Electronic Supplementary Material (ESI) for Medicinal Chemistry Communications

Electronic Supplementary Information

Rationally designed *Escherichia coli* cytosine deaminase mutants with improved specificity towards the prodrug 5-fluorocytosine for potential gene therapy applications

V. Kohila [a], *Amit Jaiswal* [b], *Siddhartha Sankar Ghosh* [*, a, b]

[a] Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati-39, Assam, India

[b] Centre for Nanotechnology, Indian Institute of Technology Guwahati, Guwahati-39, Assam, India

*Corresponding author phone: +0361-258-2206; fax: +0361-258-2249;

email: sghosh@iitg.ernet.in

Experimental Section

Materials. pORF codA::upp plasmid was purchased from *Invivogen*. Oligonucleotides used to amplify the bCD gene, Glutathione-agarose beads, L-glutathione-reduced and Bradford reagent were procured from Sigma Aldrich Pvt. Ltd., USA. Restriction enzymes used for cloning and screening of mutants were purchased from New England Biolabs (Beverly, MA).

Bacterial strains. *E.coli* DH5 alpha was used for cloning and *E.coli* BL21 (DE3) was used for the expression studies of the WT and the mutant enzymes.

Construction of the WT CD vector. The cytosine deaminase gene was PCR amplified from pORF codA:upp (*Invivogen*) vector using the primers 5'

CGCGGATCCATGGTGTCGAATAACGC3' and 5' CCGCTCGAGTGACGCCATTAGCTCCGCTG 3' with BamHI and XhoI overhangs under the conditions of denaturation at 94 ºC for 30 s, annealing at 55 ºC for 1min, extension at 72 ºC for 1min. The amplified CD gene (1.3 Kb) was cloned in pGEMT-Easy vector (*Promega*) and further subcloned into pGEX-4T2, a bacterial expression vector with N-terminal GST tag and designated as pGEX-CD.

Hot spots Identification. Identification of mutagenesis hot spots present in the single subunit of WT CD (426 residues) was done by Hotspot wizard[S1]. The software takes the protein PDB file (code: 1K6W) as input and gives the output in the order of mutability of functional residues (residues lining in the active site pocket) as well as the active site as hot spots.

Prediction of protein stability upon point mutation. Stability of the protein upon single amino acid substitutions in the hotspot residues was performed using CUPSAT [S2]. The software predicts the stability of the protein upon point mutations for already existing structures and custom structures as well[S2]. The stable mutants obtained were chosen for further modeling studies.

Molecular Modeling. Predicted stable mutants were modeled; taking WT bCD (1K6W) as template with MODELLER[S3], a homology modeling program which models the proteins by satisfaction of spatial restraints. Totally 5 iterations were given to create models for each mutant. The model with lowest Discrete Optimized Protein Energy (DOPE) score was considered as the top model and chosen for docking analysis.

Macromolecule preparation for Docking. The crystal structure of bCD (1K6W) was obtained from RCSB PDB and all the hydrogen atoms (polar and non-polar hydrogen atoms) were added using AutoDock[S4]. After removing the water molecules from the WT structure, the mutants were modeled and used for docking. In both cases the charge of the metal ion (Fe) was set to zero.

Ligand preparation. The structure of the ligands cytosine and 5-FC were drawn using Java Molecular Editor (JME) [S5] and transferred to PRODRG[S6], a small molecule topology generator to convert the ligands in required format for docking studies.

Docking Procedure. Docking of bCD (without the GST tag) with cytosine and 5-FC was performed using AutoDock 4.2.3[S4]. Lamarckian Genetic Algorithm (LGA) was employed to obtain the energetically favored binding modes. Grid box with the dimensions of 116 × 104 × 48 and 0.602 Å grid spacing were set to cover the entire protein. Docking parameters were fixed as 250 individual populations, 2500000 energy evaluations, and a mutation rate of 0.02 with the output level of 0. One hundred docked conformations were generated for each ligand. Similar conformations of the ligand were combined as a single cluster. Maximum root mean square tolerance of 2 Å was set. The free energy of binding score obtained from AutoDock includes the van der waals, electrostatics and desolvation energy. The results were analyzed based on known interactions between the macromolecule and the ligands as well as the free energy of binding of the ligands.

Site Directed Mutagenesis. Site directed mutagenesis was used to individually introduce the following substitutions: F186W, F82C, S126N, R91T and S126R. The mutagenesis reaction was carried out with the QuikChange site directed mutagenesis kit obtained from Stratagene (La Jolla,CA). The pGEX-CD expression vector was used as template for the mutagenesis reactions. Table S1 lists the synthetic oligonucleotides used for *in vitro* site directed mutagenesis.

List of synthetic oligonucleotide primers used for site directed mutagenesis

Mutation	Oligonucleotide Sequence
F186W	5'-GTGGGGGCGATTCCGCATTGGGAATTTACCCGTGAATACG-3'
F82C	5'-GAATCAGTCCGGCACGCTGTGTGAAGGCATTG-3'
S126N	5'-GCGTACCCATGTCGATGTTAATGATGCAACGCTAACTGCGC-3'
R91T	5'-TTGAACGCTGGGCCGAGACCAAAGCGTTATTAACCC-3'
S126R	5'-TGTGCGTACCCATGTCGATGTTAGGGATGCAACGCTA-3'

DNA sequencing. Sequencing was carried out to confirm the WT CD and the introduction of mutation sites. All the sequencing reactions were performed at Xcelris labs ltd, India. The multiple sequence alignment of the WT and the mutants are illustrated in Fig. S1.

```
S126N_A SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
S126R_A SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
F186W_A SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
R91T_A  SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
1K6W_A  SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
F82C_A  SNNALQTIINARLPGEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50
        **************************************************

S126N_A GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIERWAERKALLTHDDV 100
S126R_A GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIERWAERKALLTHDDV 100
F186W_A GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIERWAERKALLTHDDV 100
R91T_A  GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIERWAETKALLTHDDV 100
1K6W_A  GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLFEGIERWAERKALLTHDDV 100
F82C_A  GLVIPPFVEPHIHLDTTQTAGQPNWNQSGTLCEGIERWAERKALLTHDDV 100
        ****************************** ******** *********

S126N_A KQRAWQTLKWQIANGIQHVRTHVDVNDATLTALKAMLEVKQEVAPWIDLQ 150
S126R_A KQRAWQTLKWQIANGIQHVRTHVDVRDATLTALKAMLEVKQEVAPWIDLQ 150
F186W_A KQRAWQTLKWQIANGIQHVRTHVDVSDATLTALKAMLEVKQEVAPWIDLQ 150
R91T_A  KQRAWQTLKWQIANGIQHVRTHVDVSDATLTALKAMLEVKQEVAPWIDLQ 150
1K6W_A  KQRAWQTLKWQIANGIQHVRTHVDVSDATLTALKAMLEVKQEVAPWIDLQ 150
F82C_A  KQRAWQTLKWQIANGIQHVRTHVDVSDATLTALKAMLEVKQEVAPWIDLQ 150
        ************************ ************************

S126N_A IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHFEFTREYGVESLHKT 200
S126R_A IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHFEFTREYGVESLHKT 200
F186W_A IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHWEFTREYGVESLHKT 200
R91T_A  IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHFEFTREYGVESLHKT 200
1K6W_A  IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHFEFTREYGVESLHKT 200
F82C_A  IVAFPQEGILSYPNGEALLEEALRLGADVVGAIPHFEFTREYGVESLHKT 200
        ********************************** **************

S126N_A FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
S126R_A FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
F186W_A FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
R91T_A  FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
1K6W_A  FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
F82C_A  FALAQKYDRLIDVHCDEIDDEQSRFVETVAALAHHEGMGARVTASHTTAM 250
        **************************************************

S126N_A HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
S126R_A HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
F186W_A HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
R91T_A  HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
1K6W_A  HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
F82C_A  HSYNGAYTSRLFRLLKMSGINFVANPLVNIHLQGRFDTYPKRRGITRVKE 300
        **************************************************
```

```
S126N_A MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
S126R_A MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
F186W_A MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
R91T_A  MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
1K6W_A  MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
F82C_A  MLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQLMGYGQINDGL 350
        **************************************************

S126N_A NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
S126R_A NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
F186W_A NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
R91T_A  NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
1K6W_A  NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
F82C_A  NLITHHSARTLNLQDYGIAAGNSANLIILPAENGFDALRRQVPVRYSVRG 400
        **************************************************

S126N_A GKVIASTQPAQTTVYLEQPEAIDYKR 426
S126R_A GKVIASTQPAQTTVYLEQPEAIDYKR 426
F186W_A GKVIASTQPAQTTVYLEQPEAIDYKR 426
R91T_A  GKVIASTQPAQTTVYLEQPEAIDYKR 426
1K6W_A  GKVIASTQPAQTTVYLEQPEAIDYKR 426
F82C_A  GKVIASTQPAQTTVYLEQPEAIDYKR 426
        **************************
```

Fig. S1. Sequence alignment of the WT bCD (1K6W) and the mutants. F82C mutation is shown as red box. R91T mutation is shown as green box. S126N and S126R mutations are given in blue box. F186W mutation site is shown as black box. ClustalX-2.1[S7] program was used to generate the multiple sequence alignment.

Expression of WT CD and the mutants. WT CD and the mutant enzymes were expressed in the *E. coli* strain BL21 (DE3) *tdk* ˉ. A 50 mL culture of LB medium was inoculated with 3mL primary culture and induced for overnight with 0.5mM IPTG at 37 °C. The cells were harvested at 6000 rpm for 10 min at 4°C, and the cell pellet was resuspended with pre chilled 1× PBS (pH 7.4) and disrupted twice at constant cell disruptor at 20 kpsi. The disrupted sample was incubated with 1% triton X–100 for 30-40 min on ice. Then the sample was centrifuged at 15,000 rpm for 30 min at 4°C. The supernatant which contains the soluble fraction of the protein of interest was collected and passed through a 0.45µ membrane filter.

Purification of WT CD and the mutants. The WT and the mutant enzymes were purified near homogeneity by affinity column chromatography using immobilized glutathione-agarose beads. In short, 5mL of filtered lysate was added to 2 mL of beads and allowed to mix on a rocker for 30 min in chilled condition to obtain the maximum binding of protein of interest to the resin.

After draining out the flow through, the column was washed at least with ten column volumes of wash buffer (1× PBS, pH 7.4) to remove the non-specific proteins. Then the purified protein was eluted with 5 ml of elution buffer (10 mM glutathione reduced in 50 mM Tris-HCl, pH 9.5). The purified protein was dialyzed in dialysis buffer (1× PBS, pH 7.4) at 4 °C for overnight. After dialysis, the protein samples were collected and kept at -20 °C. Protein concentrations for both the WT and the mutants were determined using Bradford method using BSA as standard. Purified proteins were analyzed using SDS-PAGE, and silver staining was performed to visualize the purified fractions.

Enzymatic assay of the WT and mutant proteins with cytosine. CD assays were carried out with varying substrate concentrations. Enzyme assays were performed using the spectrophotometer based method[S8]. A main stock of 25 mM cytosine was made in 1× PBS, pH 7.4. The degradation of substrate cytosine was monitored at 286 nm and the readings were taken immediately after collecting the sample every 2 min for a total of 18 min. The procedure was repeated using the same amount of enzyme for every cytosine concentration. Lineweaver-Burk plot was used to determine K_m values for each mutant and WT CD enzyme. The catalytic efficiency (K_{cat}/K_m) was also determined for each mutant and WT CD. Assays were performed three to five times for each substrate concentration.

Enzymatic assay of the WT and mutant with 5-fluorocytosine. Kinetic values for the WT and the mutants were obtained with 5-FC substituted for cytosine as the substrate. A stock of 25 mM 5-FC was made in sterile water. The reaction mixture containing the desired concentration of 5-FC and water was placed at 37°C. The reaction was started immediately after addition of the enzyme. A small 100 µL aliquots of the reaction were collected every 2 min over a 18 min time period and immediately quenched in 0.1 M HCl. Readings were then taken at 290 and 255 nm and the concentrations of 5-FC and 5-FU (in mM) were calculated using the formula [5-FC] = $0.119(A_{290}) - 0.025(A_{255})$ and [5-FU] = $0.185(A_{255}) - 0.049(A_{290})$ [S9]. Assays were repeated three to five times. Similar to cytosine the kinetic parameters were determined using the double reciprocal plots.

Table S1. Mutability scores given for the residues over the bCD structure (1K6W) by HotSpot Wizard

Mutability Score	Residues
1	ILeu 85, His 246, His 63, His 61, Asp 246, Gln 150, Arg 293, Arg 120, Pro 184, Thr 80, Ala 153, Leu 81, Ala 204, Gly 176, Gly 323, Val 308, Gly 115, Gly 401, Ala 358, His 214, Asp 313, Asp 212, Gln 156, Glu 227, Asp 219, Thr 354, Ser 196, Pro 155, Pro 276, His 185, Asp 178, Gly 181, Asp 65, Asp 220, Gly 158, Gly 79, Gly 284, Gly 371, Asp 317, Thr 296, Phe 154, Asp 124, Arg 285, Asp 287, Glu 217, Ala 374, Ser 245, Glu 187, His 122, Ala 233, Ser 223, Ser 397, Arg 103, Ala 172, Pro 290, Arg 292, Leu 52, Asn 279, Glu 83, Gly 294, Asn 76, Gly 367, ILeu 183, Pro 318, Arg 224, Val 152, Arg 297, Ser 78, Leu 328, Leu 282, Lys 92, Ser 252, Val 298
2	Val 180, Ala 230, Gly 305, Val 242, His 336, Thr 69, Ala 249, Gly 84, Ala 104, Gln 283, Asn 326, Trp 319, Asn 275, Thr 248, Ala 135, Ala 244, Leu 261, Tyr 366, Val 394, Glu 300, Thr 96, Thr 189, Ala 387, Asp 127, Asn 307, Tyr 257, Ala 177, Leu 64, Pro 55, Met 250, ILeu 22, Gln 117, Glu 59, Gly 311, Lys 291, Ser 161, Thr 121, Gly 237, Val 179, Ser 268, ILeu 218, Ala 381, Asp 314, Val 123, Ala 256
3	Gly 27, Ala 11, Tyr 320, Val 58, Thr 67, Gln 50, ILeu 29, ILeu 377, Trp 88, Gly 193, Met 327, Pro 56, Val 315, Leu 133, ILeu 32, Leu 130, Asn 254, Leu 277, His 312, Ala 48, Pro 321, Arg 389, Ala 325, Pro 73, Thr 324, Asp 99, Phe 57, Val 278, Glu 157, Leu 340, Cys 215, Tyr 253, Thr 247, Glu 138, Ala 405, Leu 331, Tyr 207, Leu 13, His 356, Gln 339, Ala 182, ILeu 211, Leu 203, Leu 24, Thr 66, ILeu 151, Lys 140, Phe 225, His 251, Ala 132, Asn 375, Gln 21, ILeu 353, Val 53, Gly 165, Val 139, ILeu 116
4	Val 125, Phe 82, Arg 190, Thr 243, Arg 241, Asn 10, ILeu 368, Arg 260, ILeu 62, Asp 47, Gln 6, Met 341, Gly 334, Ser 126, Leu 379, Arg 91, Val 194, Thr 228, Leu 173, Leu 168, Asp 148, Val 119, Leu 169, Met 333, Lys 299, Thr 407, Gly 71, Val 273, Arg 399, ILeu 112, Leu 197, Val 226, Leu 352, Val 330, Leu 265, Phe 272, Leu 149, Glu 191, Lys 206, Val 100, Thr 107, Glu 171, Thr 288, Asn 271, Cys 309, Lys 266, Glu 221, Gly 239, His 281, Met 301, Phe 186, Gln 111, Val 213, Ala 144, Val 398, Gly 255, Ala 274, Met 136, Val 337, Asp 386, Gly 342, Leu 5, Pro 380, ILeu 8, Ser 304
5	Phe 201, ILeu 346, Arg 87, Val 229, Leu 361, Leu 322, Ala 70, Leu 46, Leu 264, ILeu 159, Val 403, Gln 391, ILeu 404, Leu 137, Met 238, Asp 98, Asp 208, Thr 200, Ser 357, Leu 363, Leu 108, ILeu 270, Thr 360, Gln 205, Arg 263, Thr 258, Val 143, Ala 128, Arg 12, Met 267, ILeu 280, Ser 259, Thr 42, ILeu 306, Ala 240

6	Pro 60, Glu 303, Glu 86, Pro 145, Phe 310, Gly 15, Arg 395, Gln 329, Glu 90, Pro 163, Gln 141, Gln 68, Lys 134, Leu 350, Tyr 289, Ala 89, Phe 316, Leu 388, Thr 131, ILeu 54, Ala 113, Ala 231, Cys 338, Leu 376, His 235, Val 392, ILeu 9, Leu 95, Asn 74, Glu 142, Leu 335, ILeu 378, Glu 170, Asn 383
7	Leu 210, Arg 174, Pro 409, Trp 20, Leu 232, His 332, Thr 7, Phe 262, Asn 362, Arg 209, Ala 31, Gly 349, Arg 359, ILeu 295, Ala 202, Ser 36, Lys 402, Asn 114, Lys 101, Leu 175, ILeu 147, His 198, Phe 286, Gln 345
8	Thr 129, Lys 109, Ala 370, Tyr 162, Phe 188, Gln 222, Asn 351, Ser 30, Ala 93, Asp 26, His 118, Leu 160, Asn 164, Phe 385, Arg 390, Lys 28, Gln 106, Tyr 192, Pro 393
9	ILeu 422, Ala 34, His 355, Val 38, Glu 16, Thr 412, Glu 195, Glu 236, Ala 167, Asn 372, His 23, Trp 110, Gly 269, Asn 44, Asp 348, Val 414, Gln 364, Gly 384, Pro 40, Gly 18, Asp 423, Ala 369, Ser 406, Leu 19, Gln 77, Tyr 343, Ser 373, Ala 410, Ala 4, Met 39, Gly 400, Pro 14, Ser 45, Tyr 424, Gln 25, Trp 105, Gln 102, Tyr 396, His 234, Gly 344, Arg 426, Glu 420, Leu 416, Gly 51, Glu 43, Leu 302, Glu 17, Asp 33, Gln 418, Trp 75, Glu 417, Gln 72, Asp 365, Glu 382, Glu 166, Lys 425, Ala 421, Leu 94, His 97, Pro 419, Gln 408, ILeu 41, Gly 37, Glu 49, Trp 146, Thr 413, Asn 347, Lys 199, Tyr 415, Gln 411, Gln 35

Table S2. Effect of the protein stability upon mutation in the hotspot residues

Residue No.	Residue	Overall stability	Substituted amino acid
162	Tyr	Stabilizing	-
		Destabilizing	Ala, Val, Leu, ILeu, Met, Thr, Cys, Asp, Glu, Asn, Gln, Ser, Gly, Arg, His, Trp, Phe, Pro, Lys
129	Thr	Stabilizing	Ser, Met, Lys, Cys, Asp
		Destabilizing	Ala, Val, Leu, ILeu, Glu, Asn, Gln, Arg, His, Pro, Trp, Tyr, Phe, Gly
286	Phe	Stabilizing	Gly, Ala, Val, Pro, Ser, Thr, Gln, Lys, Asn, Cys, Glu, Asp, Arg, His
		Destabilizing	Met, Leu, ILeu, Trp, Tyr
95	Leu	Stabilizing	Trp, Ser

		Destabilizing	Ala, Val, ILeu, Glu, Asn, Gln, Arg, His, Pro, Tyr, Phe, Gly, Met, Lys, Cys, Asp,Thr
131	Thr	Stabilizing	Ala, Leu, ILeu, Pro, Trp, Phe, Glu
		Destabilizing	Val, Lys, Asp, Arg, His, Asn, Gly, Cys, Met, Ser, Tyr, Gln
186	Phe	Stabilizing	Met, Pro, Trp, Ser, Thr, Lys, Asn, Cys, Asp, His
		Destabilizing	Ala, Val, Gly, Leu, ILeu, Arg, Tyr, Glu, Gln
91	Arg	Stabilizing	Ala, Met,Thr, Phe, Lys, Glu, Asp
		Destabilizing	Val, Leu, ILeu, Gly, Gln, Asn, Ser, Pro, His, Trp, Tyr, Cys
126	Ser	Stabilizing	Met, Pro, Gln, Lys, Asn, Glu, Asp, Arg, His
		Destabilizing	Ala, Val, Gly, Leu, ILeu, Phe, Trp, Tyr, Cys, Thr
82	Phe	Stabilizing	Ala, Val, Leu, Pro, Trp, Gln, Lys, Tyr, Cys, Glu, Asp, Arg, His
		Destabilizing	ILeu, Gly, Met, Ser, Asn, Thr

Table S3. Binding energy scores and the number of clusters obtained from AutoDock for the WT and the mutants docked with cytosine and 5-FC

Mutant	**Binding energy (kcal/mol)**		**No. of Clusters**	
	With cytosine	With 5-FC	With cytosine	With 5-FC
F82A	-4	-3.92	1	3

F82C	-4.2	-4.21	11	13
F82D	-4.11	-3.91	2	2
F82E	-4.02	-3.81	6	6
F82H	-4.24	-4.03	4	1
F82K	-3.82	-3.8	1	3
F82L	-4.14	-3.98	24	10
F82P	-4.32	-3.79	6	2
F82Q	-3.96	-3.91	21	5
F82R	-4.34	-3.98	7	5
F82V	-4.58	-4.15	15	12
F82W	-4.14	-4.03	7	4
F82Y	-3.81	-3.37	2	1
R91A	-4.08	-3.98	9	3
R91D	-4.47	-3.98	10	9
R91E	-3.85	-3.87	2	7
R91F	-4.36	-3.98	19	11
R91K	-4.32	-4.11	18	6
R91M	-4.46	-4.18	16	12
R91T	-3.92	-4	1	3
L95S	-4.26	-3.99	1	4
L95W	-	-	-	-
S126D	-3.99	-3.8	3	7
S126E	-4.41	-4.07	14	15

S126H	-4.28	-4.23	16	6
S126K	-4.35	-4.11	6	1
S126M	-4.43	-4.04	8	6
S126N	-4.17	-4.19	14	1
S126P	-4.05	-3.98	6	8
S126Q	-4.49	-4.24	11	8
S126R	-	-3.6	-	1
T129C	-4.3	-4.26	24	6
T129D	-4.07	-3.89	12	1
T129K	-3.87	-3.48	3	2
T129M	-3.64	-3.62	2	3
T129S	-4.05	-3.87	15	13
T131A	-4.29	-4.1	1	8
T131E	-4.35	-3.97	8	13
T131F	-4.21	-4.08	2	8
T131I	-3.74	-3.63	1	4
T131L	-4.09	-3.72	2	7
T131P	-4.41	-4.08	9	4
T131W	-3.78	-3.62	2	3
F186C	-4.46	-4.09	12	5
F186D	-4.34	-4.05	7	12
F186H	-4.53	-4.3	21	17
F186K	-4.16	-3.84	13	10

F186M	-4.26	-3.96	13	4
F186N	-4.38	-3.7	3	2
F186P	-4.47	-4.23	16	13
F186R	-4.33	-3.91	12	5
F186S	-4.28	-4.11	15	8
F186T	-4.47	-4.02	8	5
F186W	-4.18	-4.25	6	6
F286A	-4.44	-4.19	4	5
F286C	-4.44	-4.1	18	6
F286D	-4.46	-4.24	17	6
F286E	-4.07	-4.02	6	4
F286G	-4.4	-4.1	8	7
F286H	-4.54	-4.49	37	4
F286K	-4.09	-4.04	9	11
F286N	-4.15	-4.07	9	7
F286P	-3.95	-3.73	14	6
F286Q	-3.99	-3.81	13	8
F286R	-4.11	-3.87	2	3
F286S	-4.67	-4.42	13	10
F286T	-4.2	-3.88	7	8
F286V	-4.52	-4.06	9	7
F310A	-4.34	-4.03	8	4
G311A	-4.11	-3.4	1	2

H312A	-4.06	-4.13	6	8
D313A	-3.78	-3.83	8	11
D314A	-3.65	-3.55	3	1
D314G	-3.67	-3.66	3	6
D314H	-	-	-	-
D314K	-	-	-	-
D314R	-	-	-	-
D314S	-	-3.81	-	3
V315A	-3.95	-3.79	4	2
F316A	-4.36	-4.13	16	5
D317A	-4.49	-4.22	5	5
P318A	-4.01	-3.9	19	4
W319A	-4.13	-3.45	4	1
Y320A	-4.02	-4.14	3	2
Q102R	-4.48	-4.06	13	12
WT	-4.02	-4.03	19	18

Docking of different mutants with cytosine. The ligand cytosine was docked with 68 mutants using the docking parameters mentioned above, and the resulting binding energy and the numbers of clusters are given in Table S3. The ligand orientations, positioning of catalytic residues and hydrogen bond interactions between the ligand and the protein are shown in Fig. S2a-f.

Among the evaluated 68 mutants, 13 mutants displayed increased binding energy towards cytosine than the WT (Table S3). Out of these 13 mutants, T129M showed the highest binding

energy of -3.64 kcal/mol. 52 mutants showed decreased docking scores towards cytosine than the WT (Table S3). Amongst them, mutant F286S displayed the least binding free energy of -4.67 kcal/mol.

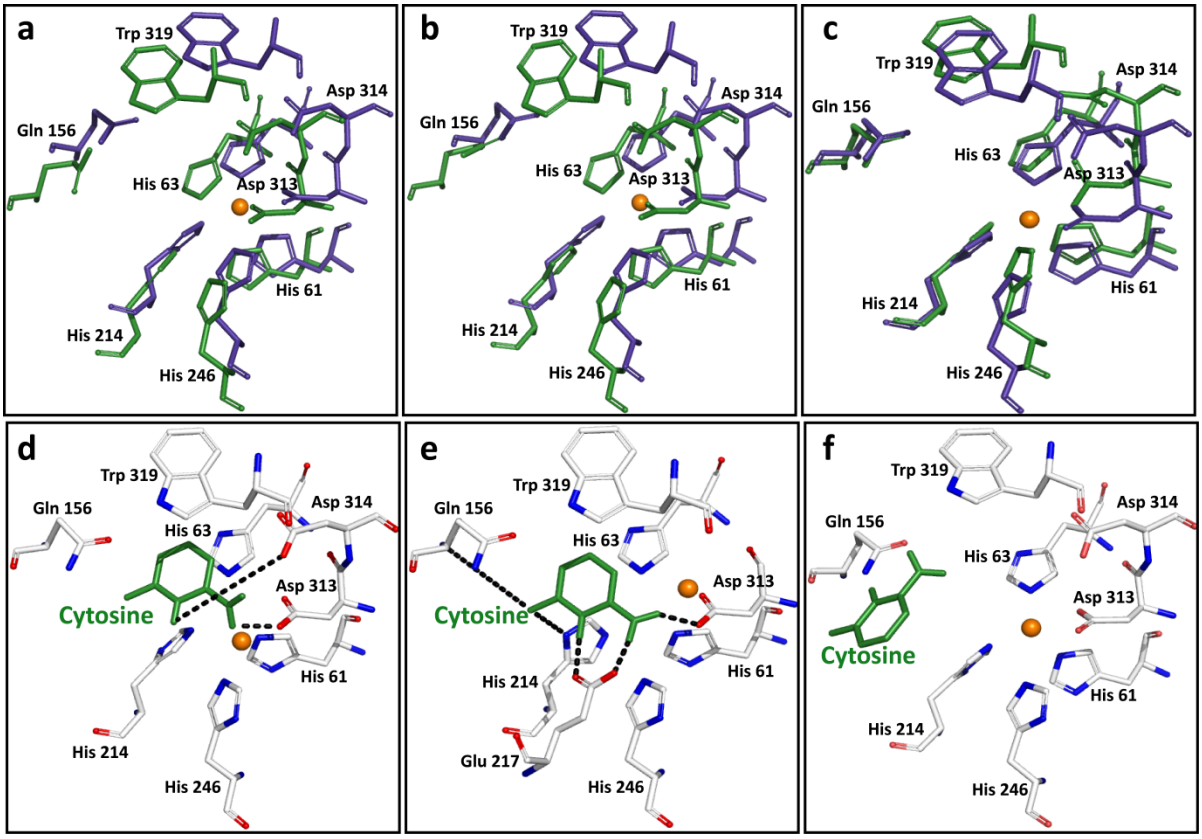

Fig. S2. (Top panel) Superimposed images of **a)** WT and mutant L95W, **b)** WT and mutant S126R. Since the metal ion and all the surrounding active site residues (Trp319, Asp 313, Asp 314, Gln 156 and His 63) should be in highly ordered manner to maintain the structural integrity of the enzyme[S10], the alteration of structure of Trp319 in the L95W and S126R mutants could be responsible for inactivating the enzymes due to the incapability of the ligand to bind to the active site of the mutants, **c)** The catalytic residues of WT and the mutant (F82E) are nearly merging with each other in their three dimensional orientation. The minor alterations observed might not affect the mutant F82E to show similar catalytic activity relative to the WT towards cytosine which was also reflected from the same binding energy (-4.02 kcal/mol) ascribed to the mutant as that of WT. In all the figures, WT and mutant residues are shown as blue and green sticks respectively. Metal ion is shown as orange sphere. (Bottom panel), **d)** Mutant T129M. Absence

of any hydrogen bond interaction with residues participating in substrate recognition and catalytic activity of the enzyme (Gln 156, Glu 217) might be the cause for diminished activity of the T129M mutant enzyme towards cytosine, **e)** Mutant F286S. The ability of the mutant F286S to form hydrogen bonding with Gln 156 residue helps in the proper orientation of ligand in the active site and the Glu 217 and Asp 313 residues facilitate the electrostatic interactions with the metal ion. As a consequence, mutant F286S could show increased sensitivity towards cytosine, **f)** Ligand orientation in case of WT. Cytosine is indicated as green sticks. Hydrogen bonds are shown as black dashed lines and metal ion as orange sphere. All parts of this Fig. were generated with PyMOL[S11].

Docking of different mutants with 5-FC. The prodrug, 5-FC was also docked with 68 mutants using the same docking parameters and the resulting binding energy and the number of clusters are listed in Table S3. The ligand orientations, the positioning of catalytic residues and the hydrogen bond interactions between the ligand and the protein are shown in Fig. S3a-e. The substrate binding residues of 5-FC is similar to those of cytosine. Parallel to cytosine, the ligand 5-FC also failed to bind to the active site of the mutant L95W due to structural destabilization of the mutant (Fig. S2a). Based on the binding free energy score, 36 mutants displayed lower affinity towards 5-FC as compared to WT (Table S3). Mutant F82Y was ascribed with the highest binding energy of -3.37 kcal/mol with 5-FC. Twenty-nine of the 68 mutants attained lesser scores with 5-FC as compared with WT (Table S3). Mutant F286H obtained the least binding energy of -4.49 kcal/mol when docked with 5-FC.

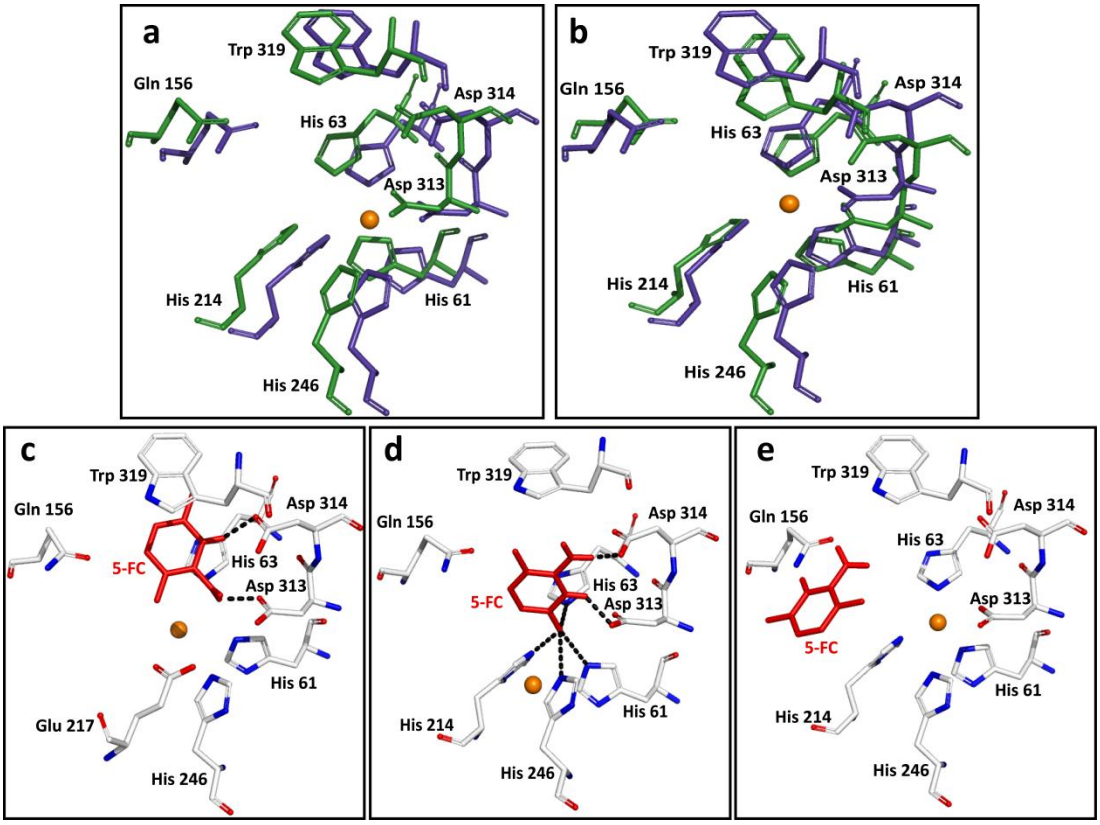

Fig. S3. (Top panel) Superimposed images of **a)** WT and the mutant F82H, **b)** WT and the mutant F82W. F82H and F82W mutants showed same binding energy of -4.03 kcal/mol in consistent with the WT, as a consequence of which they are predicted to show similar activity towards 5-FC as that of the WT protein. Absence of any significant alterations in the active site residues in the superimposed images of the mutant and the WT further supporting this observation. WT and mutant residues are shown as blue and green sticks respectively. Metal ion is shown as orange sphere. (Bottom panel) **c)** In the mutant F82Y, there is a possibility of steric clash between the critical substrate binding residue Trp319 and the ligand when compared with the WT (Fig. S3e) which may reduce the activity of the protein towards 5-FC, **d)** In case of the mutant F286H, the ligand's orientation can successfully avoid the steric constraints with Trp 319 residue to that of WT (Fig. S3e). 5-FC is indicated as red sticks. Hydrogen bonds are shown as black dashed lines and metal ion as orange sphere. Both panels of this Fig. were generated with PyMOL[S11].

References:

S1. A. Pavelka, E. Chovancova, J. Damborsky, *Nucleic Acids Res*., 2009, **37**, W376-W383.

S2. V. Parthiban, M. M. Gromiha, D. Schomburg, *Nucleic Acids Res*., 2006, **34**, W239-242.

S3. N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, *Curr*. *Protoc*. *Bioinf*., 2006, **15**, 5.6.1-5.6.30.

S4. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart , R. K. Belew and A. J. Olson, *J*. *Comput*. *Chem*., 1998, **19,** 1639-1662.

S5. P. Ertl, Java Molecular Editor. http://davapc1.bioch.dundee.ac.uk/prodrg/jme_window.html

S6. A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallogr D*., 2004, **D60,** 1355-1363.

S7. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M . Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*., 2007, **23**, 2947-2948.

S8. P. L. Ipata and G. Cercignani, *Methods Enzymol*., 1978, **51**, 395-400.

S9. M. S. Hayden, P. S. Linsley, A. R. Wallace, H. Marquardt and D. E. Kerr, *Protein Expres. Purif*., 1998, **12**, 173–184.

S10. G. C. Ireton, G. McDermott, M. E. Black, B. L. Stoddard, *J*. *Mol*. *Biol*., 2002, **315**, 687-697.

S11. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC