# Supporting Information

Murphy et al. 10.1073/pnas.0811070106

## SI Text

**Pharmacology of Human Prodrug-Activating Enzymes.** A designed cytosine deaminase with a sequence close to that of a human protein might retain low immunogenicity while providing prodrug-activating ability. Cellular expression of cytosine deaminase activity causes susceptibility to the toxic effects of the prodrug 5-fluoro-cytosine (5FC), which, after deamination to 5-fluoro-uracil, is metabolized to the cytotoxin 5-fluoro-dUMP (1). Because the human genome does not encode a free cytosine deaminase (2), 5FC is used clinically for treatment of infections by organisms that do express cytosine deaminase, such as fungi (3). 5FC has also been used in suicide gene chemotherapy of human tumors into which microbial cytosine deaminase has been introduced by means of either antibody targeting or transfection (4). However, prodrug-activating enzymes derived from microbes have high immunogenicity and hence are inappropriate for some suicide gene therapy regimens when an initial phase of viability is required, such as preemptive suicide gene therapy for graft vs. host disease (5–7). We consider human guanine deaminase (hGDA) to be the best candidate enzyme for specificity alteration to cytosine and 5FC by means of our de novo loop design algorithm. hGDA is the only free nucleobase deaminase in the human genome (8, 9). It shares significant structural homology with bacterial cytosine deaminase (bCD) (10, 11), although not enough to allow for direct loop grafting. Other aspects of cytosine and guanine that differ remain to be investigated.

**Materials and Methods.** ***Computational Method.*** Our general method for enzyme specificity alteration comprises 4 stages. First, a model of the desired substrate bound to the enzyme in the transition state configuration is created. Knowledge of the reaction and structure of the enzyme is incorporated at this stage, although subsequent stages are independent of the precise method used to create the transition state model.

Second, protein side-chain functional groups important for specific interactions between the enzyme and the desired substrate are optimally positioned in the model. These ideal positions can be derived from crystal structures containing the desired substrate or from energetic considerations. When the side chain of a given residue is assigned to an ideal position, the backbone is also implicitly assigned to a set of possible positions, depending on the rotameric configuration of the side chain, analagous to the "inverse rotamers" described in a previous study (12). In general, it is possible that no residue in a given scaffold protein will have its backbone atoms located at any of these positions.

Third, we attempt to remodel the polypeptide backbone of the scaffold protein in such a way that it is able to host the newly introduced side chain in its ideal position. A segment of the scaffold is chosen for modification, and backbone discontinuities are introduced at the beginning and the end of that segment. The position of the segment is transformed so that the backbone atoms of a randomly chosen residue in the segment are relocated to one of the implicitly assigned positions determined by the desired side-chain interaction. A small random number of residues are added to or deleted from the segment at this stage, because the optimal loop length is unknown but not necessarily equal to that of the scaffold. At this point, the backbone of the model will be discontinuous both at the beginning and end of the segment. To reform a closed chain, several general methods from protein structure prediction are applied. Fragment insertion (13), in which backbone segments of unrelated protein crystal structures are randomly combined, is used to generate alternative configurations that differ substantially from the original structure. These configurations are accepted or rejected according to the Metropolis criterion (14), using a low-resolution energy function reflecting chain closure, atomic overlap, and backbone torsional angle preference. Gradient-based minimization (15, 16) and cyclic coordinate-descent (17) are used to optimize chain closure beyond what can be achieved with crystallographically observed torsional angles. A kinematic framework that represents both intramolecular torsional angles as well as intermolecular rigid-body transformations was used to generate atomic coordinates for energy evaluation (18–20). Such a framework allows us to explicitly specify the rigid-body transformation desired between substrate and side chain and thus focus sampling of backbone configurations exclusively in that conformational subspace (Fig. S7).

Fourth, standard computational design protocols, in conjunction with a high-resolution energy function (21), are used to identify sequences that maximally stabilize the generated backbone structures and thus optimize transition state binding. De novo loop-modeling methods (22) are used to corroborate that designed sequences computationally fold to the desired structure when the constraint between substrate and side chain is removed. Some sequences are observed to fold to configurations that do not satisfy the substrate–side-chain constraint and are thus not further evaluated.

***Modeling of hGDA with cytosine and ammelide.*** A model of a hypothetical cytosine deaminase was created, using the structure of hGDA as the enzyme scaffold and using the structure of bCD as a source of ideal interactions between enzyme and substrate. In this model, cytosine was superimposed on the 6-membered ring of guanine in the orientation corresponding to the transition state for deamination. The interaction between cytosine N1/O2 and glutamine 156 observed in bCD was used to position the side chain of a new amide-containing residue (either asparagine or glutamine). Alternate backbone configurations of hGDA residues 211–220 were sampled, including configurations containing insertions and deletions of residues in that loop. The optimal loop identified by this protocol involved sequence changes only to residues 213–218.

Mutations of residues other than those in the loop were also predicted by our design protocol. The mutation M171S was predicted to stabilize the conformation of the 213–216 loop, and thus improve transition state binding. However, experimental characterization showed that this mutation did not significantly affect activity. Loops that were longer than the wild type by 2 residues were also predicted to adopt conformations compatible with substrate binding and catalysis. However, experimental characterization showed that mutants with this longer loop length were not active.

Cytosine differs from guanine not only in the N1/O2 moiety, but also the C5/C6 moiety of the 6-membered ring. Interactions with that moiety could not be made on the basis of changes to hGDA residues 213–218. Results of the loop modeling protocol targeting N1/O2 of cytosine or ammelide did not depend on other aspects of the substrate. Therefore, for subsequent experimental characterization, we assayed activity toward ammelide, which is similar to the N1/C6/O6 moiety of guanine in that region.

***Cloning.*** cDNA for hGDA (NM004293) was obtained from Origene Technologies (TC111122; Origene Technologies). Oligonucleotides were obtained from Integrated DNA Technologies (IDT). The wild-type gene was amplified with primers pET29-

hGDA-fwd and pET29-hGDA-rev and ligated into pET29b+ (Novagen) between the NdeI and XhoI restriction sites, giving construct pET29b-hGDA-wt.

Overlap assembly PCR (23) using primers des-fwd and des-rev was used to introduce mutations for pET29b-hGDA-des. Point mutants of pET29b-hGDA-wt and pET29b-hGDA-des were made by means of Kunkel mutagenesis (24, 25) using primers shown in Table S2.

hGDA-des was also ligated into pET15b between the NdeI and BamHI restriction sites after amplification by primers pET15-hGDA-fwd and pET15-hGDA-rev, to allow for proteolytic removal of the hexahistidine affinity tag and further purification.

***Expression and purification.*** Proteins were expressed in an autoinduction media (26) at 18 °C for 24 h, after growth to log phase at 37 °C for 8 h. Cells were pelleted and lysed by using the BugBuster Protein Extraction Reagent (Novagen) or by sonication. Proteins were purified by using 2 mL of NiNTA His*Bind Resin (Qiagen) and dialyzed at 4 °C against 100 volumes of PBS (pH 7.5). Relative rates of activity were determined by using protein expressed from pET29b+ after this stage of purification.

Thrombin cleavage was also performed on expressed pET15-hGDA-des to remove the hexahistidine tag in a stage of further purification before determination of Michaelis–Menten parameters. One-half unit of biotinylated thrombin (Novagen) per milligram of purified protein was added during dialysis. The digest was allowed to proceed for 24 h, at which point, 1 mL of Ni NTA His*Bind Resin and 200 $\mu$L of Streptavidin Agarose (Novagen) were added, as well as imidazole to a final concentration of 20 mM. Flow through containing purified protein was collected after 1 h.

Purified protein was concentrated to ≈5 mg/mL by using Amicron Ulta Centrifugal Filter Devices, 10,000 MWCO (Millipore). Protein concentration was determined by using a Nanodrop-1000 spectrophotometer (Thermo Fisher Scientific) to measure the absorbance at 280 nm. A molar extinction coefficient of 40840 $M^{-1}$ $cm^{-1}$ was calculated by using the amino acid sequence and ProtParam online tools (27).

***Ammelide deamination assay.*** Ammelide and cyanuric acid were obtained from Sigma–Aldrich. Relative rates of activity for the designed, wild type, and other control variants were assayed in triplicate at concentrations of 20 $\mu$M enzyme and 500 $\mu$M substrate in PBS (pH 7.5) at 25 °C. Reactions were stopped at 6, 24, and 48 h by addition of acetonitrile to a final concentration of 78% (vol/vol). Michaelis–Menten parameters were determined by assaying in triplicate at 10 $\mu$M enzyme and 50, 75, 100, 125, 250, 500, and 750 $\mu$M substrate for 12 and 18 h, in PBS (pH 7.5) at 25 °C.

Product formation was determined by HPLC, using a protocol derived from ref. 28. An Agilent 1200 Series HPLC (Agilent) was used with a Zorbax NH2 Analytical Column, 4.6 × 250 mm, 5 $\mu$m (Agilent). Isocratic elution with 22% 5 mM sodium phosphate (pH 6.0) and 78% acetonitrile was performed at 1 mL/min, and absorption at 210 nm was monitored. Ammelide eluted between 12.7 and 13.0 min, whereas cyanuric acid eluted between 5.7 and 6.0 min.

Kinetic parameters were determined by fitting the Michaelis–Menten equation with nonlinear least-squares regression in R (29).

***Crystallization.*** pET15b-hGDA-des was expressed as described above. Cell pellets were lysed by sonication in 25 mM Hepes (pH 7.5), 200 mM sodium chloride, 5% glycerol, 10 mM imidazole, with 1 mM TCEP, 30 $\mu$M ammelide, 20 $\mu$M zinc acetate, and 1 mM PMSF. After centrifugation and filtration as described above, lysate was flowed across Talon resin (Clontech) and subsequently washed and eluted. Wash and elution buffers were identical to sonication buffer, but had 20 and 300 mM imidazole, respectively. Eluted protein was dialyzed into 25 mM Hepes (pH 7.5), 200 mM sodium chloride, 5% glycerol, with 1 mM TCEP, and 100 $\mu$M ammelide. Overnight digestion with biotinylated thrombin was performed during dialysis, and streptavidin capture was performed as described above. Size-exclusion chromatography using a Superdex 200 26/60 column (GE Healthcare) was performed with buffer identical to dialysis buffer except for having 300 mM sodium chloride. Purified protein was concentrated to 10 mg/mL, and during this process, buffer conditions were adjusted to 25 mM Hepes (pH 7.5), 100 mM sodium chloride, 2% glycerol, 1 mM TCEP, 200 $\mu$M ammelide. Protein was crystallized by vapor-phase equilibration in the hanging-drop geometry against a mother liquor of 100 mM Mes (pH 6.5), 200 mM sodium chloride, and 1.8 M ammonium sulfate. Crystals appeared in 1–4 days and were flash cooled in 2 stages with mother liquor plus 12% glycerol and subsequently 22% glycerol.

***Crystallography.*** Crystals belonged to the space group I212121, with dimensions $a$ = 86.83 $b$ = 107.31 $c$ = 233.62. X-ray diffraction data were collected to 2.4-Å resolution at beamline 5.0.1 at the ALS (Advanced Light Source, Lawrence Berkeley Laboratory, Berkeley, CA) at wavelength 0.9774. Data were processed and scaled by using HKL-2000 (30). The structure was built by using the CCP4 suite of programs (31, 32). Phases were determined by molecular replacement using a truncated model of wild-type hGDA derived from 2UZ9, by using PHASER. Model building and refinement were performed with COOT and REFMAC, respectively, excluding a random 5% of the data for cross-validation. Final statistics are provided in Table S4. Coordinates were deposited into the RCSB Protein Data Bank (accession no. 3E0L).

***Calculation of RMSDs.*** RMSDs reported in the text were calculated by superimposing static, nondesigned portions of the structures (that is, excluding residues 90–109, 213–216, and 412–422 in the calculation of the transformation for superimposition). Distances between C$\alpha$ atoms for the specified regions were computed after this superimposition was applied.

**DNA Sequence of hGDA.**

```
atgtgtgccgctcagatgccgcccctggcgcacatcttccgagg
gacgttcgtccactccacctggacctgccccatggaggtgctgcgg
gatcacctcctcggcgtgagcgacagcggcaaaatagtgtttttag
aagaagcatctcaacaggaaaaactggccaaagaatggtgcttcaa
gccgtgtgaaataagagaactgagccaccatgagttcttcatgcct
gggctggttgatacacacatccatgcctctcagtattcctttgctg
gaagtagcatagacctgccactcttggagtggctgaccaagtacac
atttcctgcagaacacagattccagaacatcgactttgcagaagaa
gtatataccagagttgtcaggagaacactaaagaatggaacaacca
cagcttgttactttgcaacaattcacactgactcatctctgctcct
tgccgacattacagataaatttggacagcgggcatttgtgggcaaa
gtttgcatggatttgaatgacacttttccagaatacaaggagacca
ctgaggaatcgatcaaggaaactgagagatttgtgtcagaaatgct
ccaaaagaactattctagagtgaagcccatagtgacaccacgtttt
tccctctcctgctctgagactttgatgggtgaactgggcaacattg
ctaaaacccgtgatttgcacattcagagccatataagtgaaaatcg
tgatgaagttgaagctgtgaaaaacttataccccagttataaaaac
tacacatctgtgtatgataaaaacaatcttttgacaaataagacag
tgatggcacacggctgctacctctctgcagaagaactgaacgtatt
ccatgaacgaggagcatccatcgcacactgtcccaattctaattta
tcgctcagcagtggatttctaaatgtgctagaagtcctgaaacatg
aagtcaagatagggctgggtacagacgtggctggtggctattcata
ttccatgcttgatgcaatcagaagagcagtgatggtttccaatatc
cttttaattaataaggtaaatgagaaaagcctcaccctcaaagaag
tcttcagactagctactcttggaggaagccaagccctggggctgga
tggtgagattggaaactttgaagtgggcaaggaatttgatgccatc
ctgatcaaccccaaagcatccgactctcccattgacctgttttatg
gggacttttttggtgatatttctgaggctgttatccagaagttcct
ctatctaggagatgatcgaaatattgaagaggtttatgtgggcgga
aagcaggtggttccgttttccagctcagtgtaa
```

**Amino Acid Sequence of hGDA.**

MCAAQMPPLAHIFRGTFVHSTWTCPMEVLRDHLLGVSDSGKIVF
LEEASQQEKLAKEWCFKPCEIRELSHHEFFMPGLVDTHIHASQYSF
AGSSIDLPLLEWLTKYTFPAEHRFQNIDFAEEVYTRVVRRTLKNGT
TTACYFATIHTDSSLLLADITDKFGQRAFVGKVCMDLNDTFPEYKE
TTEESIKETERFVSEMLQKNYSRVKPIVTPRFSLSCSETLMGELGN
IAKTRDLHIQSHISENRDEVEAVKNLYPSYKNYTSVYDKNNLLTNK
TVMAHGCYLSAEELNVFHERGASIAHCPNSNLSLSSGFLNVLEVLK
HEVKIGLGTDVAGGYSYSMLDAIRRAVMVSNILLINKVNEKSLTLK
EVFRLATLGGSQALGLDGEIGNFEVGKEFDAILINPKASDSPIDLF
YGDFFGDISEAVIQKFLYLGDDRNIEEVYVGGKQVVPFSSSV

**Amino Acid Sequence of hGDA-des.**

MCAAQMPPLAHIFRGTFVHSTWTCPMEVLRDHLLGVSDSGKIVF
LEEASQQEKLAKEWCFKPCEIRELSHHEFFMPGLVDTHIHASQYSF
AGSSIDLPLLEWLTKYTFPAEHRFQNIDFAEEVYTRVVRRTLKNGT
TTACYFATIHTDSSLLLADITDKFGQRAFVGKVCMDLNDTFPEYKE
TTEESIKETERFVSEMLQKNYSRVKPIVTPGNGVSETLMGELGNIA
KTRDLHIQSHISENRDEVEAVKNLYPSYKNYTSVYDKNNLLTNKTV
MAHGCYLSAEELNVFHERGASIAHCPNSNLSLSSGFLNVLEVLKHE
VKIGLGTDVAGGYSYSMLDAIRRAVMVSNILLINKVNEKSLTLKEV
FRLATLGGSQALGLDGEIGNFEVGKEFDAILINPKASDSPIDLFYG
DFFGDISEAVIQKFLYLGDDRNIEEVYVGGKQVVPFSSSV


1. Longley DB, et al. (2003) 5-fluorouracil: Mechanisms of action and clinical strategies. Nat Rev Cancer 3(5):330–338.
2. Loffler M, et al. (2005) Pyrimidine pathways in health and disease. Trends Mol Med 11(9):430–437.
3. Vermes A, et al. (2000) Flucytosine: A review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions. J Antimicrob Chemother 46(2):171–179.
4. Rooseboom M, et al. (2004) Enzyme-catalyzed activation of anticancer prodrugs. Pharmacol Rev 56(1):53–102.
5. Bordignon C, et al. (1995) Transfer of the HSV-tk gene into donor peripheral blood lymphocytes for in vivo modulation of donor anti-tumor immunity after allogeneic bone marrow transplantation. Hum Gene Ther 6(6):813–819.
6. Verzeletti S, et al. (1998) Herpes simplex virus thymidine kinase gene transfer for controlled graft-versus-host disease and graft-versus-leukemia: Clinical follow-up and improved new vectors. Hum Gene Ther 9(15):2243–2251.
7. Sprangers B, et al. (2007) Can graft-versus-leukemia reactivity be dissociated from graft-versus-host disease?Front Biosci 12:4568–4594.
8. Zielke CL, Suelter CH (1971) Purine, purine nucleoside, and purine nucleotide aminohydrolases. The Enzymes, ed Boyer PD (Academic, London) 3rd Ed.
9. Barthelmes J, et al. (2007) BRENDA, AMENDA and FRENDA: The enzyme information system in 2007. Nucleic Acids Res 35 (Database issue):D511–D544.
10. Ireton GC, McDermott G, et al. (2002) The structure of Escherichia coli cytosine deaminase. J Mol Biol 315(4):687–697.
11. Moche M, et al. (2007) Human Guanine Deaminase (Guad) in Complex with Zinc and its Product Xanthine. 10.2210/pdb2uz9/2uz9.
12. Zanghellini A, et al. (2006) New algorithms and an in silico benchmark for computational enzyme design. Protein Sci 15(12):2785–2794.
13. Simons KT, Kooperberg C, et al. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268(1):209–225.
14. Metropolis N, et al. (1953) Equation of state calculation by fast computing machines. J Chem Phys 21(6):1087–1092.
15. Abe H, Braun W, et al. (1984) Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. Comput Chem 8(4):239–247.
16. Wedemeyer WJ, Baker D (2003) Efficient minimization of angle-dependent potentials for polypeptides in internal coordinates. Proteins 53(2):262–272.
17. Canutescu AA, Dunbrack RL, Jr (2003) Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein Sci 12(5):963–972.
18. Mazur AK, Abagyan RA (1989) New methodology for computer-aided modelling of biomolecular structure and dynamics. 1. Non-cyclic structures. J Biomol Struct Dyn 6(4):815–832.
19. Bradley P, Baker D (2006) Improved beta-protein structure prediction by multilevel optimization of nonlocal strand pairings and local backbone conformation. Proteins 65(4):922–929.
20. Das R, Baker D (2008) Macromolecular modeling with Rosetta. Annu Rev Biochem 77:363–382.
21. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA 97(19):10383–10388.
22. Rohl CA, et al. ((2004) Modeling structurally variable regions in homologous proteins with Rosetta. Proteins 55(3):656–677.
23. Sambrook J, Russell DW (2001) Molecular Cloning. A Laboratory Manual (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), 3rd Ed, pp 13.36–13.39.
24. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 82(2):488–492.
25. Kunkel TA, et al. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. Methods Enzymol 204:125–139.
26. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41(1):207–234.
27. Gasteiger E, et al. (2005) Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook, ed Walker JM (Humana Press, Totowa, NJ), pp 571–607.
28. Sugita T, et al. (1990) Determination of melamine and three hydrolytic products by liquid chromatography. Bull Environ Contam Toxicol 44(4):567–571.
29. Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. J Comput Graph Stat 5(3):299–314.
30. Otwinowski Z, eds (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology (Academic, New York).
31. Collaborative Computational Project, Number 4. (1994) The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D 50(Pt 5):760–763.
32. Potterton E, et al. (2003) A graphical user interface to the CCP4 program suite. Acta Crystallogr D 59(Pt 7):1131–1137.

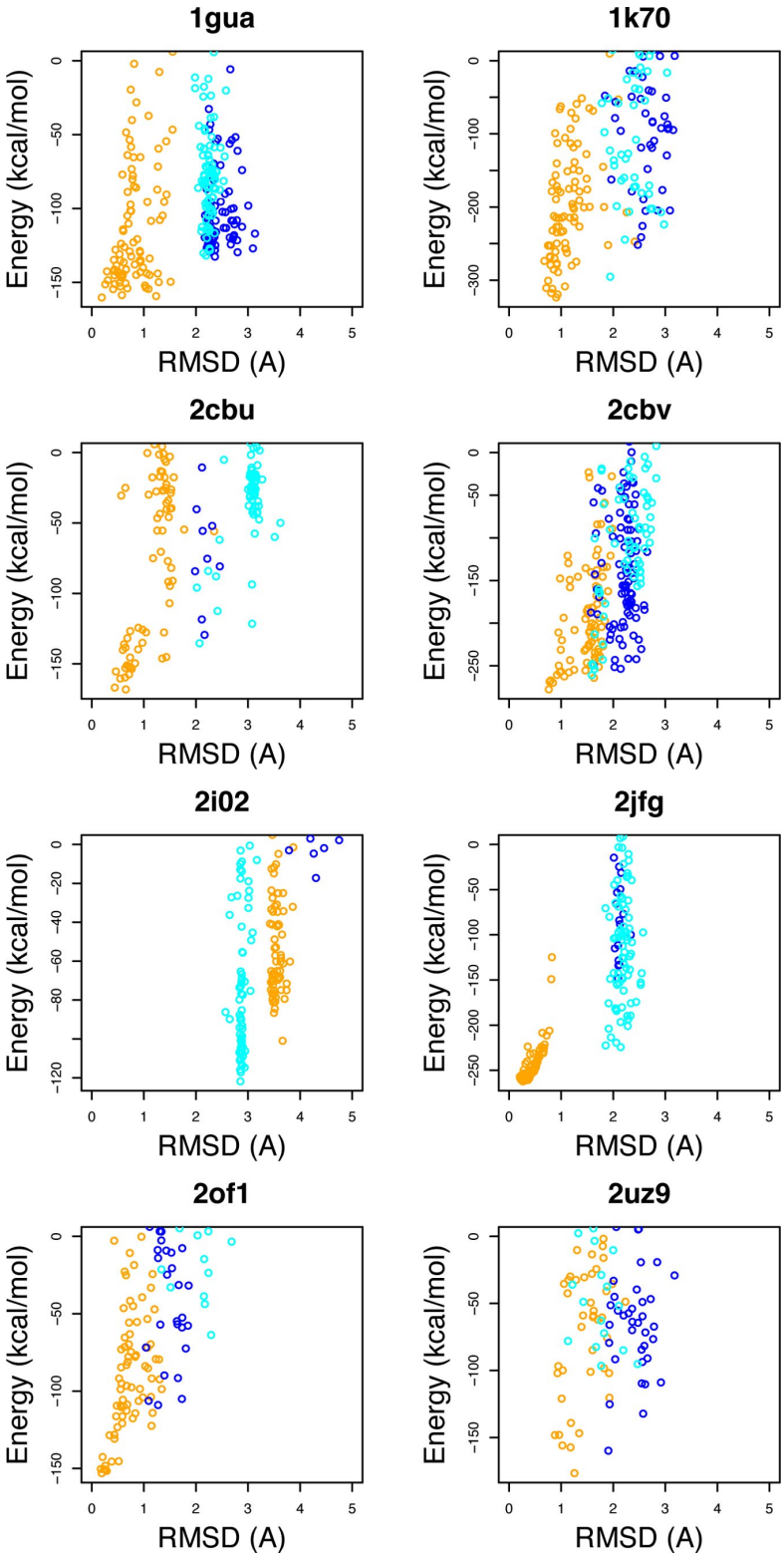


**Fig. S1.** Benchmark energy vs. RMSD plots. The score and RMD to the native structure are plotted for each of 200 models generated during benchmarking for loops of the native length. Note that this includes loops of the correct anchor position (3 residues before, 3 residues after, orange), as well as loops of incorrect anchor position that still have the native loop length (2, 4, cyan; 4, 2, blue). RMSD to the native structure is not defined for loops of nonnative lengths and thus is not plotted. In most cases, subangstrom structures of the correct loop length can be identified energetically. In the case of 2IO2, inadequate sampling contributed to the failure to identify the correct loop length, because no structure was generated with a score comparable with that of the crystal structure conformation (−235.38 kcal/mol).

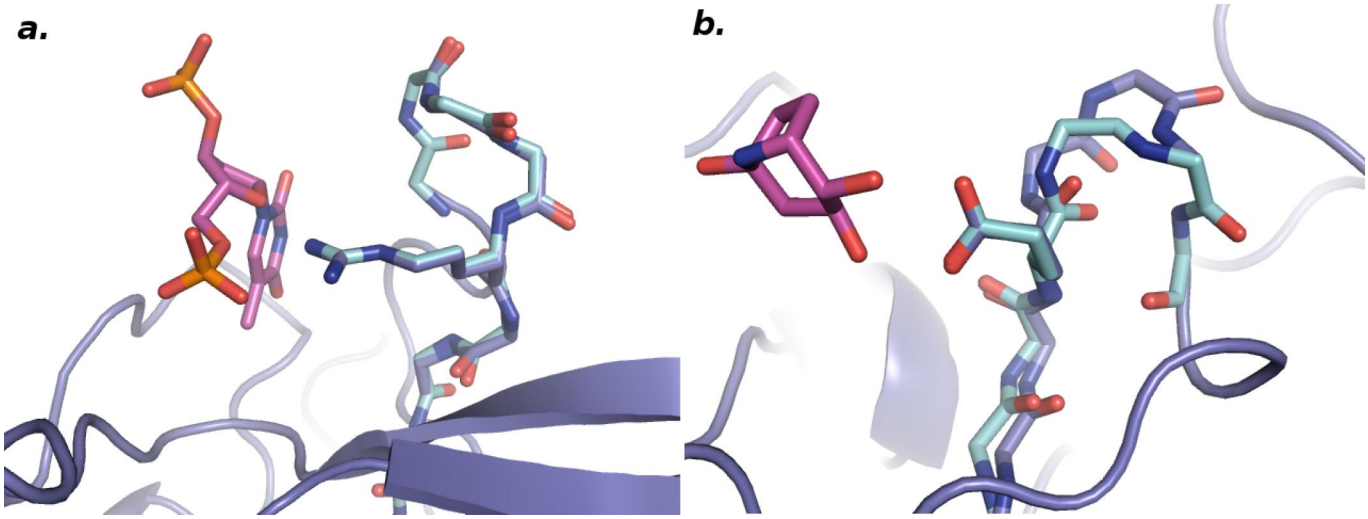


**Fig. S2.** Native structure recapitulation. The native structure (blue) and lowest-energy model (cyan) for 2 cases from the native structure recapitulation test are shown. In each, the terminal side-chain moiety of the anchor residue is in the same orientation relative to the ligand (magenta). (A) The lowest-energy model generated for a cocrystal of a staphylococcal thermonuclease with thymidine diphosphate (2OF1) corresponded to the native loop length and was nearly identital (0.22-Å C$\alpha$ RMSD) to the native loop configuration. (B) In contrast, the lowest-energy model generated for a cocrystal of a $\beta$-glucosidase from Thermatoga maritima with calystegine B2, a conformationally restricted inhibitor (2CBV) was 1 residue shorter than the native loop length. The RMSD between these 2 structures is not defined. Nonnative lengths can be construed as incorrect in recapitulation experiments but constitute potentially accurate predictions in specificity alteration experiments.

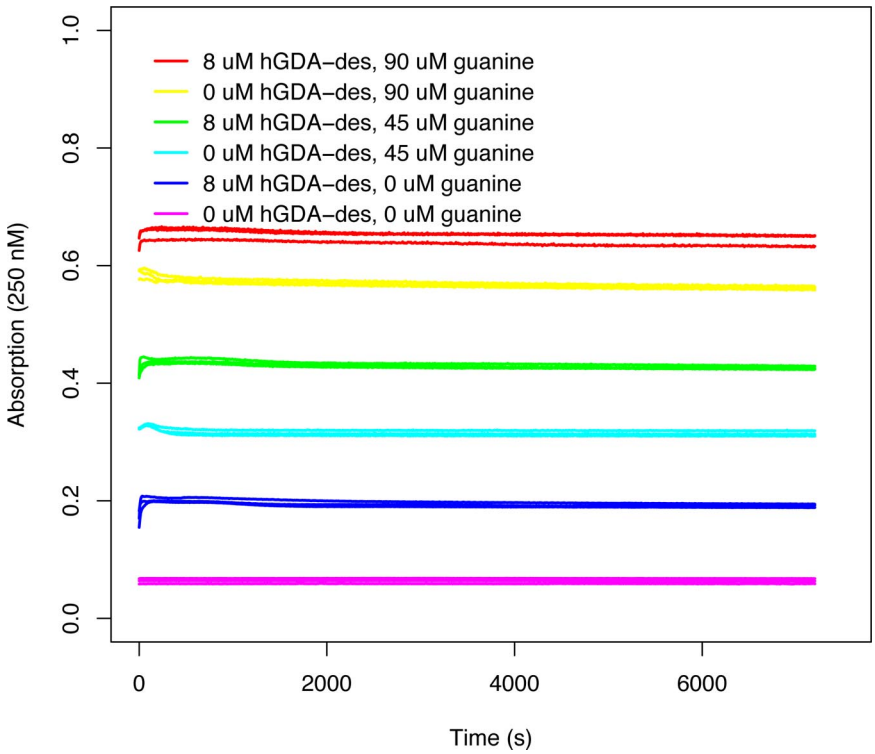


**Fig. S3.** Wild-type activity of hGDA-des. After purification by 1 step of metal-affinity chromatography as above, hGDA-des was concentrated to 5 mg/mL and assayed for guanine deaminase activity. Guanine (Sigma–Aldrich) was prepared at 100 $\mu$M and 50 $\mu$M in PBS (pH 7.5). Reactions were performed in triplicate by adding 10 $\mu$L of 80 $\mu$M hGDA-des to 90 $\mu$L of 100 $\mu$M or 50 $\mu$M guanine in a half-area UV transparent-bottom 96-well plate (Corning Inc). Absorption at 250 nm was measured for 2 h with a SpectraMax microplate spectrophotometer (Molecular Devices). No significant change in signal was observed. The change in absorption at 250 nm for conversion of 100 $\mu$M guanine to 100 $\mu$M xanthine (Sigma–Aldrich) at 100-$\mu$L volume is $-0.2$ AU. An upper bound on the activity of hGDA-des toward guanine can be estimated from these data. The change in A250 is $<0.01$ AU at 1,000 sec for reactions performed at 8 $\mu$M hGDA-des and 90 $\mu$M guanine. This corresponds to $<5$ $\mu$M product formed and a velocity of $<6e{-}4$ s$-1$. In comparison, the velocity of hGDA-wt in these conditions would be 15 s$-1$, for a (vwt/vdes)g $>$ 2.5e4 at 90 $\mu$M guanine. As shown above, the relative velocity (vdes/vwt)ad $=$ 1.0e2 at 500 $\mu$M ammelide, for a total specificity switch of (vad/vg)des $\times$ (vg/vad)wt $>$ 2.5e6. Wild-type activity measured by the same procedure is consistent with literature values.

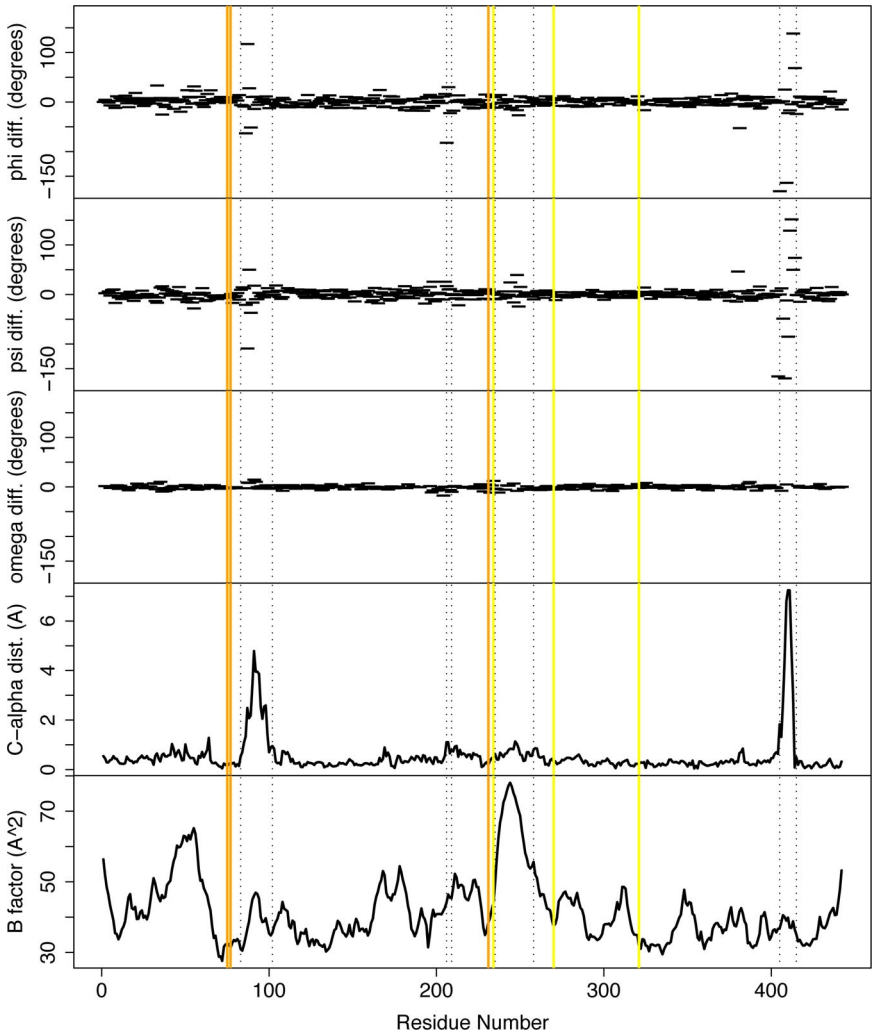


**Fig. S4.** Parametric comparison of hGDA-des computational model and crystal structure A chain. For each residue in the designed protein, 5 parameters relevant to interpreting the differences between the computational model and the crystal structure are plotted. The angular difference between backbone $\phi$, $\psi$, and $\omega$ angles for each residue in the structure is shown in the upper 3 images. C$\alpha$ distance after superimposition is shown in the fourth image, and B-factor for each C$\alpha$ atom is shown in the fifth image. Vertical lines represent residues of note in the structure. Orange lines indicate the location of metal-binding histidines H82, H84, H238. Yellow lines indicate the location of catalytic residues D328, E241, H277. Sequential pairs of dotted black lines indicate the location of residues 90–109 (active site lid), 213–216 (designed loop), 242–265 (high B-factor region adjacent to designed loop), 412–422 (loop adjacent to active site lid).

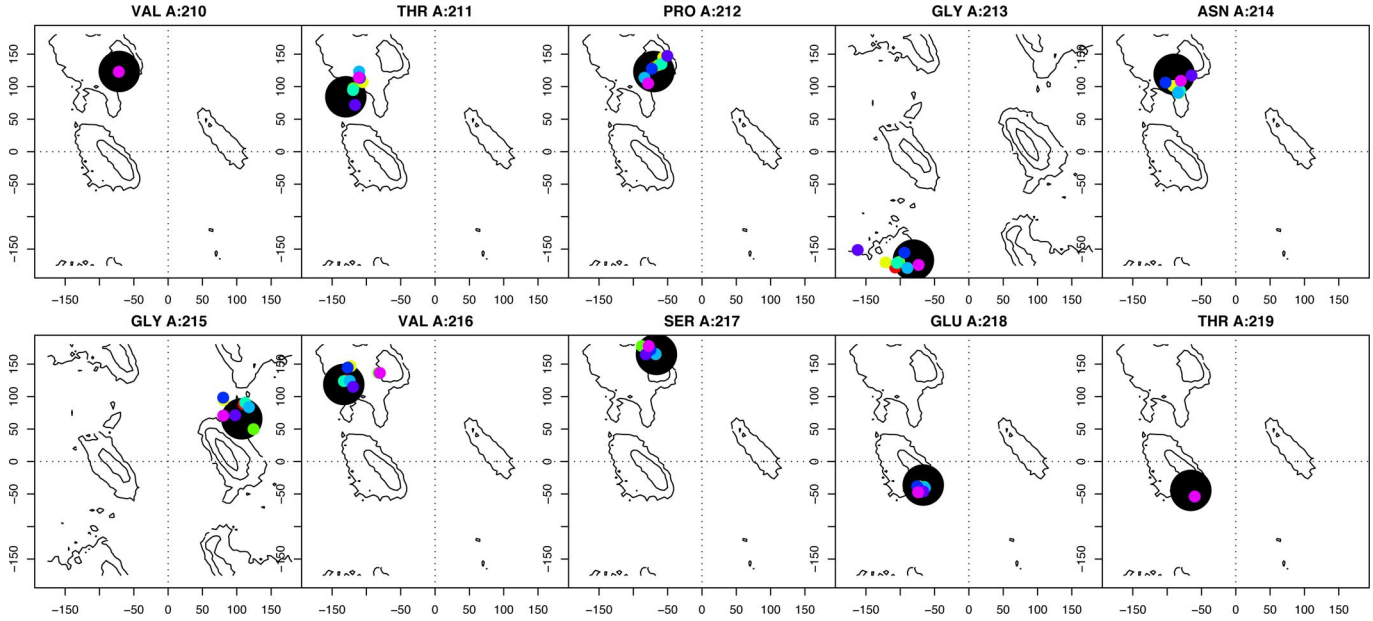


**Fig. S5.** Ramachandran plot for designed residues of hGDA-des crystal structure and computational models. For each residue of the designed loop, the $\phi$-$\psi$ angles of the backbone of the crystal structure are in the same region of the Ramachandran plot as the computational models. The value of the $\phi$-$\psi$ angles of the crystal structure are depicted as a larger black dot, whereas the value of the $\phi$-$\psi$ angles of several computational models consistent with transition state binding are shown as smaller colored dots. The $\phi$-$\psi$ angles of surrounding residues were not perturbed by the design. Contours represent the log frequency of $\phi$-$\psi$ angles in protein loop regions in high-resolution crystal structures.

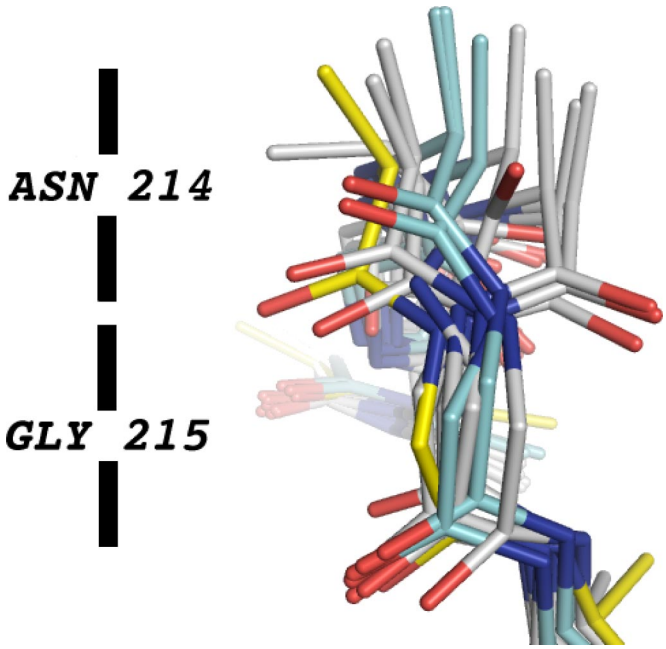


**Fig. S6.** Backbone structure of designed loop. The 2 chains in the crystal structure are shown in cyan. The backbone structure of the computational design model, which is optimally compatible with transition state binding, is shown in yellow. An isoenergetic ensemble of structures, produced at the corroborative loop-modeling stage, is shown in gray. This ensemble was modeled in the absence of ligand to assess the prebound conformation of the designed loop. The ensemble is shifted slightly from the optimal binding conformation and is consistent with the conformation of the crystal structure, which also does not contain substrate.

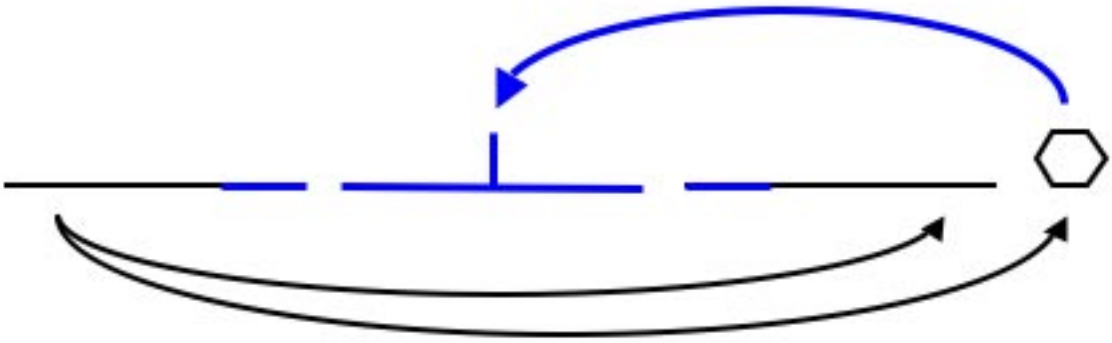

**Fig. S7.** MiniRosetta fold tree. The structure of the miniRosetta fold tree used for kinematics during the course of the loop remodeling and design phase of the protocol is depicted. The polypeptide backbone is depicted as a horizontal line; the constrained side chain is depicted as a vertical line; the ligand is depicted as a hexagon. Jumps between the polypeptide and backbone are depicted as curved arcs. Rigid degrees of freedom are shown in black, whereas moveable degrees of freedom are shown in blue. Chain breaks in the backbone are shown as gaps in the horizonal line representing the polypeptide backbone.

**Table S1. Benchmark and design parameters and results**

| PDB | Ligand | | | Anchor | | | Loop | | |
|---|---|---|---|---|---|---|---|---|---|
| ID code | Type | Index | Atom | Type | Index | Atom | Begin | End | MinRMSD | MinE RMSD |
| 1gua | GNP | B:170 | HN1 | ASP | A:119 | OD1 | A:116 | A:122 | 0.19 | 0.19 |
| 1k70 | HPY | B:501 | O2 | GLN | A:153 | 1HE2 | A:150 | A:156 | 0.68 | 0.91 |
| 2cbu | CTS | B:1447 | O4 | GLN | A:18 | 1HE2 | A:15 | A:21 | 0.44 | 0.65 |
| 2cbv | CGB | B:1447 | HO5 | GLU | A:402 | OE2 | A:399 | A:405 | 0.76 | 0.76 |
| 2i02 | FMN | B:200 | O1P | LYS | A:55 | H | A:52 | A:58 | 3.40 | 3.66 |
| 2jfg | ADP | B:1441 | N7 | ASN | A:270 | 1HD2 | A:267 | A:273 | 0.21 | 0.27 |
| 2of1 | THP | B:501 | O5P | ARG | A:81 | 2HH1 | A:78 | A:84 | 0.17 | 0.19 |
| 2uz9 | XAN | B:1453 | N7 | ARG | A:206 | HE | A:203 | A:209 | 0.88 | 1.26 |
| 2uz9 | HPY | B:501 | O2 | ASN | (Designed) | 1HD2 | A:204 | A:212 | n/a | n/a |
| 2uz9 | HPY | B:501 | O2 | GLN | (Designed) | 1HE2 | A:204 | A:212 | n/a | n/a |

The benchmark and design results used to generate the data in Fig. 1 and Fig. S1 were produced by using the above parameters. The PDB file of the specified accession number was downloaded from www.rcsb.org. The ligand was assigned chain B, the polypeptide was assigned chain A and was renumbered starting from 1 to account for missing density. The residues between the Loop Begin and End were excised from a model of the structure. The rigid-body orientation between the Ligand Atom and Anchor Atom was preserved, and 100 structures were generated for each loop length. The residue numbers reported in the manuscript have not been renumbered to account for missing density. For each of the 200 loops generated for the native loop length and anchor position (3 residues before, 3 residues after), 2 RMSDs are reported above. The MinRMSD column shows the lowest RMSD of any of the 200 loops, whereas the MinE RMSD column shows the RMSD of the minimum energy structure. Many of these RMSDs are below 1 A, indicating that energy is a sensitive test for proximity to the native structure, and that our sampling protocol tightly constrains generated structures to near-native configurations.

**Table S2. Sequences of oligonucleotides used for cloning and mutagenesis of hGDA**

| Oligonucleotide | Sequence |
|---|---|
| pET29b(+)-hGDA-fwd (Nde) | GAC-CAC-ATA-TGT-GTG-CCG-CTC-AGA-TGC-CG |
| pET29b(+)-hGDA-rev (Xho) | GCA-ACC-TCG-AGC-GAG-CCC-ACT-GAG-CTG-GAA-AAC-GGA-AC |
| pET15b-hGDA-fwd (Nde) | GAC-CAC-ATA-TGT-GTG-CCG-CTC-AGA-TGC-CG |
| pET15b-hGDA-rev (BamHI) | GCA-ACG-GAT-CCT-TAC-ACT-GAG-CTG-GAA-AAC-GGA-AC |
| des-fwd | CCA-GGC-AAT-GGC-GTG-TCT-GAG-ACT-TTG-ATG-GG |
| des-rev | CAA-AGT-CTC-AGA-CAC-GCC-ATT-GCC-TGG-TGT-CAC-TAT-GGG-C |
| des-N213Q-rev | CAA-AGT-CTC-AGA-CAC-GCC-CTG-GCC-TGG-TGT-CAC-TAT-GGG-C |
| des-N213A-rev | CAA AGT CTC AGA CAC GCC CGC GCC TGG CTGT CAC TAT GGG |
| wt-R213N-rev | GGA-AAA-ATT-TGG-TGT-CAC-TAT-GGG-C |
| wt-R213Q-rev | GGA-AAA-CTG-TGG-TGT-CAC-TAT-GGG-C |
| wt-F214N-rev | GAG-GGA-ATT-ACG-TGG-TGT-CAC-TAT-GG |
| wt-F214Q-rev | GAG-GGA-CTG-ACG-TGG-TGT-CAC-TAT-GG |
| E243A-rev | CGA-TTC-GCA-CTT-ATA-TGG-CTC-TGA-ATG-TG |

**Table S3. Slope of kinetic traces of Fig. 5*A***

| Clone | Slope, $\mu$ M/h | 95% CI | ●●● |
|---|---|---|---|
| wt | 0.05004 | −0.01825 | 0.1183 |
| wt-E243A | 0.08115 | −0.04104 | 0.2033 |
| wt-R213N | 0.8677 | 0.7198 | 1.015 |
| wt-R213Q | 0.1294 | 0.03710 | 0.2216 |
| wt-F214N | 0.1436 | 0.03899 | 0.2481 |
| wt-F214Q | 0.04544 | −0.01317 | 0.1040 |
| des | 5.424 | 5.121 | 5.725 |
| des-E241A | 0.05626 | −0.01274 | 0.1252 |
| des-N214Q | 0.5031 | 0.3530 | 0.6531 |
| des-N214A | 0.4462 | 0.3051 | 0.5873 |

**Table S4. Data collection and refinement statistics**

| | |
|---|---|
| Wavelength, Å | 0.9774 |
| Resolution, Å | 117–2.37 |
| (Highest shell) | 2.45–2.37 |
| Total reflections | 44,921 |
| Unique reflections | 43,645 |
| Completeness, % | 97.3 |
| $R_{merge}$ | 8.6 |
| Average $I/\sigma I$ | 11.8 |
| Refinement statistics | |
| Resolution, Å | 117–2.37 |
| No. of reflections | 44,845 |
| Test set (5%) | 2,205 |
| $R_{cryst}$ | 20.5 |
| $R_{free}$ | 24.8 |
| RMS deviations | |
| Bonds, Å | 0.007 |
| Angles,° | 1.027 |
| No. of atoms | 7,011 |
| Protein/water/metal | 6,965/44/2 |
| Average B-factors, Å$^2$ | 41.34 |
| Protein/water/metal | 41.40/31.73/36.72 |