Molnupiravir synthesis using cytidine as an inexpensive and readily available starting material.

Acknowledgements

We would like to thank the Bill & Melinda Gates Foundation for funding and are particularly grateful to Dr. John Dillon, Dr. Trevor Laird and Dr. Silpa Sundaram for their helpful comments and advice during this project. We also acknowledge Dr. Aisling Ní Cheallaigh for helpful discussions on product isolation. We are grateful to Manchester SYNBIOCHEM Centre (BB/M017702/1) and the Henry Royce Institute for Advanced Materials (funded through EPSRC grants EP/R00661X/1, EP/S019367/1, EP/P025021/1 and EP/P025498/1) for access to their facilities.

Data Availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Source data are available from the corresponding author upon reasonable request.

References

1. Cox, R. M., Wolf, J. D. & Plemper, R. K. Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nat. Microbiol.* **6**, 11-18 (2021).
2. Painter, G. R., Bluemling, G. R., Natchus, M. G. & Guthrie, D. *N*4-hydroxycytidine and derivatives and anti-viral uses related thereto. WO2019113462 (2018).
3. Painter, G. R., Perryman, D. & Bluemling, G. R. 4'-Halogen containing nucleotide and nucleoside therapeutic compositions and uses related thereto. WO2019173602 (2019).
4. Vasudevan, N. *et al.* A concise route to MK-4482 (EIDD-2801) from cytidine. *Chem. Commun*. **56**, 13363-13364 (2020).
5. Ahlqvist, G. P. *et al.* Progress toward a large-scale synthesis of Molnupiravir (MK-4482, EIDD-2801) from cytidine. *ACS Omega* **6**, 10396-10402 (2021).
6. Steiner, A. *et al.* A high-yielding synthesis of EIDD-2801 from uridine. *Eur. J. Org. Chem.* 6736-6739 (2020).
7. Benkovics, T. *et al.*, https://doi.org/10.26434/chemrxiv.13472373.v1 (2020).

8. Paymode, D. J. *et al.*, Toward a practical, two-step process for Molnupiravir: direct hydroxamination of cytidine followed by selective esterification. *Org. Process Res. Dev.* **25**, 1822–1830 (2021).
9. Trimble, R. B. & Maley, F. Metabolism of 4-*N*-hydroxy-cytidine in *Escherichia coli*. *J. Bacteriol.* **108**, 145–153 (1971).
10. Frick, L., Mac Neela, J. P. & Wolfenden, R. Transition State Stabilization by Deaminases: Rates of Nonenzymatic Hydrolysis of Adenosine and Cytidine. *Bioorg. Chem.* **15**, 100-108 (1987).
11. Cohen, R. M. & Wolfenden, R. Cytidine Deaminase from *Escherichia coli*: purification, properties, and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *J. Biol. Chem.* **246**, 7561-7565 (1971).
12. Cohen, R. M. & Wolfenden, R. The equilibrium of hydrolytic deamination of cytidine and *N*4-methylcytidine. *J. Biol. Chem.* **246**, 7566-7568 (1971)
13. Xiang, S., Short, S. A., Wolfenden, R. & Carter Jr., C. W. Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry* **34**, 4516-4523 (1995).
14. Xiang, S., Short, S. A., Wolfenden, R. & Carter Jr., C. W. The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization. *Biochemistry* **36**, 4768-4774 (1997).
15. Ferrero, M. & Gotor, V. Biocatalytic selective modifications of conventional nucleosides, carbocyclic nucleosides, and C-nucleosides. *Chem. Rev.* **100**, 4319-4348 (2000).
16. Romero-Téllez, S., Lluch, J. M., González-Lafont, À. & Masgrau, L. Comparing hydrolysis and transglycosylation reactions catalyzed by *Thermus thermophilus* β-glycosidase. A combined MD and QM/MM study. *Front. Chem*. **7**, 200 (2019).
17. Moulis, C., Guieysse, D., Morel, S., Séverac, E. & Remaud-Siméon, M. Natural and engineered transglycosylases: Green tools for the enzyme-based synthesis of glycoproducts. *Curr. Opin. Chem. Biol.* **61**, 96-106 (2021).
18. Rogers, T. A. & Bommarius, A. S. Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes. *Chem. Eng. Sci.* **65**, 2118-2124 (2010).
19. Burton, S. G., Cowan, D. A. & Woodley, J. M. The search for the ideal biocatalyst. *Nat. Biotechnol.* **20**, 37-45 (2002).
20. Woodley, J. M. Accelerating the implementation of biocatalysis in industry. *Appl. Microbiol. Biotechnol.* **103**, 4733-4739 (2019).
21. Birmingham W. R. *et al.* Toward scalable biocatalytic conversion of 5-hydroxymethylfurfural by galactose oxidase using coordinated reaction and enzyme engineering. *Nat. Commun.* **12**, 4946 (2021).

Methods

Materials.

All chemicals and biological materials were obtained from commercial suppliers. *N*-Hydroxy-cytidine from Cambridge Bioscience; cytidine, uridine, Lysozyme, DNase I and kanamycin were purchased from Sigma Aldrich; polymyxin B sulfate from AlfaAesar; LB agar, LB media, 2×YT media and arabinose from Formedium; *Escherichia coli* BL21(DE3), Q5 DNA polymerase, T4 DNA ligase and restriction enzymes from New England BioLabs; and oligonucleotides were synthesized by Integrated DNA Technologies.

Construction of pET29b(+)_CD.

The *E. coli*-optimised gene (synthesised by Integrated DNA Technologies) for cytidine deaminase (CD) from *E. coli* was cloned into pET29b(+) using *NdeI* and *XhoI* restriction sites to yield pET29b(+)_CD.

Protein production and purification.

For expression of CD and variants, chemically competent *E. coli* BL21 (DE3) were transformed with the relevant pET29b(+)_CD constructs. Single colonies of freshly transformed cells were cultured for 18 h in 5 mL LB medium containing 50 μg mL^{-1} kanamycin. Starter cultures (500 μL) were used to inoculate 50 mL 2×YT medium supplemented with 50 μg mL^{-1} kanamycin. Cultures were grown at 37 °C, 200 r.p.m. to an optical density at 600 nm (OD600) of around 0.5. Protein expression was induced with the addition of IPTG to a final concentration of 0.1 mM. Induced cultures were incubated for 20 h at 30 °C and the cells were subsequently collected by centrifugation (3,220 x *g* for 10 min). Pelleted cells were resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, pH 7.5 containing 20 mM imidazole) and lysed by sonication. Cell lysates were cleared by centrifugation (27,216 x *g* for 30 min) and supernatants were subjected to affinity chromatography using Ni-NTA Agarose (Qiagen). Purified protein was eluted using 50 mM HEPES, 300 mM NaCl, pH 7.5 containing 250 mM imidazole. Proteins were desalted using 10DG desalting columns (Bio-Rad) with PBS pH 7.4 and analyzed by SDS PAGE. Proteins were aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C. Protein concentrations were determined by measuring the absorbance at 280 nm and assuming an extinction coefficient of 29910 M^{-1} cm^{-1}.

Mass spectrometry.

Purified protein samples were buffer-exchanged into 0.1% acetic acid using a 10k MWCO Vivaspin (Sartorius) and diluted to a final concentration of 0.5 mg ml^{-1}. Mass spectrometry was performed using a 1200 series Agilent LC, 5 μL injection into 5% acetonitrile (with 0.1% formic acid) and desalted inline for 1 min. Protein was eluted over 1 min using 95% acetonitrile with 5% water. The resulting multiply charged spectrum was analysed using an Agilent QTOF 6510 and deconvoluted using Agilent MassHunter Software.

Library construction.

Saturation mutagenesis. Positions were individually randomized using NNK codon degeneracy. DNA libraries were constructed by overlap extension PCR. The linear library fragments and the pET29b(+) vector were digested using *NdeI* and *XhoI* endonucleases, gel-purified and subsequently ligated using T4 DNA ligase.

Shuffling by overlap extension PCR.

After each round of evolution, beneficial diversity was combined by DNA shuffling of fragments generated by overlap extension PCR. Primers were designed that encoded either the parent amino acid or the identified mutation. These primers were used to generate short fragments which were gel-purified and mixed appropriately in overlap extension PCR to generate genes containing all possible combinations of mutations. Genes were cloned as described above.

Library screening.

For protein expression and screening, all transfer and aliquotting steps were performed using Hamilton liquid-handling robots. Chemically competent *E. coli* BL21(DE3) cells were transformed with the ligated libraries. Freshly transformed clones were used to inoculate 150 µL of LB medium supplemented with 50 µg mL^{-1} kanamycin in Corning® Costar® 96-well microtitre round bottom plates. For reference, each plate contained 6 freshly transformed clones of the parent template and 2 clones containing an empty pET29b(+) vector. Plates were incubated overnight at 30 °C, 80 % humidity in a shaking incubator at 850 r.p.m. 20 µL of overnight culture was used to inoculate 480 µL 2xYT medium supplemented with 50 µg mL^{-1} kanamycin. The cultures were incubated at 30 °C, 80 % humidity with shaking at 850 r.p.m. until an OD600 of about 0.5 was reached, and IPTG was added to a final concentration of 0.1 mM. Induced plates were incubated for 20 h at 30 °C, 80 % humidity with shaking at 850 r.p.m. Cells were harvested by centrifugation at 2,900 x *g* for 10 min. The supernatant was discarded, the pelleted cells were resuspended in 400 µL lysis buffer (PBS pH 7.4 buffer supplemented with 1.0 mg mL^{-1} lysozyme, 0.5 mg mL^{-1} polymixin B and 10 µg mL^{-1} DNase I) and incubated for 2 h at 30 °C, 80 % humidity with shaking at 850 r.p.m. Cell debris was removed by centrifugation at 2,900 x *g* for 10 min. 10 µL clarified lysate was transferred to 96-well microtitre plates containing 170 µL 1% NH_2OH. Reactions were initiated with the addition of 20 µL cytidine (50 mM final concentration) in PBS pH 7.4. The reaction was monitored spectrophotometrically at 340 nm, over 30 minutes using a CLARIOstar plate reader (BMG Labtech). Reaction rates of individual variants were normalized to the average of the 6 parent clones. Following each round, the most active variants were rescreened as purified proteins using the HPLC assay. Proteins were produced and purified as described above, however starter cultures were inoculated from glycerol stocks prepared from the original overnight cultures.

General procedure for analytical scale biotransformation.

To compare the activity of CD and its variants, analytical scale biotransformations were performed using **2** (750 mM) and the relevant biocatalyst (25 µM) in hydroxylamine (10 % in water, pH 7). For HPLC analysis, reactions were quenched with the addition of 2 volumes of acidified acetonitrile. Samples were vortexed and precipitated proteins were removed by centrifugation (14,000 x *g* for 5 minutes).

Preparative scale biotransformation.

Cytidine (5 g, 750 mM, 180 g L^{-1}) was dissolved in hydroxylamine (26.6 mL, 10 % in water, pH 7) in a 50 mL falcon tube and cooled to 4 °C. CAT3 purified enzyme (0.2 mL, 7.5 µM, 0.25 g L^{-1} final concentration) was added to the reaction mixture. The reaction was left on a tabletop roller for 3 hours at 4 °C during which time the *N*-hydroxy-cytidine product crystallized *in situ*. The reaction mixture was centrifuged at 2900 x *g* for 5 minutes and the supernatant removed. The remaining solid was washed with acidified water (10% AcOH) and then freeze dried overnight to give a free flowing white powder (4.85g, 91% yield, >98% purity). ^{1}H NMR (400 MHz, D_2O) δ 7.01 (d, *J* = 8.2 Hz, 1H), 5.77 (d, *J* = 5.7 Hz, 1H), 5.67 (d, *J* = 8.2 Hz, 1H), 4.21 (t, *J* = 5.6 Hz, 1H), 4.10 (t, *J* = 5.0 Hz, 1H), 4.01 – 3.93 (m, 1H), 3.75 (dd, *J* = 12.7, 3.2 Hz, 1H), 3.66 (dd, *J* = 12.6, 4.6 Hz, 1H). ^{13}C NMR (101 MHz, D_2O) δ 150.91, 146.84, 131.69, 98.43, 88.00, 84.00, 72.54, 69.76, 61.05. MS: 260 $(M+H)^+$.

Chromatographic analysis.

HPLC analysis was performed on a 1200 Series Agilent LC system with an InfinityLab Poroshell 120 EC-C18, 4.6 mm x 100 mm, 4µm LC Column (Agilent). Substrates and products were eluted over 14 minutes using a gradient of 2-20% acetonitrile in 20 mM NH_4OAc pH 4.5 at 0.3 mL min^{-1}. Peaks were assigned by comparison to chemically synthesized standards and the peak areas were integrated using Agilent OpenLab software.

Supplementary Figures

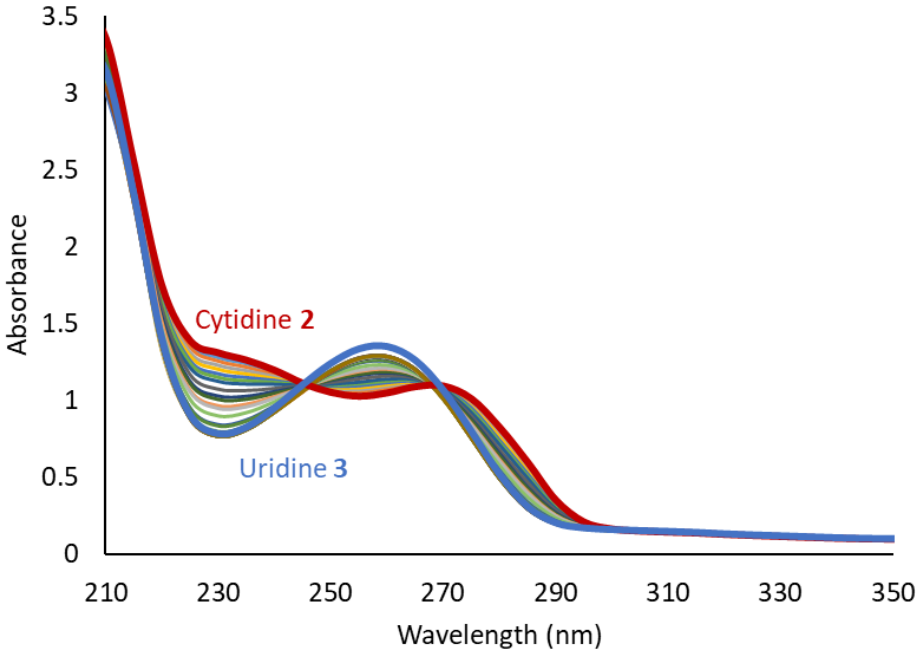


Figure S1: UV-Vis spectra showing the hydrolysis of cytidine **2** (100 µM) to uridine **3** catalysed by CD (10 nM). Spectra of authentic standards of **2** and **3** are shown in red and blue, respectively.

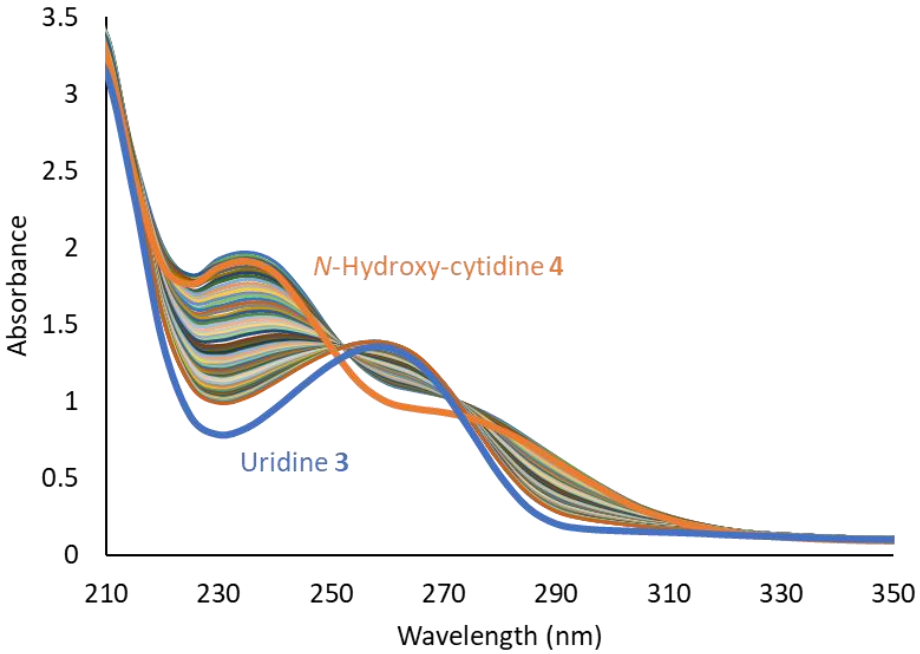


Figure S2: UV-Vis spectra showing the hydrolysis of *N*-hydroxy-cytidine **4** (100 µM) to uridine **3** catalysed by CD (100 nM). Spectra of authentic standards of **2** and **3** are shown in orange and blue, respectively.

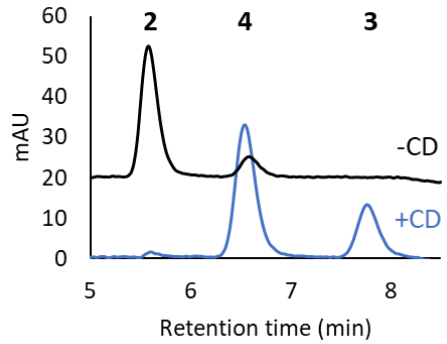


Figure S3: HPLC traces showing cytidine **2** (200 mM) conversion to *N*-hydroxy-cytidine **4** and uridine **3** by CD (20 µM) in the presence of 50% NH_2OH pH 7 (blue). Control reactions in the absence of CD show minimal conversion (black). Retention times of **2**, **3** and **4** are identical to authentic standards.

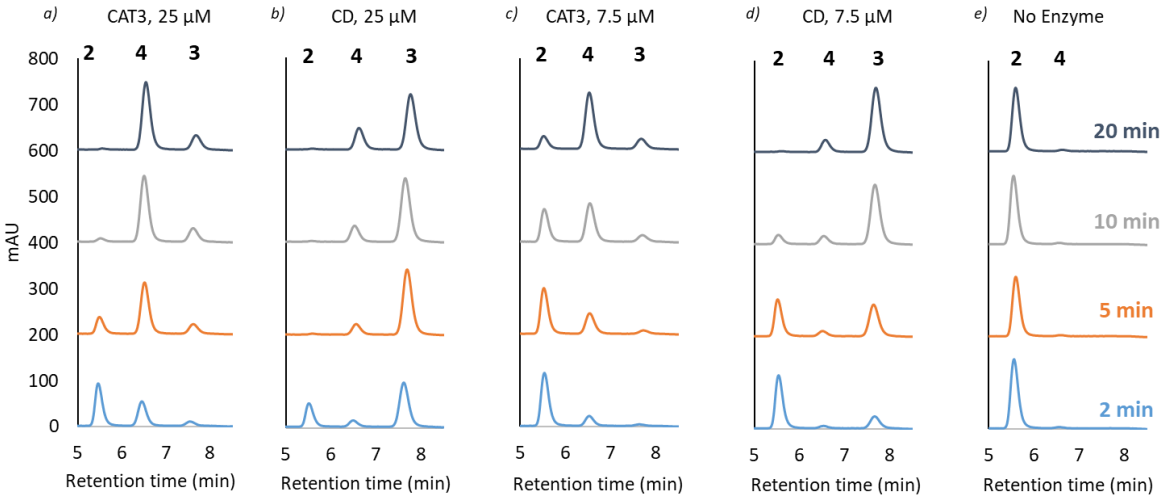
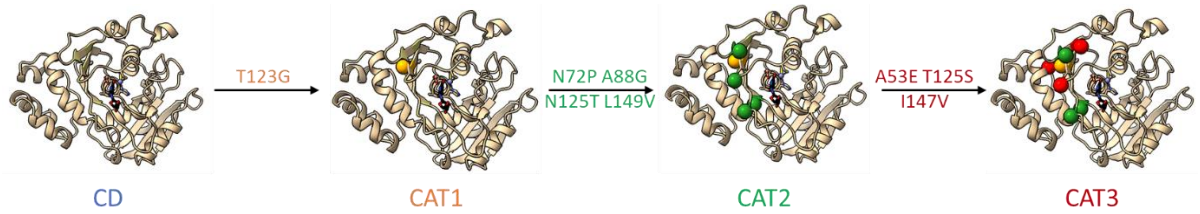


Figure S4: HPLC traces showing cytidine **2** (750 mM) conversion to *N*-hydroxy-cytidine **4** and uridine **3** at 2, 5, 10 and 20 minute time points catalyzed by a) CAT3 (25 µM), b) CD (25 µM), c) CAT3 (7.5 µM), d) CD (7.5 µM), e) no enzyme, in the presence of 50% NH_2OH pH 7. CAT3 operates predominantly *via* pathway A (direct conversion of **2** to **4**) whereas CD operates *via* pathway B (hydrolysis of **2** to **3** then equilibrium distribution of **4** and **3**). Retention times of **2**, **3** and **4** are identical to authentic standards.

Supplementary Table 1: Directed evolution workflow.



Round	Description	Number of clones screened	Beneficial mutations	Best variant
1	Saturation mutagenesis at 42 positions	3,700	A88W, T123G, N125S, N125G	**CAT1** = CD + T123G
2	Saturation mutagenesis at 10 positions	970	N72P, A88G, N125T, L149V	**CAT2** = CAT1 + N72P, A88G, N125T, L149V
3	Saturation mutagenesis at 36 positions	2,300	A53E, T125S, I147V, R152L	**CAT3** = CAT2 + A53E, T125S, I147V

Supplementary Table 2: Mass Spectra of CD and CAT3.

Variant	Expected Mass	Observed Mass
CD	32605.1	32604.2
CAT3	32533.0	32532.6

Supplementary Table 3: List of primers.

NAME	SEQUENCE
FLANKING PRIMERS	
CD_NDEI_F	CATGCATGCATATGCACCCGCGTTTT
CD_XHOI_R	CATGCATGCTCGAGGGCTAACAGCA
ROUND 1 SATURATION MUTAGENESIS	
CD_RD1_A61NNK_F	CTGCTTCCACTTGCTNNKGCTTGTGCACGTACC
CD_RD1_A61NNK_R	AGCAAGTGGAAGCAGTG
CD_RD1_C63NNK_F	CCACTTGCTGCAGCTNNKGCACGTACCCCTTTG
CD_RD1_C63NNK_R	AGCTGCAGCAAGTGGAA
CD_RD1_T66NNK_F	GCAGCTTGTGCACGTNNKCCTTTGTCCAACTTCAATGT
CD_RD1_T66NNK_R	ACGTGCACAAGCTGCA
CD_RD1_L68NNK_F	TGTGCACGTACCCCTNNKTCCAACTTCAATGTCGGT
CD_RD1_L68NNK_R	AGGGGTACGTGCACAAG
CD_RD1_S69NNK_F	GCACGTACCCCTTTGNNKAACTTCAATGTCGGTGCG
CD_RD1_S69NNK_R	CAAAGGGGTACGTGCACA
CD_RD1_N70NNK_F	CGTACCCCTTTGTCCNNKTTCAATGTCGGTGCGA
CD_RD1_N70NNK_R	GGACAAAGGGGTACGTG
CD_RD1_F71NNK_F	ACCCCTTTGTCCAACNNKAATGTCGGTGCGATTGC
CD_RD1_F71NNK_R	GTTGGACAAAGGGGTACG
CD_RD1_N72NNK_F	CCTTTGTCCAACTTCNNKGTCGGTGCGATTGCG
CD_RD1_N72NNK_R	GAAGTTGGACAAAGGGGT
CD_RD1_V73NNK_F	TTGTCCAACTTCAATNNKGGTGCGATTGCGCG
CD_RD1_V73NNK_R	ATTGAAGTTGGACAAAGGG
CD_RD1_G74NNK_F	TCCAACTTCAATGTCNNKGCGATTGCGCGTGG
CD_RD1_G74NNK_R	GACATTGAAGTTGGACAAAGG
CD_RD1_A75NNK_F	AACTTCAATGTCGGTNNKATTGCGCGTGGTGTA
CD_RD1_A75NNK_R	ACCGACATTGAAGTTGGA
CD_RD1_A88NNK_F	ACCTGGTATTTCGGGNNKAACATGGAGTTTATCGGGG
CD_RD1_A88NNK_R	CCCGAAATACCAGGTTCC
CD_RD1_N89NNK_F	TGGTATTTCGGGGCANNKATGGAGTTTATCGGGGC
CD_RD1_N89NNK_R	TGCCCCGAAATACCAGG
CD_RD1_M90NNK_F	TATTTCGGGGCAAACNNKGAGTTTATCGGGGCGACT
CD_RD1_M90NNK_R	GTTTGCCCCGAAATACCA
CD_RD1_E91NNK_F	TTCGGGGCAAACATGNNKTTTATCGGGGCGACTATG
CD_RD1_E91NNK_R	CATGTTTGCCCCGAAATA
CD_RD1_F92NNK_F	GGGGCAAACATGGAGNNKATCGGGGCGACTATG
CD_RD1_F92NNK_R	CTCCATGTTTGCCCCGA
CD_RD1_Q98NNK_F	ATCGGGGCGACTATGNNKCAAACCGTTCATGCTGAA
CD_RD1_Q98NNK_R	CATAGTCGCCCCGATAAACTC
CD_RD1_Q99NNK_F	GGGGCGACTATGCAGNNKACCGTTCATGCTGAACA
CD_RD1_Q99NNK_R	CTGCATAGTCGCCCC
CD_RD1_T100NNK_F	GCGACTATGCAGCAANNKGTTCATGCTGAACAGAGC
CD_RD1_T100NNK_R	TTGCTGCATAGTCGCC
CD_RD1_V101NNK_F	ACTATGCAGCAAACCNNKCATGCTGAACAGAGCGC
CD_RD1_V101NNK_R	GGTTTGCTGCATAGTCGC
CD_RD1_H102NNK_F	ATGCAGCAAACCGTTNNKGCTGAACAGAGCGCG

CD_RD1_H102NNK_R	AACGGTTTGCTGCATAGT
CD_RD1_A103NNK_F	CAGCAAACCGTTCATNNKGAACAGAGCGCGATC
CD_RD1_A103NNK_R	ATGAACGGTTTGCTGCAT
CD_RD1_Q105NNK_F	ACCGTTCATGCTGAANNKAGCGCGATCAGCCAT
CD_RD1_Q105NNK_R	TTCAGCATGAACGGTTTG
CD_RD1_S106NNK_F	GTTCATGCTGAACAGNNKGCGATCAGCCATGCG
CD_RD1_S106NNK_R	CTGTTCAGCATGAACGGT
CD_RD1_A107NNK_F	CATGCTGAACAGAGCNNKATCAGCCATGCGTGG
CD_RD1_A107NNK_R	GCTCTGTTCAGCATGAAC
CD_RD1_I122NNK_F	AAGGCATTAGCCGCGNNKACCGTGAATTACACGCC
CD_RD1_I122NNK_R	CGCGGCTAATGCCTTC
CD_RD1_T123NNK_F	GCATTAGCCGCGATCNNKGTGAATTACACGCCATGC
CD_RD1_T123NNK_R	GATCGCGGCTAATGCCT
CD_RD1_V124NNK_F	TTAGCCGCGATCACCNNKAATTACACGCCATGCG
CD_RD1_V124NNK_R	GGTGATCGCGGCTAATG
CD_RD1_N125NNK_F	GCCGCGATCACCGTGNNKTACACGCCATGCGGA
CD_RD1_N125NNK_R	CACGGTGATCGCGG
CD_RD1_Y126NNK_F	GCGATCACCGTGAATNNKACGCCATGCGGACA
CD_RD1_Y126NNK_R	ATTCACGGTGATCGCGG
CD_RD1_T127NNK_F	ATCACCGTGAATTACNNKCCATGCGGACACTGT
CD_RD1_T127NNK_R	GTAATTCACGGTGATCGC
CD_RD1_P128NNK_F	ACCGTGAATTACACGNNKTGCGGACACTGTCG
CD_RD1_P128NNK_R	CGTGTAATTCACGGTGATC
CD_RD1_C129NNK_F	GTGAATTACACGCCANNKGGACACTGTCGCCAG
CD_RD1_C129NNK_R	TGGCGTGTAATTCACGGT
CD_RD1_G130NNK_F	AATTACACGCCATGCNNKCACTGTCGCCAGTTTATG
CD_RD1_G130NNK_R	GCATGGCGTGTAATTCAC
CD_RD1_H131NNK_F	TACACGCCATGCGGANNKTGTCGCCAGTTTATGAATG
CD_RD1_H131NNK_R	TCCGCATGGCGTGTAATT
CD_RD1_C132NNK_F	ACGCCATGCGGACACNNKCGCCAGTTTATGAATGAGTT
CD_RD1_C132NNK_R	GTGTCCGCATGGCG
CD_RD1_R133NNK_F	CCATGCGGACACTGTNNKCAGTTTATGAATGAGTTGAACTCA
CD_RD1_R133NNK_R	ACAGTGTCCGCATGG
CD_RD1_Q134NNK_F	TGCGGACACTGTCGCNNKTTTATGAATGAGTTGAACTCAGG
CD_RD1_Q134NNK_R	GCGACAGTGTCCGC
CD_RD1_M136NNK_F	CACTGTCGCCAGTTTNNKAATGAGTTGAACTCAGGCT
CD_RD1_M136NNK_R	AAACTGGCGACAGTGTCC
CD_RD1_H148NNK_F	TTGGATCTTCGTATCNNKCTGCCGGGACGTGAA
CD_RD1_H148NNK_R	GATACGAAGATCCAAGCCT
CD_RD1_L149NNK_F	GATCTTCGTATCCACNNKCCGGGACGTGAAGC
CD_RD1_L149NNK_R	GTGGATACGAAGATCCAAGC
CD_RD1_P150NNK_F	CTTCGTATCCACCTGNNKGGACGTGAAGCCCAT
CD_RD1_P150NNK_R	CAGGTGGATACGAAGATCC
	ROUND 1 SHUFFLING
CD_RD1_A88W_F	ACCTGGTATTTCGGGTGGAACATGGAGTTTATCGGGG
CD_RD1_A88_R	CCCGAAATACCAGGTTCC
CD_RD1_T123G_N125S_F	GCATTAGCCGCGATCGGGGTGTCGTACACGCCATGCGGA
CD_RD1_T123_R	GATCGCGGCTAATGCCT

CD_RD1_T123G_N125G_F	GCATTAGCCGCGATCGGGGTGGGGTACACGCCATGCGGA
ROUND 2 SATURATION MUTAGENESIS	
CD_RD2_N125NNK_F	GCCGCGATCGGGGTGNNKTACACGCCATGCGGA
CD_RD2_N125NNK_R	CACCCCGATCGCGG
CD_RD2_I76NNK_F	TTCAATGTCGGTGCGNNKGCGCGTGGTGTAAG
CD_RD2_I76NNK_R	CGCACCGACATTGAA
CD_RD2_V124NNK_F	TTAGCCGCGATCGGGNNKAATTACACGCCATGCG
CD_RD2_V124NNK_R	CCCGATCGCGGCTAA
CD_RD2_I122NNK_F	AAGGCATTAGCCGCGNNKGGGGTGAATTACACGC
CD_RD2_I122NNK_R	CGCGGCTAATGCCTT
ROUND 2 SHUFFLING	
CD_RD2_A88G_F	ACCTGGTATTTCGGGGGCAACATGGAGTTTATCGGGG
CD_RD2_A88_R	CCCGAAATACCAGGTTCC
CD_RD2_N72P_F	CCTTTGTCCAACTTCCCTGTCGGTGCGATTGCG
CD_RD2_N72_R	GAAGTTGGACAAAGGGGT
CD_RD2_N125D_F	GCCGCGATCGGGGTGGATTACACGCCATGCGGA
CD_RD2_N125_R	CACCCCGATCGCGGC
CD_RD2_N125T_F	GCCGCGATCGGGGTGACGTACACGCCATGCGGA
CD_RD2_L149V_F	GATCTTCGTATCCACGTGCCGGGACGTGAAGC
CD_RD2_L149_R	GTGGATACGAAGATCCAAGC
ROUND 3 SATURATION MUTAGENESIS	
CD_RD3_P72NDT_F	CCTTTGTCCAACTTCNDTGTCGGTGCGATTGC
CD_RD3_P72VHG_F	CCTTTGTCCAACTTCVHGGTCGGTGCGATTGC
CD_RD3_P72TGG_F	CCTTTGTCCAACTTCTGGGTCGGTGCGATTGC
CD_RD3_P72_R	GAAGTTGGACAAAGGGG
CD_RD3_G88NDT_F	ACCTGGTATTTCGGGNDTAACATGGAGTTTATCGGG
CD_RD3_G88VHG_F	ACCTGGTATTTCGGGVHGAACATGGAGTTTATCGGG
CD_RD3_G88TGG_F	ACCTGGTATTTCGGGTGGAACATGGAGTTTATCGGG
CD_RD3_G88_R	CCCGAAATACCAGGTT
CD_RD3_T125NDT_F	GCCGCGATCGGGGTGNDTTACACGCCATGCGGA
CD_RD3_T125VHG_F	GCCGCGATCGGGGTGVHGTACACGCCATGCGGA
CD_RD3_T125TGG_F	GCCGCGATCGGGGTGTGGTACACGCCATGCGGA
CD_RD3_T125_R	CACCCCGATCGCGG
CD_RD3_D50NNK_F	ACGGGACTGGACGAANNKGCTCTTGCGTTTGCAC
CD_RD3_D50_R	TTCGTCCAGTCCCGT
CD_RD3_A53NNK_F	GACGAAGATGCTCTTNNKTTTGCACTGCTTCCAC
CD_RD3_A53_R	AAGAGCATCTTCGTCCA
CD_RD3_F54NNK_F	GAAGATGCTCTTGCGNNKGCACTGCTTCCACTT
CD_RD3_F54_R	CGCAAGAGCATCTTCG
CD_RD3_L57NNK_F	CTTGCGTTTGCACTGNNKCCACTTGCTGCAGCT
CD_RD3_L57_R	CAGTGCAAACGCAAGA
CD_RD3_A60NNK_F	GCACTGCTTCCACTTNNKGCAGCTTGTGCACG
CD_RD3_A60_R	AAGTGGAAGCAGTGCA
CD_RD3_A61NNK_F	CTGCTTCCACTTGCTNNKGCTTGTGCACGTACC
CD_RD3_A61_R	AGCAAGTGGAAGCAGT
CD_RD3_C63NNK_F	CCACTTGCTGCAGCTNNKGCACGTACCCCTTTG
CD_RD3_C63_R	AGCTGCAGCAAGTGG

CD_RD3_A64NNK_F	CTTGCTGCAGCTTGTNNKCGTACCCCTTTGTCC
CD_RD3_A64_R	ACAAGCTGCAGCAAG
CD_RD3_T66NNK_F	GCAGCTTGTGCACGTNNKCCTTTGTCCAACTTCCC
CD_RD3_T66_R	ACGTGCACAAGCTG
CD_RD3_N70NNK_F	CGTACCCCTTTGTCCNNKTTCCCTGTCGGTGC
CD_RD3_N70_R	GGACAAAGGGGTACGT
CD_RD3_F71NNK_F	ACCCCTTTGTCCAACNNKCCTGTCGGTGCGATT
CD_RD3_F71_R	GTTGGACAAAGGGGTAC
CD_RD3_V73NNK_F	TTGTCCAACTTCCCTNNKGGTGCGATTGCGC
CD_RD3_V73_R	AGGGAAGTTGGACAAAGG
CD_RD3_G74NNK_F	TCCAACTTCCCTGTCNNKGCGATTGCGCGTG
CD_RD3_G74_R	GACAGGGAAGTTGGACA
CD_RD3_I76NNK_F	TTCCCTGTCGGTGCGNNKGCGCGTGGTGTAAG
CD_RD3_I76_R	CGCACCGACAGGGA
CD_RD3_R78NNK_F	GTCGGTGCGATTGCGNNKGGTGTAAGCGGAACC
CD_RD3_R78_R	CGCAATCGCACCGA
CD_RD3_F86NNK_F	AGCGGAACCTGGTATNNKGGGGGCAACATGGAG
CD_RD3_F86_R	ATACCAGGTTCCGCTT
CD_RD3_G87NNK_F	GGAACCTGGTATTTCNNKGGCAACATGGAGTTTATCG
CD_RD3_G87_R	GAAATACCAGGTTCCGCT
CD_RD3_N89NNK_F	TGGTATTTCGGGGGCNNKATGGAGTTTATCGGGGC
CD_RD3_N89_R	GCCCCCGAAATACCA
CD_RD3_M90NNK_F	TATTTCGGGGGCAACNNKGAGTTTATCGGGGCGA
CD_RD3_M90_R	GTTGCCCCCGAAATAC
CD_RD3_A103NNK_F	CAGCAAACCGTTCATNNKGAACAGAGCGCGATC
CD_RD3_A103_R	ATGAACGGTTTGCTGC
CD_RD3_S106NNK_F	GTTCATGCTGAACAGNNKGCGATCAGCCATGC
CD_RD3_S106_R	CTGTTCAGCATGAACGG
CD_RD3_A121NNK_F	GAGAAGGCATTAGCCNNKATCGGGGTGACGTAC
CD_RD3_A121_R	GGCTAATGCCTTCTCC
CD_RD3_I122NNK_F	AAGGCATTAGCCGCGNNKGGGGTGACGTACACG
CD_RD3_I122_R	CGCGGCTAATGCCT
CD_RD3_V124NNK_F	TTAGCCGCGATCGGGNNKACGTACACGCCATGC
CD_RD3_V124_R	CCCGATCGCGGCTA
CD_RD3_Y126NNK_F	GCGATCGGGGTGACGNNKACGCCATGCGGAC
CD_RD3_Y126_R	CGTCACCCCGATCG
CD_RD3_T127NNK_F	ATCGGGGTGACGTACNNKCCATGCGGACACTGT
CD_RD3_T127_R	GTACGTCACCCCGAT
CD_RD3_I147NNK_F	GGCTTGGATCTTCGTNNKCACGTGCCGGGA
CD_RD3_I147_R	ACGAAGATCCAAGCCT
CD_RD3_H148NNK_F	TTGGATCTTCGTATCNNKGTGCCGGGACGTG
CD_RD3_H148_R	GATACGAAGATCCAAGCCT
CD_RD3_P150NNK_F	CTTCGTATCCACGTGNNKGGACGTGAAGCCCAT
CD_RD3_P150_R	CACGTGGATACGAAGATCC
CD_RD3_G151NNK_F	CGTATCCACGTGCCGNNKCGTGAAGCCCATGCT
CD_RD3_G151_R	CGGCACGTGGATACG
CD_RD3_R152NNK_F	ATCCACGTGCCGGGANNKGAAGCCCATGCTTTACG
CD_RD3_R152_R	TCCCGGCACGTGGA

CD_RD3_E153NNK_F	CACGTGCCGGGACGTNNKGCCCATGCTTTACGTGACT
CD_RD3_E153_R	ACGTCCCGGCACGT
CD_RD3_H155NNK_F	CCGGGACGTGAAGCCNNKGCTTTACGTGACTACTTGC
CD_RD3_H155_R	GGCTTCACGTCCCG
CD_RD3_Y160NNK_F	CATGCTTTACGTGACNNKTTGCCTGATGCCTTC
CD_RD3_Y160_R	GTCACGTAAAGCATGGG
CD_RD3_N248NNK_F	GCGCTGATTCTTTTANNKTTGAAGGGCTATGACTACC
CD_RD3_N248_R	TAAAAGAATCAGCGCTCCT
CD_RD3_L249NNK_F	CTGATTCTTTTAAATNNKAAGGGCTATGACTACCC
CD_RD3_L249_R	ATTTAAAAGAATCAGCGCTCC
	ROUND 3 SHUFFLING
CD_RD3_T125S_F	GCCGCGATCGGGGTGAGTTACACGCCATGCGGA
CD_RD3_T125_R	CACCCCGATCGCGGC
CD_RD3_A53E_F	GACGAAGATGCTCTTGAGTTTGCACTGCTTCCAC
CD_RD3_A53_R	AAGAGCATCTTCGTCCA
CD_RD3_I147V_F	GGCTTGGATCTTCGTGTTCACGTGCCGGGA
CD_RD3_I147_R	ACGAAGATCCAAGCCT
CD_RD3_R152L_F (I147)	ATCCACGTGCCGGGACTGGAAGCCCATGCTTTACG
CD_RD3_R152L_F (I147V)	GTTCACGTGCCGGGACTGGAAGCCCATGCTTTACG
CD_RD3_R152_R (I147)	TCCCGGCACGTGGAT
CD_RD3_R152_R (I147V)	TCCCGGCACGTGAAC

DNA and protein sequence of CAT3

Mutations from wild-type CD: A53E N72P A88G T123G T125S I147V L149V

ATGCACCCGCGTTTTCAGACGGCGTTCGCCCAGTTAGCTGACAATCTGCAGAGTGCCTTGGAGCCTATCTTGG
CAGATAAGTATTTTCCGGCCCTGCTGACCGGGGAGCAAGTTTCCTCCTTGAAAAGCGCTACGGGACTGGACGA
AGATGCTCTTGAGTTTGCACTGCTTCCACTTGCTGCAGCTTGTGCACGTACCCCTTTGTCCAACTTCCCTGTCGG
TGCGATTGCGCGTGGTGTAAGCGGAACCTGGTATTTCGGGGGCAACATGGAGTTTATCGGGGCGACTATGCA
GCAAACCGTTCATGCTGAACAGAGCGCGATCAGCCATGCGTGGCTGAGCGGGGAGAAGGCATTAGCCGCGA
TCGGGGTGAGTTACACGCCATGCGGACACTGTCGCCAGTTTATGAATGAGTTGAACTCAGGCTTGGATCTTCG
TGTTCACGTGCCGGGACGTGAAGCCCATGCTTTACGTGACTACTTGCCTGATGCCTTCGGTCCCAAGGACTTA
GAAATCAAGACCCTTCTTATGGACGAACAAGACCACGGATATGCGCTTACTGGAGATGCACTGTCCCAGGCTG
CAATCGCTGCGGCCAACCGTAGCCACATGCCCTACTCCAAGAGTCCGAGCGGTGTCGCACTGGAGTGTAAAG
ATGGACGTATTTTCAGCGGCTCATACGCGGAGAATGCGGCTTTCAATCCCACTCTTCCCCCGTTACAAGGAGC
GCTGATTCTTTTAAATTTGAAGGGCTATGACTACCCAGACATTCAACGTGCAGTGCTTGCCGAGAAGGCGGAC
GCACCATTGATCCAATGGGATGCTACCAGCGCAACTCTGAAAGCTTTAGGGTGTCACAGCATCGATCGCGTGC
TGTTAGCCCTCGAGCACCACCACCACCACCAC

MHPRFQTAFAQLADNLQSALEPILADKYFPALLTGEQVSSLKSATGLDEDAL**E**FALLPLAAACARTPLSNF**P**VGAIAR
GVSGTWYFG**G**NMEFIGATMQQTVHAEQSAISHAWLSGEKALAAI**G**V**S**YTPCGHCRQFMNELNSGLDLR**V**H**V**PG
REAHALRDYLPDAFGPKDLEIKTLLMDEQDHGYALTGDALSQAAIAAANRSHMPYSKSPSGVALECKDGRIFSGSY
AENAAFNPTLPPLQGALILLNLKGYDYPDIQRAVLAEKADAPLIQWDATSATLKALGCHSIDRVLLALEHHHHHH