SUPPLEMENTARY INFORMATION

Computational design of an orthogonal nucleoside analog kinase

Lingfeng Liu[1], Paul Murphy[2], David Baker[2,3], Stefan Lutz[1]*

[1]Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA 30322, [2]Department of Biochemistry, University of Washington, Box 357350, Seattle, WA 98195 & [3]Howard Hughes Medical Institute, Seattle, WA 98195

Corresponding author: Stefan Lutz (sal2@emory.edu)

MATERIAL AND METHODS

Materials: Oligodeoxynucleotides were ordered from Integrated DNA Technologies (Coralville, IA). *Pfu* Turbo DNA polymerase (Strategene, La Jolla, CA) was used for DNA amplification. Restriction enzymes were purchased from New England Biolabs (Ipswich, MA) while pyruvate kinase and lactate dehydrogenase were obtained from Roche Biochemicals (Indianopolis, IN). All other reagents were obtained from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

Computational design: The Rosetta suite for molecular modeling (Rosetta++ 2.3.1) was used for design predictions. The coordinates for DmdNK with a bound thymidine (PDB access#: 1OT3 [1]) were chosen as starting point for the modeling. Following replacement of the substrate thymidine with the modeled structure of the nucleoside analog ddT, two rounds of design were performed.

In the first round, residues which contained atoms within 6.5 angstroms of the thymidine-O3' were designed. The threshold was chosen to include residues that are part of the ribose binding pocket and in the immediate vicinity of the 3'-position of the substrate. Six residues were selected (I29, L66, M69, Y70, E172, and V175). The design protocol used for this first round was fixed-backbone design using Monte Carlo sampling of an expanded version of the Dunbrack rotamer library according to the Rosetta full-atom energy function.[2,3] The results of the first round revealed that I29 and M69 were not situated appropriately, despite their proximity, to interact the substrate near the thymidine O3'.

For this reason, a second round of design including only L66, Y70, E172, and V175 was performed. Since substitutions to E172 had already been identified as being critical to substrate specificity, the second round calculations were performed twice, once with and once without E172, using residues (L66, Y70, and E172) and (L66, Y70, and V175) respectively. The side chains of other residues in the vicinity of the four targeted amino acids were allowed to vary in conformation only. No changes were made to the conformation of the polypeptide backbone or to the position of waters in the binding pocket.

Given the small set of residues (n=3), combinations of new amino acid identities for these residues were sampled exhaustively and assignments for each of the 8000 (20^3) combinations were made. Residues within 10Å of the designed residues were allowed to repack and minimize with respect to the side chain degrees of freedom. The position of the substrate was held fixed. Calculations were performed for transition state models of ddT and Thy independently. The predicted energy of interaction between the enzyme and ddT (ΔG_{ddT}) was used to estimate the catalyst's activity for the NA. Separately, the difference in energies of interaction ($\Delta\Delta G$) was calculated by subtracting ΔG_{Thy} from ΔG_{ddT} to provide an approximation of relative specificity. Of the 8000 possibilities, designs with maximal $\Delta\Delta G$ and low ΔG_{ddT} values were identified. Using both criteria favored designs with high activity and specificity for ddT relative to Thy.

Site-directed mutagenesis: Site-specific amino acid substitutions in DmdNK were introduced by primer overlap extension mutagenesis. PCR products were cloned into pMAL-c2x (Invitrogen, Carlsbad, CA) via EcoRI and HindIII restriction sites and transformed into *E. coli* TOP10 (Invitrogen). Transformants were cultured on LB-agar plates, supplemented with ampicillin (50 μg/mL) at 37˚C and colonies were harvested, flash-frozen, and stored at -80˚C. The correct gene sequences were confirmed by DNA sequencing.

Protein expression and purification: For in vitro characterization, WT DmdNK and selected kinase mutants were overexpressed as fusion proteins with a N-terminal maltose-binding protein (MBP). Individual kinase genes were subcloned into pMAL-C2x (NEB) and transformed into *E. coli* K12 TB1 (NEB). Cell cultures were grown in 100 mL LB media containing ampicillin (100 μg/mL) to an OD (600 nm) of ~ 0.5 at 37℃ and protein expression was induced with 0.3 mM IPTG for 12 h at 20˚C. The cells were centrifuged and pellets were resuspended in 5 mL lysis buffer (20 mM Tris-HCl, pH 7.4; 0.2 M NaCl; 1 mM EDTA; 1 mM ATP) and mixed with 50 μL protease inhibitor cocktail (Sigma) and 5 μL benzonase (Novagen, Madison, WI). Cells were lyzed by sonication and the suspension centrifuged (10,000x g, 4˚C, 20 minutes). The clear supernatant was incubated with 600 μL amylose resin (NEB) and the resin was loaded on a Prep column (BioRad, Carlsbad, CA). Following washing the resin with

10 column volumes (CV) of lysis buffer containing 1 mM maltose, bound protein was eluted with lysis buffer supplemented with 10 mM maltose, yielding protein of >95% purity as determined by SDS-PAGE (see Fig. S1). The protein concentration was measured by absorbance at 280 nm (MBP-DmdNK, ε = 106,230 M^{-1} cm^{-1}) and sample aliquots were flash-frozen for storage at -80˚C.

Steady-state kinetic assays: The kinase activity of recombinant enzymes was determined using a spectrophotometric coupled-enzyme assay.[4] Briefly, 2′-deoxyribonucleosides and NAs at 1 μM to 7 mM were prepared in reaction buffer containing 50 mM Tris-HCl (pH 8), 0.15 M NaCl, 2.5 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 0.21 mM phosphoenolpyruvate, 0.18 mM NADH, and 30 units/mL pyruvate kinase and 33 units/mL lactate dehydrogenase. Assays were performed at 37˚C, measuring the absorbance change at 340 nm in the presence of enzyme. The enzyme amount was adjusted to limit NADH turnover to 10% over the time of the experiment. All experiments were performed in triplicate and kinetic data was determined by non-linear regression analysis using the Michaelis-Menten equation in Origin®7 (OriginLab, Northampton, MA). An representative example of the primary data for the determination of kinetic parameters for RosD7 is shown in Fig. S2. Separate kinetic experiments with RosD7 lacking the MBP fusion tag did not indicate a significant change in the kinetic performance of the enzyme (data not shown).

Thermostability: Enzyme activity after 10-minute incubation at 4˚C and 37˚C, respectively, was determined using the standard kinetic assay with thymidine (0.7 mM) and ATP (1 mM). Thermostability (TS) was expressed as % residual activity = 100% * v_{max}(37˚C)/v_{max}(4˚C). All experiments were performed in triplicate.

REFERENCES

1. N. E. Mikkelsen, K. Johansson, A. Karlsson, W. Knecht, G. Andersen, J. Piskur, B. Munch-Petersen and H. Eklund, *Biochemistry*, 2003, **42**, 5706-5712.

2. B. Kuhlman, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.,* 2000, **97**, 10383-10388.

3. C.A. Rohl, C.E. Strauss, K.M. Misura, D. Baker, *Methods Enzymol.*, 2004, **383**, 66-93.

4. M. L. Gerth and S. Lutz, *Biochem. Biophys. Res. Commun.*, 2007, **354**, 802-807.

Table S1: Comparison of kinetic properties of designer kinase (RosD4) with substitutions in position 172.

enzyme		T	ddT	RS	enzyme		T	ddT	RS
RosD4	k_{cat} (s^{-1})	4.6 ± 0.1	0.4 ± 0.01		**RosD4-E172Y**	k_{cat} (s^{-1})	0.21 ± 0.01	0.28 ± 0.01	
(L66F, Y70M,		**(-3)**	**(-1.3)**		(L66F, Y70M,		**(-61)**	**(-1.9)**	
V175Y)	K_M (μM)	56 ± 2	36 ± 2		V175Y, E172Y)	K_M (μM)	86 ± 4	50 ± 6	
		(-20)	**(+3)**				**(-32)**	**(+2.3)**	
	k_{cat}/K_M	83	11	0.13		k_{cat}/K_M	2.4	5.6	2.3
	(10^3 M^{-1} s^{-1})	**(-58)**	**(+2.4)**			(10^3 M^{-1} s^{-1})	**(-2005)**	**(+1.2)**	
RosD5	k_{cat} (s^{-1})	0.08 ± 0.01	0.19 ± 0.01		**RosD4-E172M**	k_{cat} (s^{-1})	0.18 ± 0.01	0.35 ± 0.01	
(L66F, Y70M,		**(-160)**	**(-2.8)**		(L66F, Y70M,		**(-71)**	**(-1.5)**	
V175Y, E172V)	K_M (μM)	96 ± 15	35 ± 4		V175Y, E172M)	K_M (μM)	66 ± 9	52 ± 6	
		(-36)	**(+3.3)**				**(-24)**	**(+2.3)**	
	k_{cat}/K_M	0.84	5.4	6.4		k_{cat}/K_M	2.7	6.8	2.5
	(10^3 M^{-1} s^{-1})	**(-5730)**	**(+1.1)**			(10^3 M^{-1} s^{-1})	**(-1780)**	**(+1.5)**	
RosD6	k_{cat} (s^{-1})	0.21 ± 0.01	0.41 ± 0.01		**RosD4-E172A**	k_{cat} (s^{-1})	0.24 ± 0.01	0.46 ± 0.01	
(L66F, Y70M,		**(-60)**	**(-1.3)**		(L66F, Y70M,		**(-53)**	**(-1.1)**	
V175Y, E172I)	K_M (μM)	66 ± 7	35 ± 4		V175Y, E172A)	K_M (μM)	65 ± 5	50 ± 4	
		(-24)	**(+3.3)**				**(-24)**	**(+2.3)**	
	k_{cat}/K_M	3.2	12	3.7		k_{cat}/K_M	3.7	9.3	2.5
	(10^3 M^{-1} s^{-1})	**(-1500)**	**(+2.6)**			(10^3 M^{-1} s^{-1})	**(-1300)**	**(+2)**	
RosD4-E172P	k_{cat} (s^{-1})	0.16 ± 0.01	0.28 ± 0.01		**RosD4-E172T**	k_{cat} (s^{-1})	0.19 ± 0.01	0.44 ± 0.01	
(L66F, Y70M,		**(-80)**	**(-1.9)**		(L66F, Y70M,		**(-68)**	**(-1.2)**	
V175Y, E172P)	K_M (μM)	74 ± 12	42 ± 2		V175Y, E172T)	K_M (μM)	91 ± 5	47 ± 3	
		(-27)	**(+2.7)**				**(-34)**	**(+2.4)**	
	k_{cat}/K_M	2.2	6.7	3		k_{cat}/K_M	2.0	9.4	4.7
	(10^3 M^{-1} s^{-1})	**(-2190)**	**(+1.5)**			(10^3 M^{-1} s^{-1})	**(-2400)**	**(+2)**	
RosD4-E172H	k_{cat} (s^{-1})	0.19 ± 0.01	0.29 ± 0.01		**RosD4-E172G**	k_{cat} (s^{-1})	0.21 ± 0.01	0.23 ± 0.01	
(L66F, Y70M,		**(-68)**	**(-1.9)**		(L66F, Y70M,		**(-61)**	**(-2.3)**	
V175Y, E172H)	K_M (μM)	77 ± 7	47 ± 2		V175Y, E172G)	K_M (μM)	92 ± 10	54 ± 7	
		(-28)	**(+2.4)**				**(-34)**	**(+2.1)**	
	k_{cat}/K_M	2.5	6.2	2.5		k_{cat}/K_M	2.3	4.3	1.9
	(10^3 M^{-1} s^{-1})	**(-1925)**	**(+1.3)**			(10^3 M^{-1} s^{-1})	**(-2100)**	**(-1.1)**	
RosD4-E172L	k_{cat} (s^{-1})	0.22 ± 0.01	0.44 ± 0.02		**RosD4-E172F**	k_{cat} (s^{-1})	0.20 ± 0.01	0.35 ± 0.01	
(L66F, Y70M,		**(-58)**	**(-1.2)**		(L66F, Y70M,		**(-65)**	**(-1.5)**	
V175Y, E172L)	K_M (μM)	101 ± 21	58 ± 8		V175Y, E172F)	K_M (μM)	75 ± 8	41 ± 3	
		(-37)	**(+2)**				**(-28)**	**(+2.8)**	
	k_{cat}/K_M	2.2	7.6	3.5		k_{cat}/K_M	2.7	8.5	3.1
	(10^3 M^{-1} s^{-1})	**(-2190)**	**(+1.7)**			(10^3 M^{-1} s^{-1})	**(-1780)**	**(+1.8)**	

Grey-shaded table entry is parental enzyme. Numbers in parentheses are fold changes in catalytic efficiency for the particular substrate [k_{cat}/K_M (variant)/k_{cat}/K_M (DmdNK)]. RS: Relative specificity [k_{cat}/K_M (ddT)/k_{cat}/K_M (T)].

Table S2: Comparison of kinetic properties of designer kinase (RosD6) with substitutions in position 175.

enzyme		T	ddT	RS
RosD6	k_{cat} (s^{-1})	0.21 ± 0.01	0.41 ± 0.01	
(L66F, Y70M,		**(-60)**	**(-1.3)**	
V175Y, E172I)	K_M (μM)	66 ± 7	35 ± 4	
		(-24)	**(+3.3)**	
	k_{cat}/K_M	3.2	12	3.7
	(10^3 M^{-1} s^{-1})	**(-1500)**	**(+2.6)**	
RosD7	k_{cat} (s^{-1})	0.42 ± 0.02	0.65 ± 0.02	
(L66F, Y70M,		**(-31)**	**(+1.2)**	
V175W, E172I)	K_M (μM)	173 ± 32	32 ± 4	
		(-64)	**(+3.6)**	
	k_{cat}/K_M	2.4	20.5	8.5
	(10^3 M^{-1} s^{-1})	**(-2000)**	**(+4.5)**	
RosD6-Y175F	k_{cat} (s^{-1})	0.12 ± 0.01	0.34 ± 0.01	
(L66F, Y70M,		**(-108)**	**(-1.6)**	
V175F, E172I)	K_M (μM)	202 ± 79	120 ± 16	
		(-75)	**-1**	
	k_{cat}/K_M	0.6	2.8	4.7
	(10^3 M^{-1} s^{-1})	**(-8000)**	**(-1.6)**	
RosD6-Y175I	k_{cat} (s^{-1})	0.10 ± 0.01	0.38 ± 0.04	
(L66F, Y70M,		**(-130)**	**(-1.4)**	
V175I, E172I)	K_M (μM)	331 ± 131	800 ± 183	
		(-123)	**(-7)**	
	k_{cat}/K_M	0.30	0.48	1.6
	(10^3 M^{-1} s^{-1})	**(-16,000)**	**(-10)**	
RosD6-Y175L	k_{cat} (s^{-1})	0.22 ± 0.01	0.53 ± 0.04	
(L66F, Y70M,		**(-60)**	**-1**	
V175L, E172I)	K_M (μM)	367 ± 75	302 ± 73	
		(-136)	**(-2.6)**	
	k_{cat}/K_M	0.60	1.8	3
	(10^3 M^{-1} s^{-1})	**(-8000)**	**(-2.5)**	
RosD6-Y175M	k_{cat} (s^{-1})	0.13 ± 0.01	0.34 ± 0.03	
(L66F, Y70M,		**(-100)**	**(-1.6)**	
V175M, E172I)	K_M (μM)	167 ± 25	353 ± 88	
		(-62)	**(-3)**	
	k_{cat}/K_M	0.78	1.0	1.3
	(10^3 M^{-1} s^{-1})	**(-6170)**	**(-4.6)**	

Grey-shaded table entry is parental enzyme. Numbers in parentheses are fold changes in catalytic efficiency for the particular substrate [k_{cat}/K_M (variant)/k_{cat}/K_M (DmdNK)]. RS: Relative specificity [k_{cat}/K_M (ddT)/k_{cat}/K_M (T)].

Figure S1: SDS-PAGE analysis of overexpression experiment for RosD7.

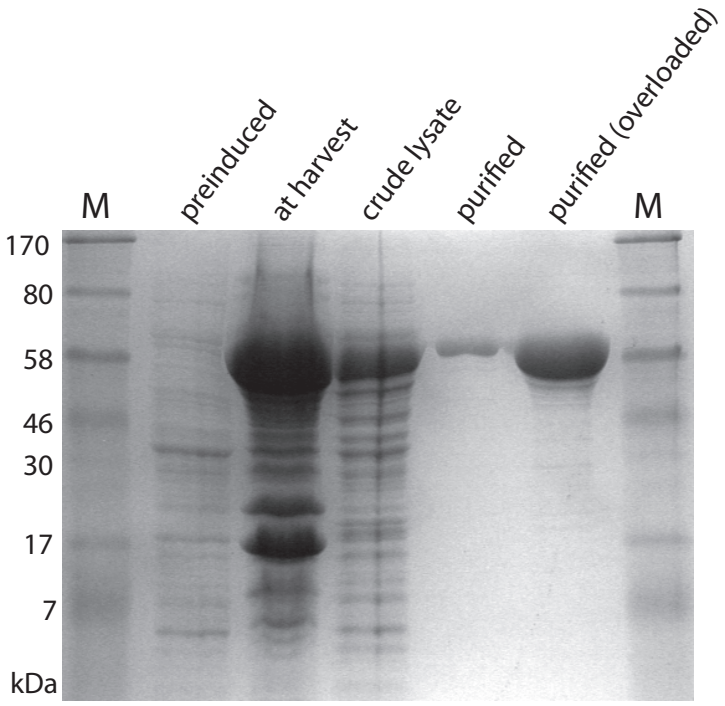

Figure S2: Primary kinetic data for RosD7 with 2',3'-dideoxythymidine (ddT) and thymidine. Kinetic parameters (K_M and v_{max}) were determined by non-linear regression analysis using the Michaelis-Menten equation in Origin®7 (OriginLab, Northampton, MA).