An Engineered Cytidine Deaminase for Biocatalytic Production of a Key Intermediate of the Covid-19 Antiviral Molnupiravir

Ashleigh J. Burke[a], William R. Birmingham[a, ‡], Ying Zhuo[a, ‡], Thomas W. Thorpe[a, ‡], Bruna Zucoloto da Costa[a], Rebecca Crawshaw[a], Ian Rowles[a], James D. Finnigan[b], Carl Young[b], Gregory M. Holgate[c], Mark P. Muldowney[c], 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, NE49 9HA, UK

[c] Sterling Pharma Solutions, Sterling Place, Dudley, Northumberland, NE23 7QG, UK

[‡] These authors contributed equally

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*-optimized gene (synthesized 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}. To determine whether the protein is modified in the presence of NH_2OH, CD and it's variants (2.5 μM) were incubated in 10% NH2OH (~3M, pH 7) for 20 minutes prior to buffer exchanging. 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.

Recombination of beneficial mutations by overlap extension PCR.

After each round of evolution, beneficial diversity was combined by rational recombination 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% (~300 mM) NH_2OH pH 7. 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 biotransformations.

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. CD1.3 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 cold acidified water (10 mL, 10% AcOH) and then freeze dried overnight to give the monohydrated product as a free flowing white powder (4.85 g, 85% yield, >98% purity). A Karl Fischer titration was used to confirm the product crystallizes as a monohydrate (6.7 wt% H_2O) as described previously in the literature (reference 7 in the manuscript). ^{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 $(M+H)^+$.

Cytidine (37.6 g, 770 mM, 188 g L^{-1}) was dissolved in hydroxylamine (200 mL, 10 % in water, pH 7) in a 500 mL flask and cooled to 4 °C. CD1.3 cell free extract (376 mg, 1 % wt) was added to the reaction mixture. The reaction was left on a tabletop shaker 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 10 minutes and the supernatant removed. The remaining solid was washed with cold acidified water (30 mL, 10% AcOH) and then freeze dried overnight to give the monohydrated product as a white powder (40.8 g, 89 % yield, >96 % purity). ^{1}H NMR (400 MHz, D_2O) δ 7.00 (d, *J* = 8.3 Hz, 1H), 5.79 (d, *J* = 5.6 Hz, 1H), 5.68 (d, *J* = 8.2 Hz, 1H), 4.23 (t, *J* = 5.7 Hz, 1H), 4.12 (t, *J* = 5.0 Hz, 1H), 3.99 (q, *J* = 4.2 Hz, 1H), 3.81 – 3.61 (m, 2H). ^{13}C NMR (101 MHz, D_2O) δ 151.27, 146.63, 131.34, 98.69, 87.95, 84.00, 72.50, 69.79, 61.07. MS: 260 $(M+H)^+$.

Cytidine (169.2 g, 0.69 mol, 1 eq) was dissolved in 10% hydroxylamine (900 mL, 2.92 mol, 4.2 eq, pH 7) in a 2L Mettler Toledo AP01 jacketed vessel and cooled to 4°C. CD1.3 crude lysate (730 mg, 0.45 % wt) was suspended in pH 7 phosphate buffer and brought to a volume of 6.75 mL, this solution was added to the reaction mix. The reaction was stirred at 300 rpm for 3 hours at 4 °C during which time the *N*-hydroxy-cytidine 4 product crystallized in situ. The reaction mixture was filtered through 20 μm cloth, and the remaining solid was washed with 10% AcOH in water (340 mL, 2 vol). The solid was then dried in a vacuum oven at 30°C for 18 hours giving a white solid (136.8 g, 71% yield, >95% purity by NMR & HPLC). ^{1}H NMR (400 MHz, D_2O) δ 7.1 (d, 1H), 5.8 (d, 1H), 5.6 (d, 1H), 4.1 (t, 1H), 4.1 (t, 1H), 3.9 (q, 1H), 3.7 (dd, 1H), 3.6 (dd, 1H).

Chromatographic analysis.

HPLC and LCMS analysis was performed on a 1200 Series Agilent LC or LC/MSD 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 or analysis of MS data and the peak areas were integrated using Agilent OpenLab software. 2'deoxycytidine and resulting products were eluted over 64 minutes using a gradient of 1.6-3% acetonitrile.

Supplementary Figures

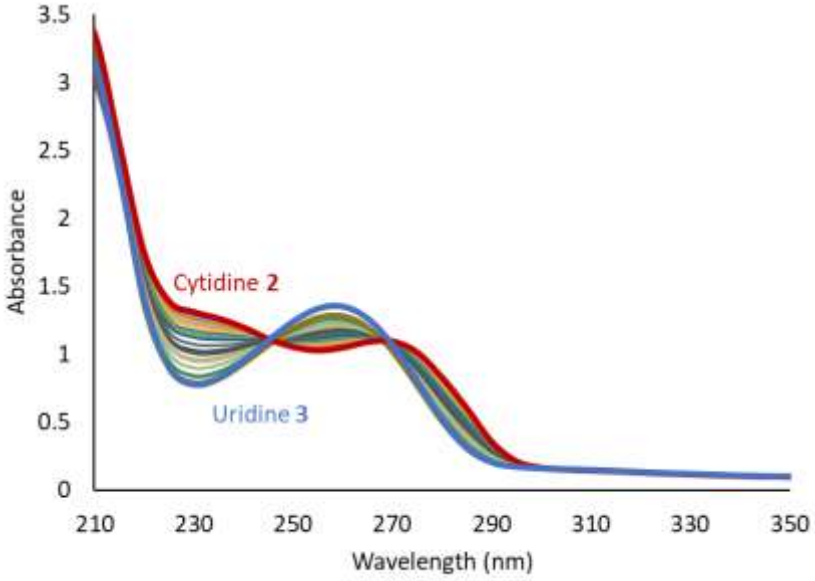


Figure S1: UV-Vis spectra showing the hydrolysis of cytidine 2 (100 µM) to uridine 3 catalyzed by CD (10 nM). Spectra of authentic standards of 2 and 3 are shown in red and blue, respectively.

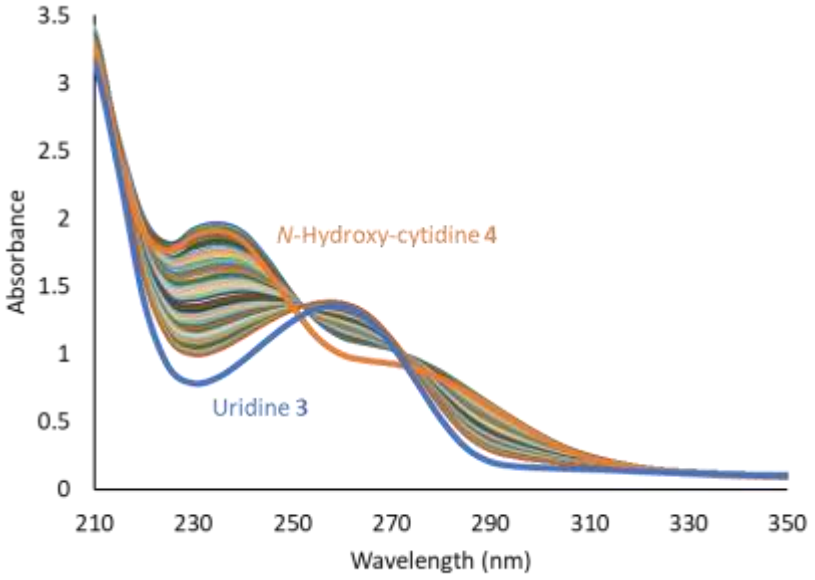


Figure S2: UV-Vis spectra showing the hydrolysis of *N*-hydroxy-cytidine 4 (100 µM) to uridine 3 catalyzed by CD (100 nM). Spectra of authentic standards of 4 and 3 are shown in orange and blue, respectively.

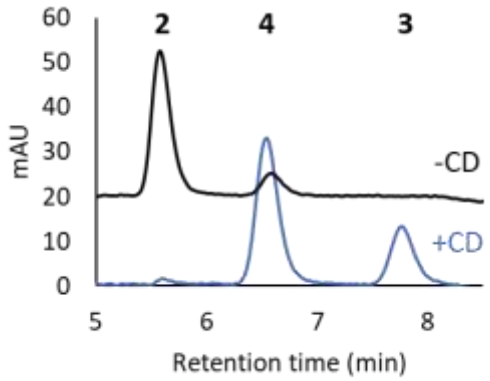
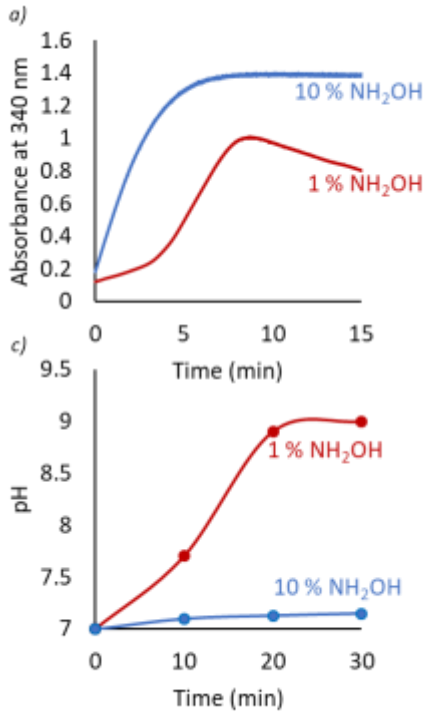


Figure S3: HPLC traces showing cytidine 2 (200 mM) conversion to *N*-hydroxy-cytidine 4 and uridine 3 by CD (20 μM) in the presence of 50% NH_2OH pH 7 (blue). Control reactions in the absence of CD show minimal conversion (black). Retention times of 2, 3 and 4 are identical to authentic standards.



b)

Variant	Expected Mass	Observed Mass	Mass After 10% NH_2OH Incubation
CD	32605.1	32604.2	32605.2
CD1.3	32533.0	32532.6	32532.8

d)

Compound	Concentration (mM)	pH	Absorbance at 340 nm
N-hydroxy-cytidine **4**	25	7	0.73
		9	2.63
Cytidine **2**	25	7	0.075
		9	0.075

Figure S4: a) The conversion of cytidine 2 (50 mM) to *N*-hydroxy-cytidine 4 catalyzed by CD1.3 (2.5 μM) in 1% NH_2OH (~300 mM, pH 7, red) and 10% NH_2OH (~3 M, pH 7, blue), monitored by increasing absorbance at 340 nm. An apparent initial lag phase is observed at low (1%) NH_2OH concentrations. b) Expected and observed masses of CD and CD1.3. The data shows there is no observed changed to the mass of the protein following incubation in 10% NH_2OH. This suggests the apparent lag phase in S4a is not caused by enzyme modification during the reaction. c) Observed pH changes under reaction conditions of 1% and 10% NH_2OH. A pH change from 7 to 9 is observed at 1% NH_2OH which is not seen at 10% NH_2OH. d) The extinction coefficient of *N*-hydroxy-cytidine 4 is sensitive to pH changes, this is not observed for cytidine 2.

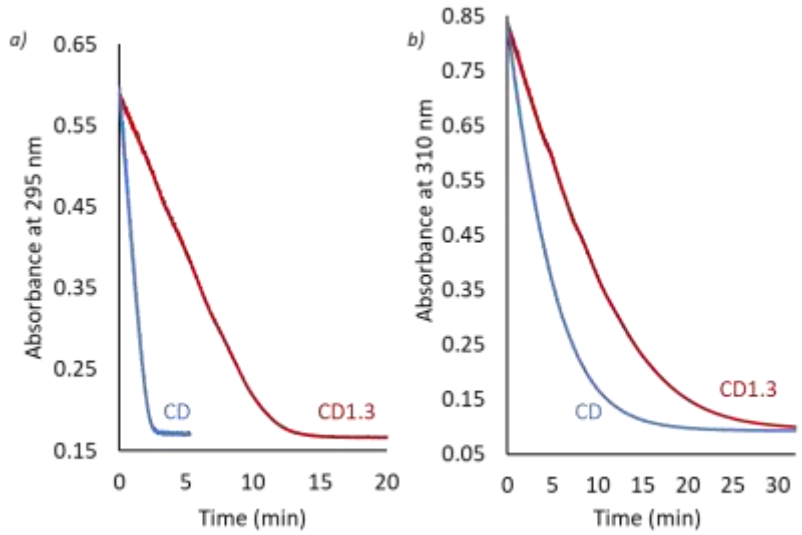
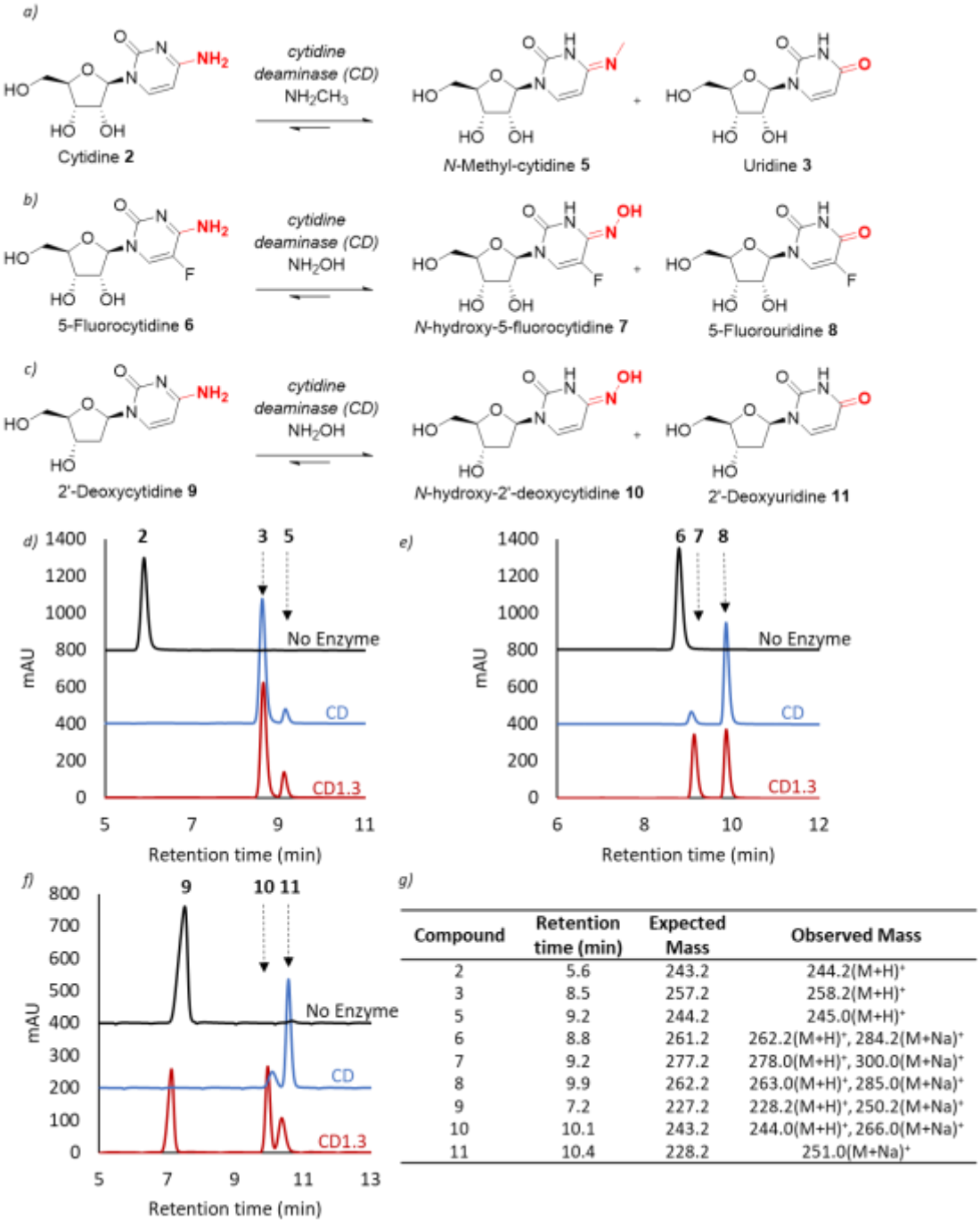


Figure S5: a) The conversion of cytidine 2 (1 mM) to uridine 3 catalyzed by CD1.3 (50 nM, red) and CD (50 nM, blue) in PBS (pH 7.4) monitored by decreasing absorbance at 295 nm. k_{obs} = 30 s^{-1} and 140 s^{-1} for CD1.3 and CD, respectively. b) The conversion of *N*-hydroxy-cytidine 4 (1 mM) to uridine 3 catalyzed by CD1.3 (500 nM, red) and CD (500 nM, blue) in PBS (pH 7.4) monitored by decreasing absorbance at 310 nm. k_{obs} = 2 s^{-1} and 5 s^{-1} for CD1.3 and CD, respectively. Reduced rates of hydrolysis contribute to improved performance of CD1.3 for the production of *N*-hydroxy-cytidine 4.



g)

Compound	Retention time (min)	Expected Mass	Observed Mass
2	5.6	243.2	244.2(M+H)$^+$
3	8.5	257.2	258.2(M+H)$^+$
5	9.2	244.2	245.0(M+H)$^+$
6	8.8	261.2	262.2(M+H)$^+$, 284.2(M+Na)$^+$
7	9.2	277.2	278.0(M+H)$^+$, 300.0(M+Na)$^+$
8	9.9	262.2	263.0(M+H)$^+$, 285.0(M+Na)$^+$
9	7.2	227.2	228.2(M+H)$^+$, 250.2(M+Na)$^+$
10	10.1	243.2	244.0(M+H)$^+$, 266.0(M+Na)$^+$
11	10.4	228.2	251.0(M+Na)$^+$

Figure S6: a) Reaction scheme showing the conversion of cytidine 2 to *N*-methylcytidine 5 and uridine 3 catalyzed by cytidine deaminase in the presence of methylamine (NH_2CH_3). b) Reaction scheme showing the conversion of 5-fluorocytidine 6 to *N*-hydroxy-5-fluorocytidine 7 and 5-fluorouridine 8 catalyzed by cytidine deaminase in the presence of NH_2OH. c) Reaction scheme showing the conversion of 2'-deoxycytidine 9 to *N*-hydroxy-2'-deoxycytidine 10 and 2'-deoxyuridine 11 catalyzed by cytidine deaminase in the presence of NH_2OH. d) HPLC traces showing the conversion of cytidine 2 (50 mM) to *N*-methylcytidine 5 and uridine 3 catalyzed by CD1.3 (1 µM, red), CD (1 µM, blue) and no enzyme (black) in the presence of NH_2CH_3 (50% w/v) after 5 minutes. CD1.3 shows improved activity for the production of 5 compared to CD. e) HPLC traces showing the conversion of 5-fluorocytidine 6 (50 mM) to *N*-hydroxy-5-fluorocytidine 7 and 5-fluorouridine 8 catalyzed by CD1.3 (1 µM, red), CD (1 µM, blue) and no enzyme (black) in the presence of 10% NH_2OH (~3 M, pH 7) after 5 minutes. CD1.3 shows improved activity for the production of 7 compared to CD. f) HPLC traces showing the conversion of 2'-deoxycytidine 9 (50 mM) to *N*-hydroxy-2'-deoxycytidine 10 and 2'-deoxyuridine 11 catalyzed by CD1.3 (1 µM, red), CD (1 µM, blue) and

no enzyme (black) in the presence of 10% NH_2OH (~3 M, pH 7) after 30 minutes. CD1.3 shows improved activity for the production of 10 compared to CD. g) Compound identities were confirmed by LCMS. Expected and observed masses of compounds are shown in a table.

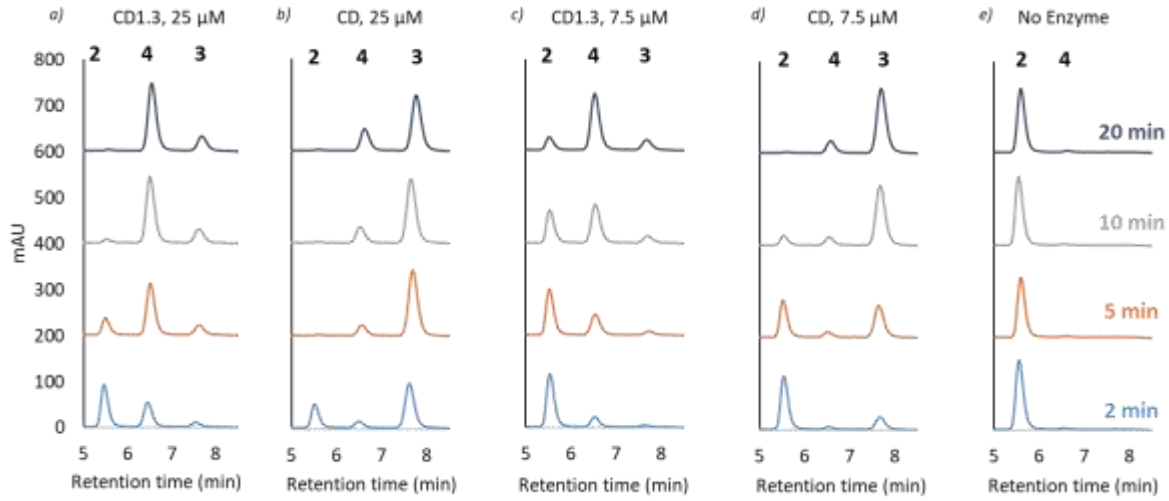


Figure S7: HPLC traces showing cytidine 2 (750 mM) conversion to *N*-hydroxy-cytidine 4 and uridine 3 at 2, 5, 10 and 20 minute time points catalyzed by a) CD1.3 (25 µM), b) CD (25 µM), c) CD1.3 (7.5 µM), d) CD (7.5 µM), e) no enzyme, in the presence of 10% (~3 M) NH_2OH pH 7. CD1.3 operates predominantly 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.

Figure S8: a) A 1 L scale biotransformation was carried out in a 2L Mettler Toledo AP01 jacketed vessel cooled to 4°C. b) *N*-Hydroxy-cytidine 4 crystallized in situ to give 137 g of a white solid in 71% isolated yield and >95% purity following filtration, washing and drying.

Supplementary Table 1: Directed evolution workflow.

Round	Description	Number of clones screened	Beneficial mutations	Best variant
1a	Saturation mutagenesis at 42 positions	3,700	A88W, T123G, N125S, N125G	CD1.1 = CD + T123G
1b	Rational combination of beneficial diversity by overlap extension PCR	7	No improved variants	CD1.1 = CD + T123G
2a	Saturation mutagenesis at 10 positions	970	N72P, A88G, N125T, L149V	CD1.1 + L149V
2b	Rational combination of beneficial diversity by overlap extension PCR	7	All combinations of variants showed some improvement in activity	CD1.2 = CD1.1 + N72P, A88G, N125T, L149V
3a	Saturation mutagenesis at 36 positions	2,300	A53E, T125S, I147V, R152L	CD1.2 + I147V
3b	Rational combination of beneficial diversity by overlap extension PCR	11	All combinations of variants showed some improvement in activity	CD1.3 = CD1.2 + A53E, T125S, I147V

Supplementary Table 2: List of primers.

NAME	SEQUENCE
FLANKING PRIMERS	
CD_NDEI_F	CATGCATGCATATGCACCCGCGTTTT
CD_XHOI_R	CATGCATGCTCGAGGGCTAACAGCA
ROUND 1A 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 1B RECOMBINATION
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 2A 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 2B RECOMBINATION
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 3A 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 3B RECOMBINATION
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 wild-type cytidine deaminase

CD - EC 3.5.4.5, UniProt accession: P0ABF6, NCBI protein accession code: WP_000553555.1

ATGCACCCGCGTTTTCAGACGGCGTTCGCCCAGTTAGCTGACAATCTGCAGAGTGCCTTGGAGCCTATCTTGGCAGAT
AAGTATTTTCCGGCCCTGCTGACCGGGGAGCAAGTTTCCTCCTTGAAAAGCGCTACGGGACTGGACGAAGATGCTCTT
GCGTTTGCACTGCTTCCACTTGCTGCAGCTTGTGCACGTACCCCTTTGTCCAACTTCAATGTCGGTGCGATTGCGCGT
GGTGTAAGCGGAACCTGGTATTTCGGGGCAAACATGGAGTTTATCGGGGCGACTATGCAGCAAACCGTTCATGCTGAA
CAGAGCGCGATCAGCCATGCGTGGCTGAGCGGGGAGAAGGCATTAGCCGCGATCACCGTGAATTACACGCCATGCGGA
CACTGTCGCCAGTTTATGAATGAGTTGAACTCAGGCTTGGATCTTCGTATCCACCTGCCGGGACGTGAAGCCCATGCT
TTACGTGACTACTTGCCTGATGCCTTCGGTCCCAAGGACTTAGAAATCAAGACCCTTCTTATGGACGAACAAGACCACG
GATATGCGCTTACTGGAGATGCACTGTCCCAGGCTGCAATCGCTGCGGCCAACCGTAGCCACATGCCCTACTCCAAGA
GTCCGAGCGGTGTCGCACTGGAGTGTAAAGATGGACGTATTTTCAGCGGCTCATACGCGGAGAATGCGGCTTTCAATC
CCACTCTTCCCCCGTTACAAGGAGCGCTGATTCTTTTAAATTTGAAGGGCTATGACTACCCAGACATTCAACGTGCAGT
GCTTGCCGAGAAGGCGGACGCACCATTGATCCAATGGGATGCTACCAGCGCAACTCTGAAAGCTTTAGGGTGTCACAG
CATCGATCGCGTGCTGTTAGCCCTCGAGCACCACCACCACCACCAC

MHPRFQTAFAQLADNLQSALEPILADKYFPALLTGEQVSSLKSATGLDEDALAFALLPLAAACARTPLSNFNVGAIARGVSGTWY
FGANMEFIGATMQQTVHAEQSAISHAWLSGEKALAAITVNYTPCGHCRQFMNELNSGLDLRIHLPGREAHALRDYLPDAFGPK
DLEIKTLLMDEQDHGYALTGDALSQAAIAAANRSHMPYSKSPSGVALECKDGRIFSGSYAENAAFNPTLPPLQGALILLNLKGYD
YPDIQRAVLAEKADAPLIQWDATSATLKALGCHSIDRVLLALEHHHHHH

DNA and protein sequence of CD1.3

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

ATGCACCCGCGTTTTCAGACGGCGTTCGCCCAGTTAGCTGACAATCTGCAGAGTGCCTTGGAGCCTATCTTGGCAGAT
AAGTATTTTCCGGCCCTGCTGACCGGGGAGCAAGTTTCCTCCTTGAAAAGCGCTACGGGACTGGACGAAGATGCTCTT
GAGTTTGCACTGCTTCCACTTGCTGCAGCTTGTGCACGTACCCCTTTGTCCAACTTCCCTGTCGGTGCGATTGCGCGT
GGTGTAAGCGGAACCTGGTATTTCGGGGGCAACATGGAGTTTATCGGGGCGACTATGCAGCAAACCGTTCATGCTGAA
CAGAGCGCGATCAGCCATGCGTGGCTGAGCGGGGAGAAGGCATTAGCCGCGATCGGGGTGAGTTACACGCCATGCGG
ACACTGTCGCCAGTTTATGAATGAGTTGAACTCAGGCTTGGATCTTCGTGTTCACGTGCCGGGACGTGAAGCCCATGC
TTTACGTGACTACTTGCCTGATGCCTTCGGTCCCAAGGACTTAGAAATCAAGACCCTTCTTATGGACGAACAAGACCAC
GGATATGCGCTTACTGGAGATGCACTGTCCCAGGCTGCAATCGCTGCGGCCAACCGTAGCCACATGCCCTACTCCAAG
AGTCCGAGCGGTGTCGCACTGGAGTGTAAAGATGGACGTATTTTCAGCGGCTCATACGCGGAGAATGCGGCTTTCAAT
CCCACTCTTCCCCCGTTACAAGGAGCGCTGATTCTTTTAAATTTGAAGGGCTATGACTACCCAGACATTCAACGTGCAG
TGCTTGCCGAGAAGGCGGACGCACCATTGATCCAATGGGATGCTACCAGCGCAACTCTGAAAGCTTTAGGGTGTCACA
GCATCGATCGCGTGCTGTTAGCCCTCGAGCACCACCACCACCACCAC

MHPRFQTAFAQLADNLQSALEPILADKYFPALLTGEQVSSLKSATGLDEDALEFALLPLAAACARTPLSNFPVGAIARGVSGTWYF
GGNMEFIGATMQQTVHAEQSAISHAWLSGEKALAAIGVSYTPCGHCRQFMNELNSGLDLRVHVPGREAHALRDYLPDAFGPK
DLEIKTLLMDEQDHGYALTGDALSQAAIAAANRSHMPYSKSPSGVALECKDGRIFSGSYAENAAFNPTLPPLQGALILLNLKGYD
YPDIQRAVLAEKADAPLIQWDATSATLKALGCHSIDRVLLALEHHHHHH