Supplementary Information

The Role of Side Chain Entropy and Mutual Information for Improving the De Novo Design of Kemp Eliminase Enzymes KE07 and KE70

Asmit Bhowmick[1], Sudhir Sharma[2], Hallie Hhonma[3], and Teresa Head-Gordon[1,2,3,4*]

[1]Department of Chemical and Biomolecular Engineering, [2]Department of Chemistry, and [3]Department of Bioengineering, University of California Berkeley
[4]Chemical Sciences Division, Lawrence Berkeley National Labs
Berkeley, California 94720, USA

Table S1. *Convergence of Side chain entropy (SCE) and Mutual Information(MI)*. 5 independent trial simulations were done on a random backbone ensemble from R7 variant. Reported are values for SCE, MI and MI without correction. Values reported are in units of k_BT.

	Trial-1	Trial-2	Trial-3	Trial-4	Trial-5	Avg	Stdev
SCE	355.6	349.2	354.0	353.0	352.0	352.7	2.39
MI	975.6	986.7	889.6	843.6	938.6	926.8	60.0
MI - uncorrected	3397.6	3805.2	2841.8	2437.8	3311.1	3158.7	528.7

Table S2. List of primers used for mutagenesis reaction to generate the specific mutations in R7-2 template[a]

Variants	Primers
H201A	Forward 5' CA TTG CCG ATC ATT GCA GCG AGG GGA GCT GGC AAG ATG 3' Reverse 5' CAT CTT GCC AGC TCC CCT CGC TGC AAT GAT CGG CAA TG 3'
K222A	Forward 5' GT GCA GAC GCG GCT GCG GCC GAT TCG GTT TTT C 3' Reverse 5' G AAA AAC CGA ATC GGC CGC AGC CGC GTC TGC AC 3'
R16Q	Forward 5' CAT TAA TAA TGA AGG ATG GCC AGG TTG TCA AAG GTA GC 3' Reverse 5' GCT ACC TTT GAC AAC CTG GCC ATC CTT CAT TAT TAA TG 3'
N25S	Forward 5' GTA GCA ATT TTG AAA GCC TGC GTG ACT CTG 3' Reverse 5' CAG AGT CAC GCA GGC TTT CAA AAT TGC TAC 3'
L170A	Forward 5' CC GGC GAA ATT GTG GCG GGT TCA ATT GAC CGC 3' Reverse 5' GCG GTC AAT TGA ACC CGC CAC AAT TTC GCC GG 3'
Q185A	Forward 5' CC GGC GAA ATT GTG GCG GGT TCA ATT GAC CGC 3' Reverse 5' GCG GTC AAT TGA ACC CGC CAC AAT TTC GCC GG 3'
[a] Changed nucleotides are shown in red text.	

Table S3. *Mutations made in various rounds of directed evolution of KE07*. The computationally designed residues (red) and mutated residues introduced by LDE of a given round (black) have been listed in the table below. The experimental k_{cat} and K_M values and representative variant names have been taken from (Khersonsky et al., 2010)

Sequence Position (Directed Evolution Round)							
	KE07 design	**R2 11/10D**	**R3 I3/10A**	**R4 1E/11H**	**R5 10/3B**	**R6 3/7F**	**R7 10/11G**
ILE 7			Gln	Asp	Asp	Asp	Asp
ALA 9							
ILE 11							
VAL 12					Met		Met
LYS 19		Glu				Glu	
SER 48							
TRP 50							
PHE 77							Ile
HIS 84							
PHE 86			Leu				
GLU 101 (catalytic base)							
ILE 102							Phe
GLN 123		Arg					
TYR 128							
ALA 130							
LYS 146		Thr	Thr	Glu		Thr	Thr
VAL 169							
GLY 171							
LEU 176							
HIS 201							
GLY 202		Arg	Arg	Arg	Arg	Arg	Arg
MET 207							
LYS 222							
ASN 224		Asp	Asp	Asp	Asp	Asp	Asp
PHE 229			Ser				Ser
k_{cat} (s^{-1})	**0.02**	**0.02**	**0.21**	**0.70**	**0.49**	**0.60**	**1.37**
K_M (mM)	**1.40**	**0.31**	**0.48**	**2.40**	**0.59**	**0.69**	**0.54**
k_{cat} / K_M ($M^{-1}s^{-1}$)	**12.2**	**66.0**	**425**	**291**	**836**	**872**	**2590**

Table S4. *Mutations made in various rounds of directed evolution of KE70*. The computationally designed residues(red) , other mutated residues introduced by LDE of a given round (black) and residues after which insertions took place (green) have been listed in the table below. The experimental k_{cat} and K_M values and representative variant names have been taken from [1]

Sequence Position (Directed Evolution Round)	KE70 Design	R2 7/12F	R4 4/1B	R5 7/4A	R6 4/8B
HIS 17 (catalyzing)					
ALA 19					
THR 20			Ser		Ser
ALA 21					
ASP 23		Gly			
LYS 29			Asn	Asn	Asn
THR 43			Asn	Asn	Asn
ASP 45 (catalyzing)					
TYR 48		Phe	Phe	Phe	Phe
TRP 72			Cys	Cys	Cys
SER 74					Gly
GLY 101				Ser	Ser
ALA 103					
SER 138			Ala	Ala	Ala
HIS 166				Asn	Asn
VAL 168					
THR 171			Pro	Pro	
GLY 177					
ALA 178					Ser
LYS 197					Asn
THR 198					Ile
ILE 202					
ALA 204			Val	Val	Val
ASP 212		Glu			
ALA 231				Ser	
ALA 235					
SER 239			Ser	Ser	Ala
HIS 251		Tyr			
k_{cat} (s^{-1})	**0.14**	**0.32**	**1.66**	**5.38**	**5.00**
K_M (mM)	**1.11**	**0.24**	**0.18**	**0.14**	**0.09**
k_{cat} / K_M ($M^{-1}s^{-1}$)	**126**	**1330**	**9240**	**37800**	**57300**

Table S5. *Residues that were determined to be network hubs with high mutual information for KE70 as a function of LDE round*. Residues colored red were designed and residues colored blue were mutated during the course of LDE; the only exception is that residue 48 was both a designed and mutated residue. The network hubs identified for KE70, 23 and 48 were mutated in R2, hub residue 29 in R2 was mutated in R4, and hub residue 197 was mutated in R6.

Round	Higher MI in EL complex state	Higher MI in Apo state
KE70	6, 11, 14, 17, 23, 24, 38, 45, 48, 58, 67, 70, 74, 83, 100, 104, 115, 117, 121, 147, 153, 166, 167, 173, 184, 186, 188, 191, 193, 216, 217, 221, 247	27, 64, 143
R2	11, 14, 17, 29, 30, 58, 67, 84, 122, 153, 188, 189	5, 64, 68, 70, 100, 109, 135, 141, 146, 147, 191, 193, 208, 215, 222
R4	18, 25, 33, 50, 64, 116, 121, 147, 170, 174, 189, 191, 197	5, 16, 22, 35, 58, 70, 90, 123, 209, 215, 233
R5	5, 24, 41, 64, 68, 79, 82, 83, 95, 153, 187, 196, 221, 232, 247	11, 49, 52, 77, 92, 115, 147, 165, 173
KE70-R6	10, 15, 17, 22, 58, 76, 123, 165, 232	11, 18, 25, 33, 35, 45, 50, 52, 56, 59, 64, 67, 70, 83, 90, 115, 118, 148, 154, 170, 174, 198, 223, 247

Table S6. *Residues that were determined to be network hubs with high mutual information for KE07 as a function of LDE round*. Residues colored red were designed into the scaffold of 1THF and residues colored blue were mutated during the course of LDE; the only exception is 224 that was both a designed and mutated residue in KE07. The network hub residues 19 and 86 identified in the designed enzyme were mutated in R2 and R3, respectively. In R2, we observe that residues 7 and 224 are network hubs that were subsequently mutated in R3. In R3, we observe that residues 7, 146, and 224 are hubs that were subsequently mutated in R4. In further improved variants like R5, we see that residues 19 and 102 appeared as network hubs, and were subsequently mutated in later rounds.

Round	Higher MI in EL complex state	Higher MI in Apo state
KE07	16, 19, 58, 68, 85, 86, 139, 163, 174, 175, 185, 230, 232, 235	4, 10, 63, 66, 71, 87, 118, 212
R2	58, 163, 222, 224, 242	6, 7, 10, 22, 34, 64, 87, 102, 113, 185, 187, 231, 236, 244
R3	42, 52, 60, 61, 62, 174	7, 22, 24, 46, 50, 59, 71, 72, 95, 146, 154, 179, 188, 193, 208, 212, 222, 224, 228, 238
R4	6, 42, 51, 59, 62, 63, 65, 85, 146, 159, 174, 175, 185, 201, 206, 230, 232, 243, 244	31, 34, 50, 72, 95, 156, 163, 191, 235, 238
R5	26, 52, 58, 62, 63, 64, 67, 102, 132, 133, 137, 148, 155, 167, 175, 208, 212, 239, 242	10, 19, 50
R6	41, 52, 91, 185, 212, 214, 236	10, 50, 62, 86, 99, 209, 222, 235, 238, 249
KE07-R7	5, 19, 62, 63, 71, 73, 84, 87, 91, 92, 118, 148, 155, 159, 199, 244, 247	12, 16, 25, 42, 52, 74, 94, 95, 128, 132, 133, 149, 201, 202, 209, 222, 230

Table S7. Side chain dihedral angles for the residues highlighted in Figure 2 of the main text. Note that the rotamer classification has been done using the Dunbrack 2007 library.

Residue	Occupation %	χ_1	χ_2	χ_3	χ_4
Ser 48	57	176			
	43	-64			
Trp 50	91	180	-88		
	9	180	-123		
Glu 101	47	180	180	88	
	34	180	180	-60	
Tyr 128	42	180	80		
	36	180	102		
His 201	90	180	-103		
	7	180	-136		
Arg 202	99	-66	180	180	180
	< 1	-62	-75	75	78
Lys 222	72	180	180	-70	-70
	25	-65	180	180	-64

REFERENCES

1. O. Khersonsky, D. Röthlisberger, A. M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K. N. Houk, D. Baker and D. S. Tawfik, *J. Mol. Bio.*, 2011, **407**, 391-412.