ACS Synthetic Biology

Supporting information for

A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases

JKB Cahn, CA Werlang, A Baumschlager, S Brinkmann-Chen, SL Mayo, and FH Arnold

Supporting Table 1. NADP-to-NAD cofactor specificity reversal. Structures followed by (h) are homology models, while those followed by another PDB accession code use the cofactor from that protein and (m) denotes a structure of a mutant protein. Mutations indicated in italics are distal to the 2' position. For citation information, see **Supporting Material 2**.

Study	Protein	Structure	Best Variant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Banta et al. 2002)	*C. glutamicum* DKR	1A80	R238H	?	?	?
(Baroni et al. 2012)	*P. falciparum* FDNR	2OK7	Y258F	0.67	$4.7x10^{1}$	-1.2
(Bastian et al. 2011)	*E. coli* KARI	3ULK	A71S, R76D, S78D, *Q110V*	190	$5.4x10^{4}$	-0.07
(Brinkmann-Chen et al. 2013)	*S.* sp. KARI	3ULK (h)	A71S, R76D, S78D, *Q110V*	64	$1.1x10^{5}$	-2.1
(Brinkmann-Chen et al. 2013)	*S. exigua* KARI	4KQW	S61D, S63D, *I95V*	88	$7.8x10^{3}$	-1.6
(Brinkmann-Chen et al. 2013)	*M. aeolicus* KARI	4KQW (h)	G50D, S52D	120	$1.2x10^{3}$	-0.84
(Brinkmann-Chen et al. 2013)	*L. lactis* KARI	4TSK (h)	V48L, R49P, K52L, S53D, *E59K*, *T182S*, *E320K*	150	$2.3x10^{4}$	0.04
(Brinkmann-Chen et al. 2013)	*A. acidocaldarius* KARI	4TSK	R48P, S51L, S52D, *R84A*	110	$4.5x10^{2}$	-1.5
(Chen et al. 1995)	*E. coli* IDH	4AJ3	*C201I*, *C332Y*, K344D, Y345I, V351A, Y391K, R395S	200	$1.4x10^{6}$	-1.5
(Dambe et al. 2006)	*S. morelense* AFDH	2GLX	*A13G*, S33D	14	?	-0.79
(Döhr et al. 2001)	*H. sapiens* P450R	3QFS	*W676A*	0.24	$1.0x10^{3}$	-0.55
(Elmore and Porter 2002)	*R. norvegicus* P450R	1AMO	*W677A*	1.2	$5.3x10^{4}$	-1.8
(Eppink et al. 1999)	*P. fluorescens* PHBH	1K0J (m)	*R33S*, *Q34R*, *P36R*, *D37A*, Y38E	3.1	$5.0x10^{4}$	-2.0
(Fasan et al. 2011)	*B. megaterium* P450R	4DQL	R966N, K972H, Y974F, W1046D	0.62	$4.4x10^{2}$	-1.1
(Gand et al. 2016)	*S.* sp. IRED	3ZHB (h)	S37V, K40A	5.5	$1.8x10^{2}$	-1.6
(Kamerbeek et al. 2004)	*P. fluorescens* HAPMO	2YLR (h)	K439F	0.62	$4.3x10^{2}$	-2.0
(Kamerbeek et al. 2004)	*A.* sp. CHMO	4RG3	K326A	0.10	$5.5x10^{1}$	-2.9
(Katzberg et al. 2010)	*S. cerevisae* DKR	4PVD	N9E	0.86	$1.2x10^{2}$	-2.9
(Khoury et al. 2009)	*C. boidinii* XR	1K8C (h)	K272G, S273G, N274D	>1	?	-1.1
(Kristan et al. 2007)	*C. lunatus* BHSDH	3QWF	Y49D	7.8	?	-4.2
(Liang et al. 2007)	*P. stipitis* XR	1K8C (h)	K270R, N272D	2.9	9.3	-1.2
(Maddock et al. 2015)	*E. coli* CaADH	1KEV (h)	G198D, S199V, P201E, Y218A	>1	?	-4.6
(Maurer et al. 2005)	*B. megaterium* P450R	4DQL	R966D, *W1046S*	0.37	$2.3x10^{2}$	-0.40

Supporting Table 1. continued

Study	Protein	Structure	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Medina et al. 2001)	*A.* PCC7119 FDNR	2BSA	S223D	0.12	$8.1x10^3$	-5.4
(Nakanishi et al. 1997)	*M. musculus* CR	1CYD	T38D	31	$1.3x10^3$	-0.51
(Paladini et al. 2009)	*P. sativum* FDNR	4AF7 (1QGA)	*Y308S*	0.02	$3.3x10^2$	-1.6
(B. Petschacher et al. 2005)	*C. tenuis* XR	1K8C	K274R, N276D	1.2	$1.9x10^1$	-1.2
(Pick et al. 2014)	*E. coli* AdhZ3	1YQD (h)	S199N, S200N, N201D	1.2	$6.9x10^1$	-0.75
(Pick et al. 2014)	*E. coli* AdhZ2	1UUF (1YQD)	T205D, T206I, S207N	2.2	$3.8x10^1$	0.00
(Rane and Calvo 1997)	*E. coli* KARI	3ULK	R68D, K69L, K75V, R76D	31	$5.8x10^4$	-0.55
(Rodriguez-Arnedo et al. 2005)	*H. volcanii* IDH	1AI2 (h)	R291S, K343D, Y344I, V350A, Y390P	>1	?	-0.77
(Rosell et al. 2003)	*R. perezi* ADH8	1P0F	G223D, T224I, H225N	7600	$8.1x10^4$	0.55
(Schepens et al. 2000)	*S. bicolor* MDH	7MDH (1CIV)	G84D, S85I, R87Q, S88A	12	$2.1x10^4$	-0.96
(Scrutton et al. 1990)	*E. coli* GTR	1GET	*A179G*, *A183G*, V197E, R198M, K199F, H200D, R204P	8.1	$1.8x10^4$	-1.5
(Shiraishi et al. 1998)	*N. crassa* CbR		S920D, R932S	65	$7.2x10^4$	-2.6
(Takase et al. 2014)	*S.* sp. A1-R	3AFN	H37N, G38S, R39H, K40V, A41D	>1	?	-3.2
(Yaoi et al. 1996)	*T. thermophilus* IDH	2D1C	R231A, K283D, Y284I, N287G, *V288I*, I290A	68	$5.2x10^4$	-1.1
(Zeng et al. 2009)	*P. stipitis* XR	1K8C (h)	K21A, N272D	>1	?	-0.10
(L. Zhang et al. 1999)	*V. harveyi* ALDH	1EZ0	T175E	130	$4.8x10^3$	-0.76
(R. Z. Zhang et al. 2009)	*C. parapsilosis* SCR	3CTM (1CYD)	S67D, P69D	0.31	4.6	-0.18
This study	*A. thaliana* GR	3DOJ (3PEF)	R31L, T32K, K35D, C68R	2.4	$3.3x10^1$	-0.71
This study	*S. cerevisae* CinADH	1PIW	S210D, R211P, K215E, S253P	2.0	$6.6x10^1$	0.77
This study	*T. emersonii* XR	1K8C (h)	S272G, N273G, R277Y, Q280E	5.2	$4.8x10^3$	-1.5
This study	*T. maritima* FeADH	1VHD	G36E, S38N, S39G	43	$8.4x10^1$	0.47

Catalytic efficiency, *CE*, is given as k_{cat}/K_M when available, or v_{max}/K_M otherwise.

[a]Final specificity is defined as $\boldsymbol{CE_{mut}^{NAD}/CE_{mut}^{NADP}}$ [b]Specificity change is defined as $\frac{CE_{mut}^{NAD}/CE_{mut}^{NADP}}{CE_{WT}^{NAD}/CE_{WT}^{NADP}}$ [c]Log relative activity is defined as $\mathbf{log}\left(\frac{CE_{mut}^{NAD}}{CE_{WT}^{NADP}}\right)$

Supporting Table 2. NAD-to-NADP cofactor specificity reversal. Structures followed by (h) are homology models, while those followed by another PDB accession code use the cofactor from that protein and (m) denotes a structure of a mutant protein. Mutations indicated in italics are distal to the 2' position. For citation information, see **Supporting Material 2**.

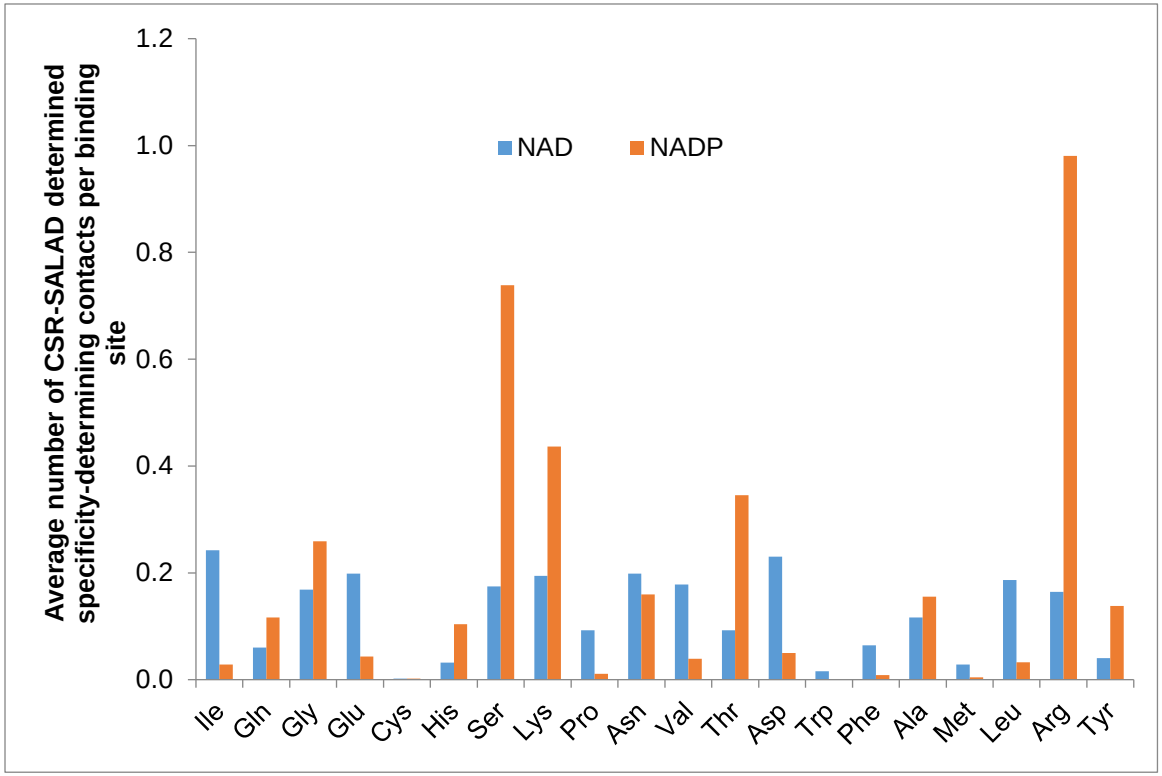
Study	Protein	Structure	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Ashida et al. 2004)	*S.* sp AlaDH	2VHW (h)	D198A	14	$3.4x10^4$	-0.96
(Bernard et al. 1995)	*L. delbruckii* LDH	1J49	D175A	1.0	$4.4x10^1$	-0.96
(Bocanegra et al. 1993)	*E.coli* DHLDH	4JQ9 (1GEU)	E205V, M206R, F207K, *D208H*, P212R	>1	?	0.67
(Bubner et al. 2008)	*P. fluorescens* M2DH	1M2W	E68K, D69A	19	$4.5x10^3$	0.77
(Capone et al. 2011)	*C. symbiosum* GDH	1BGV (4XGI)	F238S, P262S	0.32	$5.8x10^1$	-3.3
(Chen et al. 1996)	*T. thermophilus* IMDH	2ZTW	N/A	1000	$8.7x10^4$	0.19
(Clermont et al. 1993)	*B. stearothermophilus* GAPDH	3CMC	D32A, L187A, P188S	1.6	?	-1.7
(Cui et al. 2015)	*G. oxydans* Gox2181	3AWD (2WDZ)	Q20R, D43S	1.5	?	0.07
(Ehrensberger et al. 2006)	*G. oxydans* XDH	1ZEM	D38S, M39R	>1	?	0.70
(Ehsani et al. 2009)	*S. cerevisae* BDH	2D8A (h)	E221S, I222R, A223S	>1	?	0.06
(Feeney et al. 1990)	*B. stearothermophilus* LDH	1LDN	D53S	0.15	3.4	-1.3
(Friesen et al. 1996)	*P. mevalonii* HMG-CoAR	4I4B	D146A, L148K	0.14	$7.6x10^4$	-3.7
(Galkin et al. 1997)	*T. intermedius* LuDH	1LEH (h,1BW9)	D203A, I204R, *D210R*	74	?	-1.6
(Gul-Karaguler et al. 2001)	*C. methylica* FDH	2FSS (2NAD)	D195S	0.02	$6.1x10^3$	-1.5
(Hoelsch et al. 2013)	*M. vaccae* FDH	2GSD (h)	*C145S*, A198G, D221Q, *C225V*	13	?	?
(Holmberg et al. 1999)	*B. stearothermophilus* LDH	1LDN	I51K, D52S	2.2	$4.9x10^1$	-1.6
(Hong et al. 2016)	*B. cereus* AlDH	4PT0	E194S	0.47	$2.1x10^1$	-0.46
(Hsieh et al. 2006)	*H. sapiens* m-NAD-ME	1PJ3	Q362K	3.2	$2.9x10^2$	-0.53

Supporting Table 2. continued

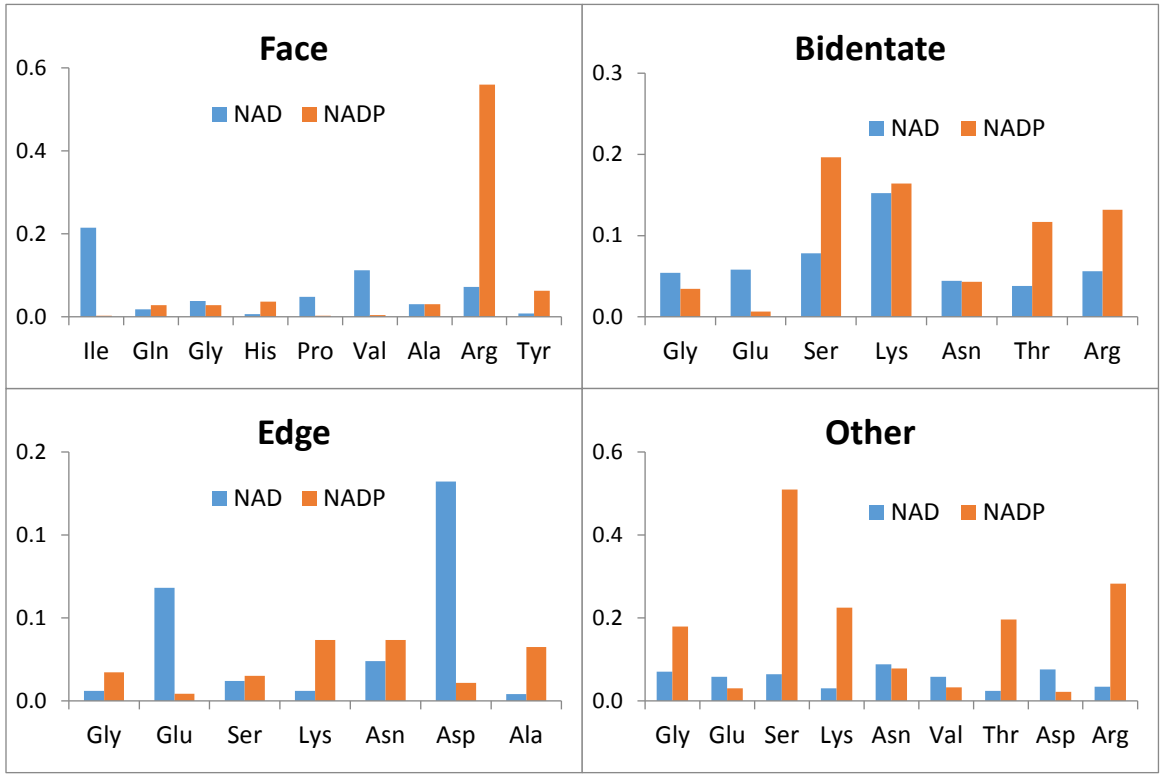
Study	Protein	Structure	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Jensen et al. 2013)	*S. maltphilia* SMFMO	4A9W (2XLP)	H194T	0.97	1.5	-0.22
(Ma et al. 2010)	*K. pneumonia* PDOR	3OX4 (h)	D41G	1.56	?	-0.66
(Marohnic et al. 2003)	*R. norvegicus* CB5R	1IB0	D239T	10	$4.1x10^4$	-0.67
(Miller et al. 2006)	*E. coli* IMDH	1CM7 (2ZTW)	*K100R*, *A229T*, *D236R*, *L248M*, D289K, I290Y, A296V, *G337Y*	370	$5.1x10^4$	-0.23
(Nishiyama et al. 1993)	*T. flavus* MDH	1BMD	E41G, I42S, P43E, Q44R, A45S, *M46F*, *K47Q*	24	$5.2x10^2$	-0.47
(Petschacher et al. 2014)	*S. mutans* NOX	2BC0 (h,2CDU)	D192A, V193R, V194H, A199R	10	$6.4x10^4$	0.56
(Serov et al. 2002)	*S. cerevisae* FDH	2NAD (h)	D196A, Y197R	2.3	?	-3.8
(Takase et al. 2014)	*S.* sp. A1-R'	4TKM	T16S, E17Q, N37H, S38G, H39R, V40K, D41A	85	$1.1x10^3$	0.67
(Watanabe et al. 2005)	*P. stipitis* XDH	1PL6 (h)	D207A, I208R, F209T	2.6	$1.1x10^4$	0.05
(Woodyer et al. 2003)	*P. stutzeri* PDH	4E5K	E175A, A176R	2.8	$3.0x10^2$	1.0
(Zheng et al. 2013)	*B. subtilis* InDH	3NT2	*A12K*, D35S, V36R	4.8	?	0.03

Catalytic efficiency, *CE*, is given as k_{cat}/K_M when available, or v_{max}/K_M otherwise.

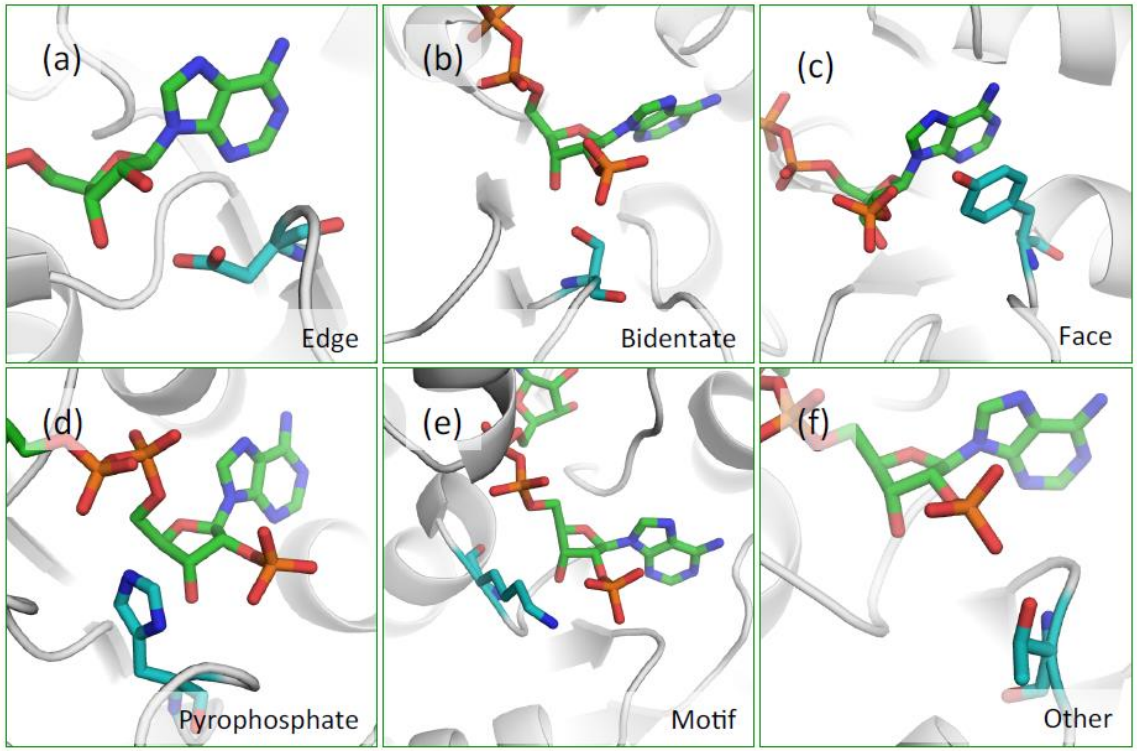
[a]Final specificity is defined as $\boldsymbol{CE_{mut}^{NAD}/CE_{mut}^{NADP}}$ [b]Specificity change is defined as $\frac{CE_{mut}^{NAD}/CE_{mut}^{NADP}}{CE_{WT}^{NAD}/CE_{WT}^{NADP}}$ [c]Log relative activity is defined as $\mathbf{log}\left(\frac{CE_{mut}^{NAD}}{CE_{WT}^{NADP}}\right)$



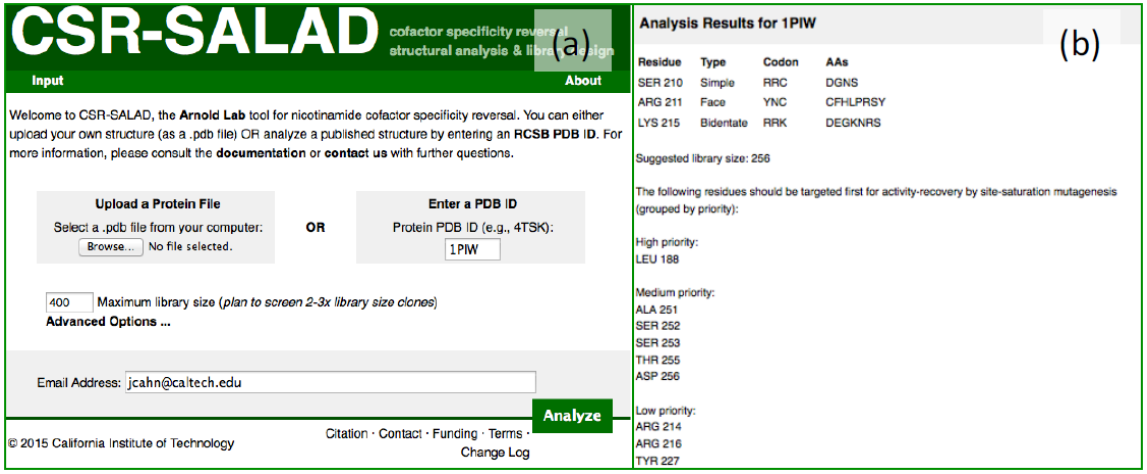
Supporting Figure 1. Average number of residues of each amino acid comprising the specificity-determining pockets for NAD-bound and NADP-bound structures in the PDB, compiled from a representative set of 499 NAD-bound structures and 463 NADP-bound structures selected on the basis of sequence identity and resolution. The criteria for selecting 'specificity-determining' residues are laid out in **Supporting Material 1**. It is important to note that NADP binding pockets on average comprise more residues, 3.58 vs 2.89 on average for NAD binding pockets.



Supporting Figure 2. The same data as **Supporting Fig. 1**, but broken out into the four most important structural classes (see **Supporting Fig. 3** and **Supporting Material 1**). For clarity, only the five most utilized amino acids for each of the two cofactors are shown for each category.



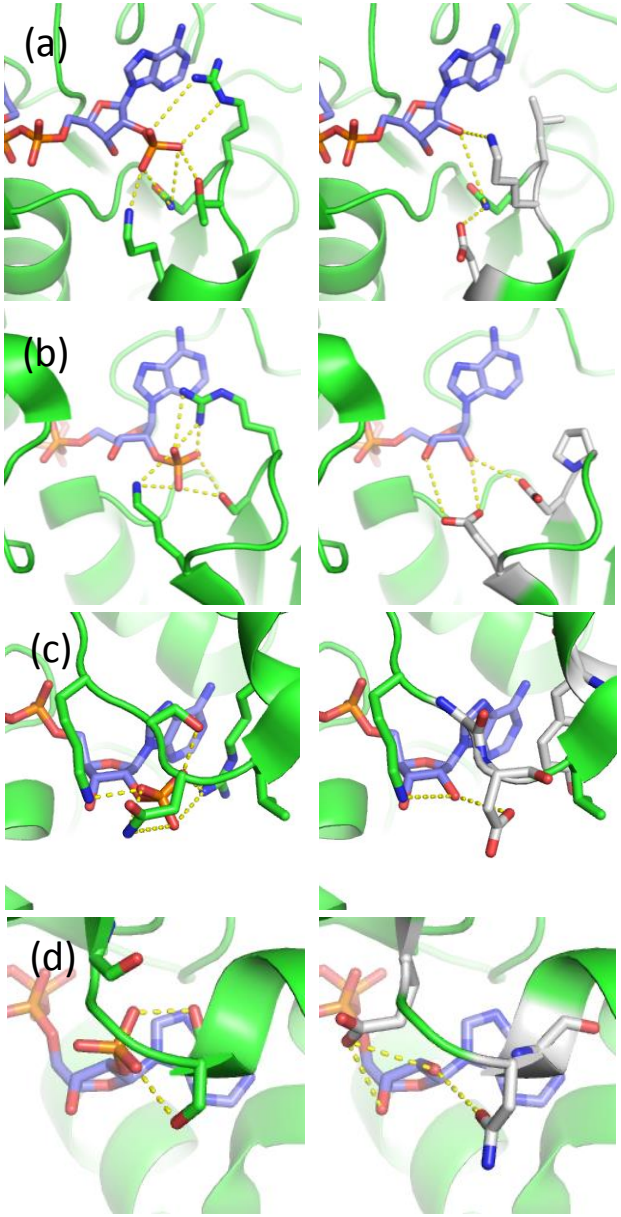
Supporting Figure 3. Examples of the six structural classifications used by CSR-SALAD. For more details, see the **Supporting Material 1**. Figures are from structures 1VC2, 1AMO, 1AMO, 1EZ0, 1CYD, and 1CYD.



Supporting Figure 4. The web interface of CSR-SALAD (a) and an example of the corresponding output (b).

Supporting Table 3. Kinetics of enzymes used for validation of CSR-SALAD. k_{cat} values are apparent values for the provided reaction conditions.

	NADH			NADPH			k_{cat}/K_M Specificity
Protein	k_{cat} (min^{-1})	K_M (μM^{-1})	k_{cat}/K_M (min^{-1} μM^{-1})	k_{cat} (min^{-1})	K_M (μM^{-1})	k_{cat}/K_M (min^{-1} μM^{-1})	NADH/NADPH
A. thaliana GR	440 ± 50	76 ± 24	5.8 ± 1.9	1600 ± 230	20 ± 6.0	80 ± 27	0.07 ± 0.03
*At*GR R31L, T32K, K35D, *C68R*	1100 ± 190	67 ± 15	15 ± 4.5	590 ± 140	91 ± 21	6.5 ± 2.1	2.4 ± 1.0
S. cerevisae CinADH	180 ± 70	55 ± 23	3.4 ± 2.0	12400 ± 50	110 ± 50	110 ± 44	0.03 ± 0.02
*Sc*CinADH S210D, R211P K215E, *S253P*	19000 ± 7900	30 ± 23	650 ± 510	19000 ± 1700	60 ± 43	320 ± 230	2.0 ± 2.2
T. emersonii XR	8.0 ± 1.0	260 ± 18	0.03 ± 0.00	2500 ± 220	85 ± 73	30 ± 25	0.0 ± 0.0
*Te*XR S272G, N273D, R277Y, *Q280E*	110 ± 40	110 ± 76	1.0 ± 0.7	40 ± 4.0	210 ± 20	0.2 ± 0.0	5.2 ± 4.0
T. maritima FeADH	1300 ± 140	750 ± 400	1.8 ± 1.0	1400 ± 270	400 ± 90	3.5 ± 1.0	0.5 ± 0.3
*Tm*FeADH G36E, S38N, S39G	1200 ± 10	110 ± 110	10 ± 10	280 ± 80	1200 ± 1500	0.2 ± 0.3	43 ± 71



Supporting Figure 5. Structures of wild-type enzymes (or homology models) (left) and models of the mutations made (right) for *A. thaliana* GR (a), *S. cerevisae* CinADH (b), *T. emersonii* XR (c), and *T. maritima* FeADH (d). Mutations are colored with white carbons and were introduced in PyMol without any structural minimization or optimization. Selected hydrophilic interactions are shown with yellow dashed lines.

Supporting Table 4. Recapitulation of beneficial mutations for NADP-to-NAD cofactor switching in default-setting CSR-SALAD libraries. Residues are colored red if the mutation is found in the switching library, orange if the residue is targeted for mutagenesis but the mutations suggested do not include the previously published mutant, blue if the residue is recommended for saturation as a recovery mutation, and black if the mutation would not be found by CSR-SALAD. Mutations indicated in italics are distal to the 2' position.

Protein	Structure	Method	Best Mutant	Switched?[a]	Highly Active?[b]
***S.* sp. A1-R**	3AFN	Loop Substitution	H37N, G38S, R39H, K40V, A41D	x	
***R. perezi* ADH8**	1P0F	Homology	G223D, T224I, H225N	x	x
***E. coli* AdhZ2**	1UUF (1YQD)	Rational	T205D, T206I, S207N	x	x
***E. coli* AdhZ3**	1YQD (h)	Semi-Rational	S199N, S200N, N201D	x	x
***S. morelense* AFDH**	2GLX	Homology	*A13G*, S33D	x	x
***V. harveyi* ALDH**	1EZ0	Semi-Rational	T175E	x	x
***C .lutantus* BHSDH**	3QWF	Homology	Y49D	x	
***E. coli* CaADH**	1KEV (h)	Homology	G198D, S199V, P201E, Y218A	x	
***A.* sp. CHMO**	4RG3	Semi-Rational	K326A		
***M. musculus* CR**	1CYD	Rational	T38D	x	x
***A.* PCC7119 FDNR**	2BSA	Rational	S223D		
***P. falciparum* FDNR**	2OK7	Rational	Y258F		
***P. sativum* FDNR**	4AF7 (1QGA)	Rational	*Y308S*		
***S. cerevisae* gDKR**	4PVD	Semi-Rational	N9E		
***E. coli* GTR**	1GET	Homology	*A179G*, *A183G*, V197E, R198M, K199F, H200D, R204P	x	
***P. fluorescens* HAPMO**	2YLR (h)	Semi-Rational	K439F		
***E. coli* IDH**	4AJ3	Homology	*C201I*, *C332Y*, K344D, Y345I, V351A, Y391K, R395S	x	
***H. volcanii* IDH**	1AI2 (h)	Homology	R291S, K343D, Y344I, V350A, Y390P	x	x

Supporting Table 4. continued

Protein	Structure	Method	Best Mutant	Switched?[a]	Highly Active?[b]
***T. thermophilus* IDH**	2D1C	Loop Substitution	R231A, K283D, Y284I, N287G, *V288I*, I290A	x	
***S.* sp IRED**	3ZHB (h)	Homology	S37V, K40A	x	
***A. acidocaldarius* KARI**	4TSK	Semi-Rational	R48P, S51L, S52D, *R84A*	x	
***E. coli* KARI**	3ULK	Semi-Rational	A71S, R76D, S78D, *Q110V*	x	x
***E. coli* KARI**	3ULK	Rational	R68D, K69L, K75V, R76D	x	x
***L. lactis* KARI**	4TSK (h)	Semi-Rational	V48L, R49P, K52L, S53D, *E59K*, *T182S*, *E320K*	x	x
***M. aeolicus* KARI**	4KQW (h)	Rational	G50D, S52D	x	x
***S. exigua* KARI**	4KQW	Rational	S61D, S63D, *I95V*	x	
***S.* sp. KARI**	3ULK (h)	Rational	A71S, R76D, S78D, *Q110V*	x	
***S. bicolor* MDH**	7MDH (1CIV)	Homology	G84D, S85I, R87Q, S88A	x	x
***B. megaterium* P450R**	4DQL	Rational	R966D, *W1046S*		x
***H. sapiens* P450R**	3QFS	Rational	*W676A*		x
***R. norvegicus* P450R**	1AMO	Rational	*W677A*	x	
***B. megaterium* P450R**	4DQL	Semi-Rational	R966N, K972H, Y974F, *W1046D*		
***P. fluorescens* PHBH**	1K0J (m)	Rational	*R33S*, *Q34R*, *P36R*, *D37A*, Y38E	x	
***C. parapsilosis* SCR**	3CTM (1CYD)	Rational	S67D, P69D		x
***C. boidinii* XR**	1K8C (h)	Computational	K272G, S273G, N274D	x	
***C. tenuis* XR**	1K8C	Rational	K274R, N276D	x	
***P. stipitis* XR**	1K8C (h)	Computational	K21A, N272D	x	x
***P. stipitis* XR**	1K8C (h)	Semi-Rational	K270R, N272D	x	

[a] A switched enzyme has $CE_{mut}^{NAD} / CE_{mut}^{NADP} > \mathbf{1}$, where catalytic effiency (*CE*) represents k_{cat}/K_M or V_{max}/K_M

[b] An active enzyme is defined as one with $\frac{CE_{mut}^{NAD}}{CE_{WT}^{NADP}} > \mathbf{0.1}$

Supporting Table 5. Recapitulation of beneficial mutations for NAD-to-NADP cofactor switching in default-setting CSR-SALAD libraries. Residues are colored red if the mutation is found in the switching library, orange if the residue is targeted for mutagenesis but the mutations suggested do not include the previously published mutant, blue if the residue is recommended for saturation as a recovery mutation, and black if the mutation would not be found by CSR-SALAD. Mutations indicated in italics are distal to the 2' position.

Protein	Structure	Method	Best Mutant	Switched?[a]	Highly Active?[b]
***K. pneumonia* PDOR**	3OX4 (h)	Computational	D41G	x	x
***S.* sp. A1-R'**	4TKM	Loop Substitution	T16S, E17Q, N37H, S38G, H39R, V40K, D41A	x	x
***S.* sp. AlaDH**	2VHW (h)	Rational	D198A	x	x
***B. cereus* AlDH**	4PT0	Homology	E194S	x	
***S. cerevisae* BDH**	2D8A (h)	Homology	E221S, I222R, A223S	x	x
***R. norvegicus* CB5R**	1IB0	Homology	D239T	x	x
***E. coli* DHLDH**	4JQ9 (1GEU)	Homology	E205V, M206R, F207K, *D208H*, P212R	x	x
***C. methylica* FDH**	2FSS (2NAD)	Rational	D195S		
***M. vaccae* FDH**	2GSD (h)	Rational	*C145S*, A198G, D221Q, *C225V*	x	
***S. cerevisae* FDH**	2NAD (h)	Rational	D196A, Y197R	x	
***B. stearothermophilus* GAPDH**	3CMC	Rational	D32A, L187A, P188S	x	
***C. symbiosum* GDH**	1BGV (4XGI)	Rational	F238S, P262S		
***G. oxydans* Gox2181**	3AWD (2WDZ)	Computational	Q20R, D43S	x	x
***P. mevalonii* HMG-CoAR**	4I4B	Rational	D146A, L148K		
***E. coli* IMDH**	1CM7 (2ZTW)	Directed Evolution	*K100R, A229T, D236R, L248M*, D289K, I290Y, A296V, *G337Y*	x	x

Supporting Table 5. continued

Protein	Structure	Method	Best Mutant	Switched?[a]	Highly Active?[b]
***E. coli* IMDH**	4XXV (h)[c]	Directed Evolution	*K100R, A229T, D236R, L248M,* D289K, I290Y, A296V, *G337Y*	x	x
***B. subtilis* InDH**	3NT2	Rational	*A12K*, D35S, V36R	x	x
***B. stearothermophilus* LDH**	1LDN	Rational	D53S		
***B. stearothermophilus* LDH**	1LDN	Homology	I51K, D52S	x	
***L. delbruckii* LDH**	1J49	Rational	D175A	x	x
***T. intermedius* LuDH**	1LEH (h,1BW9)	Homology	D203A, I204R, *D210R*	x	
***H. sapiens* m-NAD-ME**	1PJ3	Rational	Q362K	x	x
***P. fluorescens* M2DH**	1M2W	Rational	E68K, D69A	x	x
***T. flavus* MDH**	1BMD	Homology	E41G, I42S, P43E, Q44R, A45S, *M46F, K47Q*	x	x
***S. mutans* NOX**	2BC0 (h,2CDU)	Homology	D192A, V193R, V194H, A199R	x	x
***P. stutzeri* PDH**	4E5K	Semi-Rational	E175A, A176R	x	x
***S. maltphilia* SMFMO**	4A9W (2XLP)	Homology	H194T		x
***G. oxydans* XDH**	1ZEM	Rational	D38S, M39R	x	x
***P. stipitis* XDH**	1PL6 (h)	Homology	D207A, I208R, F209T	x	x

[a]A switched enzyme has $CE_{mut}^{NADP}/CE_{mut}^{NAD} > \mathbf{1}$, where catalytic effiency (*CE*) represents k_{cat}/K_M or V_{max}/K_M

[b]An active enzyme is defined as one with $\frac{CE_{mut}^{NADP}}{CE_{WT}^{NAD}} > \mathbf{0.1}$

[c]A homology model generated using a cofactor-bound homologue gave a better approximation of the geometry of the binding site, we believe, than the insertion of the cofactor into the apo structure. This is reflected in the detection of one additional amino acid.

Supporting Material 1:

Cofactor Specificity Reversal – Structural Analysis and Library Design (CSR-SALAD) Users' Manual Version 1.1

Cofactor Specificity Reversal – Structural Analysis and Library Design

(CSR-SALAD)

Users' Manual
Version 1.1
December 2015

Jackson K.B. Cahn *et al*.
csr.salad.help@gmail.com

Table of Contents

1. Introduction

The Cofactor Specificity Reversal – Structural Analysis and Library Design (CSR-SALAD) web tool has been designed with the goal of taking a common protein engineering target – the alternation of NAD(P) cofactor specificity in an enzyme – and bringing it from the specialized realm of protein engineering to a level accessible to the end-users of these proteins; to render the task as routine and dependable as can be achieved for such a complicated system as a protein.

Cofactor switching has been a well-studied problem, and one for which many results – highly successful and less so – have been published. By learning from the successes of the past, we can design libraries in a semi-rational manner that allows for mutational solutions to be found in a targeted manner, and by automating the analysis process, we hope to allow make CSR-SALAD a regular part of the bioengineering toolbox that does not require special training or skills.

2. How to use CSR-SALAD

CSR-SALAD requires an input structure in .pdb format for analysis, which must have an NAD(P)(H) ligand bound. That's it.

2.1. Fetch a file by accession code

For proteins that have already had structures published, you can access the pdb file directly using the 4-character accession code used in the RCSB PDB, PDBe, or PDBj. Accession codes are not case sensitive.

If you want to make any changes to the PDB file, you'll have to download it and make those changes manually. You can do this by going to the webpages for the PDB (www.rcsb.org/pdb/home/home.do) and downloading the files, which can be opened in any text editor.

Also, it may take a few days after structures become available on the PDB for automatic download to work. These structures can still be downloaded manually and uploaded to CSR-SALAD as described below.

2.2. Upload a file

Files can also be uploaded from your computer using the "Choose File" button on the CSR-SALAD homepage. For security reasons, any characters other than alpha-numerics, period, dash, and whitespace will be stripped from your file before processing. All user-uploaded files are stored as temporary variables, such that they are erased immediately after the output is generated.

2.2.1. File requirements and preparation

Files must be uploaded in standard .pdb format for proper recognition by the Bio.PDB Python package and CSR-SALAD. For more information on the PDB format, see http://www.wwpdb.org/documentation/file-format.

- CSR-SALAD is capable of processing .pdb files with multiple chains, but if the cofactor is in different poses between the chains, the analysis will consider all of them. If you suspect that some of the cofactor-binding poses in your structure are not physiologically relevant, it may be preferable to delete these cofactor

instances. If there are multiple cofactor molecules per chain, CSR-SALAD will not discriminate among them. Similarly, while CSR-SALAD automatically detects NAD and NADP and discriminates between them, structures with both may result in inappropriate analysis. One of the cofactor types should be deleted.

- Cofactors molecules must be named as NAD, NAI, NAP, or NDP, and atom names must be named according to the standard PDB nomenclature, as shown in Figure 1. Missing cofactor atoms at the nicotinamide end of the molecule are acceptable, but atoms of the cofactor required for the structural classification must be present.
- When a structure of your protein is not available, homology models can be created, i.e. with the SWISS-MODEL system (www.swissmodel.expasy.org), or the structure of a high-similarity protein can be used as a stand-in for your protein. Additionally, it may be useful to transfer the cofactor molecule from a protein with a similar fold if your protein has not been co-crystallized with the cofactor. When making adjustments such as these, it is important to manually inspect and adjust the positions of the side-chains (i.e. using PyMol or Coot) around the cofactor to approximate predicted physiological binding geometries.
- Because many structures use inconsistent numbering between chains, all residue numbers are considered modulo 500. That is, residue 121 is assumed to be the same residue as 621 and 1021. To our knowledge, no proteins bind cofactors with multiple residues separated by 500 residues or more, let alone by exactly 500 residues or 1000 residues, but if this poses problems for you consider renumbering certain residues to remove ambiguity.

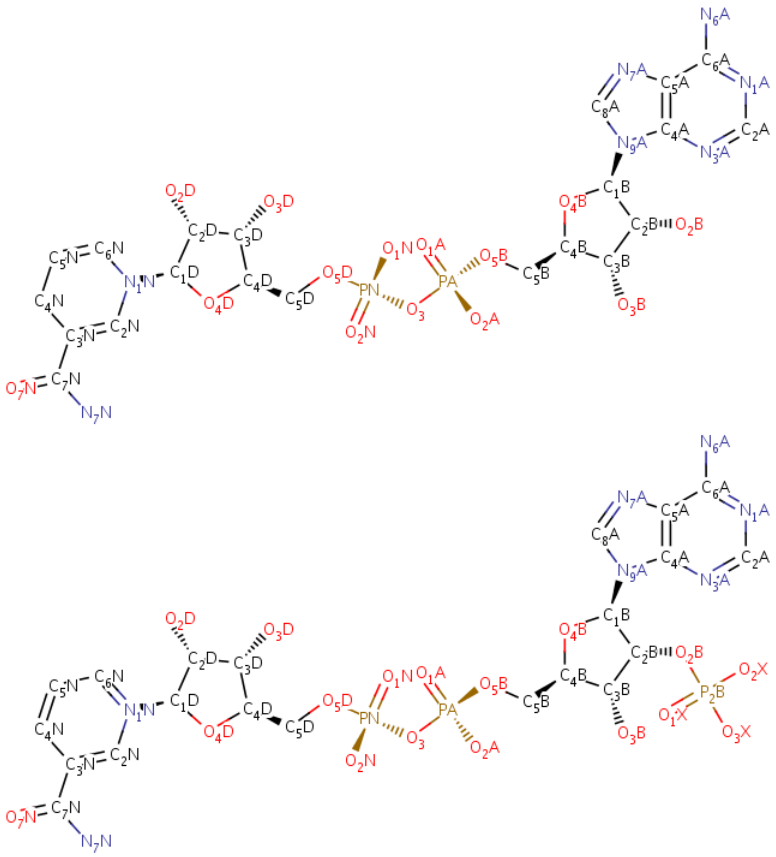


Fig. 1. Atom labels for NAD(H) and NADP(H) from the PDB.

2.3. Select library size

Before running CSR-SALAD, select a maximum size for the cofactor-switching library. This size should be determined by the throughput of your experimental setup; you should aim to screen two to three colonies for each library member to ensure complete coverage. Ideal library sizes will depend on the complexity of the binding site in your cofactor, but parameters under 40 will be rejected. Recommended library sizes might be considerably lower than the maximum entered.

2.4. Advanced options

CSR-SALAD allows for a handful of options to be selected manually. Three allow you to exclude certain classes of residues from consideration in the cofactor-switching library. Exclusion of these residue classes (below) allows for the library to be more focused on the positions likely to be most useful in switching, but also misses the chance of finding beneficial mutations in the first step of engineering.

i. Exclude residues of the glycine-rich "fingerprint motif". This motif is characteristic of the common Rossmann fold, where the glycines hydrogen-bond to the pyrophosphate that bridges the adenosine and nicotinamide moieties. While some of the (non-glycine) residues of this provide interactions with the 2' position of the cofactor that may determine specificity, mutations here have not

frequently been found to be useful in engineering specificity and are often very disruptive.

ii. Residues that, in addition to interacting with the 2' position of the cofactor, also interact with the pyrophosphate. These residues often provide a significant amount of binding energy and therefore mutations can be very detrimental to overall cofactor binding ability.

iii. Peripheral residues are those that do not have obvious interactions with the 2' position of the cofactor, but which might nevertheless play a role in cofactor specificity by means of transient interactions in alternate protein conformations or through interactions with the network of waters around the binding pocket. Mutations at these positions generally have minimal effects, but in some cases have been shown to be important for reversing cofactor specificity.

The final option generates an expansive log of the calculations made by CSR-SALAD during the library design process, focused on the criteria behind residue exclusion/inclusion and structural classification, as outlined below. This log is intended for debugging processes, such as if CSR-SALAD makes decisions you find questionable, but can also give a "peek under the hood".

2.5. Interpreting results

CSR-SALAD results come in three sections.

i. Messages: In some cases, minor problems in structure parsing may prevent completion of the CSR-SALAD analysis. These issues are reported in messages shown at the top of the output. It may be worthwhile to consider modifying the input structure or parameters to resolve these issues before rerunning CSR-SALAD.

ii. Cofactor-switching library: The library suggested for cofactor specificity reversal is presented in a four column table. The first two columns describe the residue present in the wild-type structure, giving its identity, numbering, and a description of the structural role it plays in the protein. For more detail on these structural roles, see 4.2.2.2. The second two columns give the library recommendation made by CSR-SALAD, the former being a degenerate codon (see Table 1) and the latter being the amino acids it codes for. If the codon is listed as '---' the DNA should be left unmutated.

iii. Activity-recovery hotspots: Because cofactor-switching often leads to significant decreases in overall activity, CSR-SALAD recommends positions around the cofactor where mutations ought to have large effects on binding orientation and energetics. They are presented in 1-3 groups, starting from those most likely, in our experience, to provide activating mutations. The types of amino acids presented are:

High Priority

- Residues passed over for mutagenesis on the basis of library size limitations

- Residues passed over for mutagenesis on the basis of the 'exclude' advanced options

Medium Priority

- Residues around the adenine but not involved in determining cofactor specificity (see Cahn et al. 2015, *PEDS*).

Low Priority

- Residues that made hydrogen bond contacts in the input structure with residues included in the switching library (which now potentially have unsatisfied H-bond donors or acceptors)
- Charged residues around the binding pocket that likely acted to create local charge balance with the direct 2'-specificity determining residues.

Below this output is printed the verbose log file (if requested) and the options used for generation of the library. If you provide your email address the full output will be emailed to you.

A/T/G/C	As normal	N	A C G T	B	C G T
Y	C T	R	A G	D	A G T
S	G C	W	A T	H	A C T
K	T G	M	A C	V	A C G

Table 1: The mixed base pair notations used by CSR-SALAD for the generation of degenerate codons, showing the base pairs included. Oligonucleotides containing mixed bases can be purchased from most oligonucleotide suppliers.

3. How to use CSR-SALAD results in the lab

This section is not intended to be a comprehensive methodological tutorial, but to give an introduction to the relevant experimental techniques. While the methods described herein are those used by the Arnold Group, any other methods for the high-throughput screening of mutant libraries should be compatible with CSR-SALAD.

3.1. Establishing protein expression and assay

Before any engineering can take place, it is essential to have robust protocols established for the wild-type enzyme. We generally express the proteins in BL-21 *E. coli* with a pET vector and a C-terminal His_6-tag. Sufficient expression is necessary that activity can be detected in the lysate of sub-mL cultures grown in 96-well plates, and the assay should give clear and consistent linear signal on the natively-preferred cofactor if not on the other. Expression can be optimized by media selection, expression time and temperature, and lysis conditions. Assay conditions can be optimized by varying lysate volume, cofactor and substrate concentrations, and buffer components.

3.2. Cloning libraries

Libraries can be cloned using any desired method, including SOE-PCR, exponential quick-change, or round-the-horn PCR. When sites of the switching libraries cannot be cloned in a single step, it's best to clone the smaller sub-library first, and then use it as a template to clone the full library, rather than the other way around. Also, be careful when designing primers for recovery libraries to include any mutations from the switching library to avoid accidentally reverting these to wild type.

After cloning libraries, transform and plate on agar at a density that yields distinct single colonies.

3.3. Screening libraries

Screening of libraries is best accomplished in multi-well plates, such as 96-well blocks. Pick individual colonies into single wells containing 300 µL of culture media using toothpicks or pipette tips and grow overnight with shaking to produce saturated starter cultures. Transfer 50 µL of this starter culture to 600 µL of fresh media in a new plate for expression, and store the rest of the starter plate at 4 °C for up to a week. The expression plate should be grown and induced in whatever fashion is optimal for your protein, then harvested by centrifugation, and the supernatant discarded. Pellets can be stored at -20 °C for at least a month.

Cell pellets are lysed with 200-300 µL of a buffered solution containing 2 mM $MgCl_2$, 750 mg/L lysozyme, and 10 mg/L DNaseI. One hour at 37 °C can be used for most proteins, although some respond better to lower temperatures and higher lysozyme concentrations or longer lysis durations. Alternately, an extraction reagent such as BugBuster or SoluLyse can be added to speed lysis at room temperature, although some proteins are inhibited by the detergents.

After lysis, lysates should be clarified by centrifugation and the supernatant transferred to a 96-well assay plate, where it is combined with the assay buffer (containing substrate and cofactor) and the reaction progress is monitored by the change in absorbance at 340 nm on a plate reader, or by the change in fluorescence if the plate reader is equipped for that (excitation 340 nm, emission 440 nm). Starting cofactor and substrate concentrations should be near the natural K_Ms of the enzyme. Most reactions will go to completion within 5 minutes, although some may be faster or slower. Faster reactions can be slowed by serial dilutions of the lysate.

The enzyme variants with the best rates on the desired cofactor should be rescreened before additional characterization is performed. From the starter plates, streak agar plates with the mutants of interest and pick new starter plates with multiple colonies from each 'hit'. These plates should be expressed and assayed as before to ensure that cultures were monoclonal and that activities are reproducible. DNA can then be purified from the best hits for sequencing and subsequent mutagenesis.

3.4. Kinetic assays

Protocols for kinetic assays are widely available; here we provide only a short summary of our methods. After scaled-up (0.2-2 L, depending on expression levels) growth of *E. coli* and expression of proteins, cell pellets are frozen and thawed before resuspension in His-Trap buffer A. Lysis by sonication or detergent solubilization is followed by

centrifugal clarification, and then the protein is purified on an FPLC with a Ni-NTA column, eluting with an imidazole gradient. If the protein is to be kept overnight or frozen for later characterization, it should subsequently be buffer exchanged into an imidazole-free buffer by dialysis or centrifugal concentration and redilution, but we prefer to do kinetic characterization immediately.

For the determination of cofactor K_M, protein can be mixed with an assay buffer containing substrate and with cofactor, and the progress of the reaction is monitored. We use fluorescence (excitation 340 nm, emission 440 nm) for kinetics because it is more sensitive and therefore allows for the detection of activity at lower cofactor concentrations which would be undetectable by absorbance alone. By varying cofactor concentration and keeping substrate and protein concentration constant (substrate concentration should be saturating), sufficient data can be collected to fit a Michaelis-Menten or Hill equation and solve for K_M (or K_H). Substrate affinity can be determined similarly, using constant saturating cofactor and varying the substrate concentration. However, because fluorescence detection is not precisely linear with concentration, we prefer to use a separate experiment to determine k_{cat}, where the rate of cofactor consumption (in the saturated range) is measured using absorbance, and compared to a calibration curve to determine precise activity. Protein concentration can be determined using a Bradford colorimetric or other assay.

While for many applications only in-cell activity is relevant, it is important to characterize kinetics in a more thorough fashion. In part, this can important for teasing apart the contributions of k_{cat} and K_M, but also because understanding activity in a protein-expression-independent context can allow for an understanding of the expression effects of mutations. Some mutants, which appear significantly improved in lysate can show small or even nonexistent improvements in kinetic properties if they boost expression levels, while other seemingly small improvements can become much larger when decreased protein yields are considered. Understanding the relative contributions of expression, affinity, and activity can be useful in planning further experiments or in the optimization of applied setups to maximize substrate conversion.

4. How CSR-SALAD works

4.1. Web front-end

4.1.1. Implementation

The website is hosted on the cheme.caltech.edu server, which is managed by the Department of Chemistry and Chemical Engineering at the California Institute of Technology. The CSR-SALAD webpage is implemented primarily in HTML 4.01. A PHP script parses user input and options and executes the CSR-SALAD algorithm, which is implemented as a Python script (discussed further in 4.2).

4.1.2. Privacy and security

Files and user input are retrieved using the HTML POST function and held as temporary variables. All information from a session is deleted upon completion of the execution of the script. As such, user-uploaded pdb files and email addresses are never recorded or saved on the server.

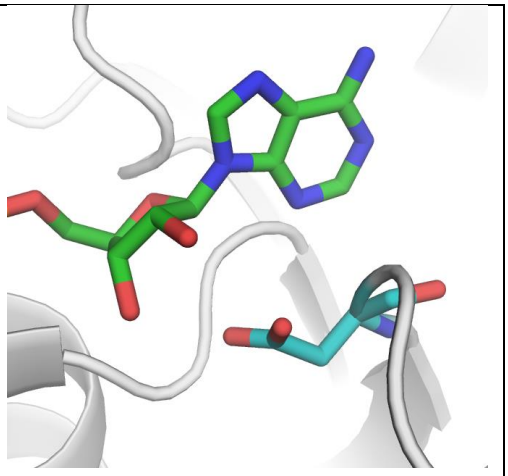
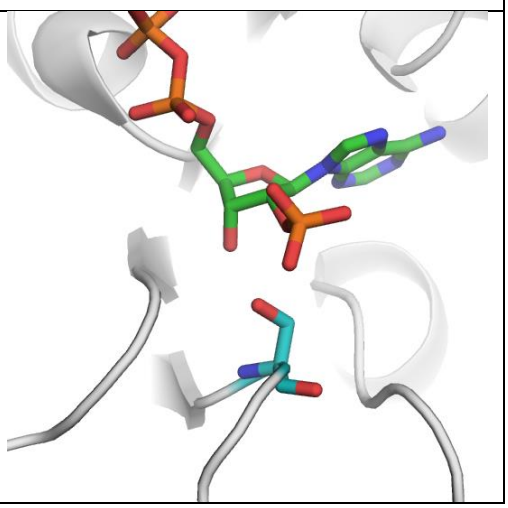
4.2. Python back-end

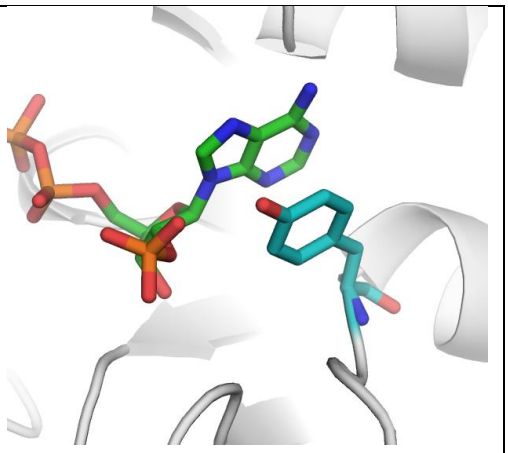
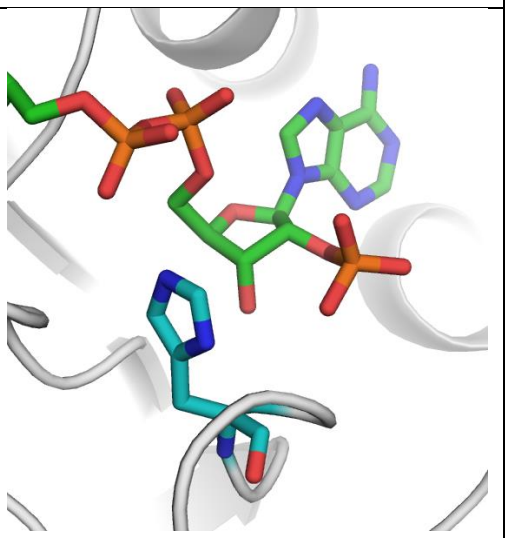
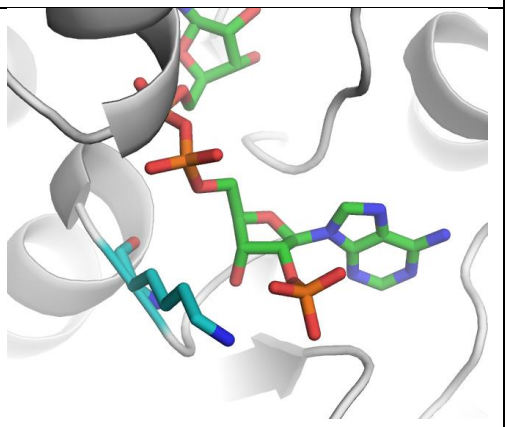
4.2.1. Implementation

CSR-SALAD is implemented in Python 2.7.x, and relies heavily on the Biopython PDB module (http://biopython.org/wiki/The_Biopython_Structural_Bioinformatics_FAQ) for structural analysis. The numpy, itertools, and operator modules are also used.

4.2.2. Structural analysis

Once the structure has been uploaded and parsed with Bio.PDB, the first step is to select and classify the residues. Before diving into the detailed geometric definitions, let's overview the classes of residues involved. Table 2 contains the basic residue classifications used by CSR-SALAD. Several of them have multiple names, based on differential geometry definitions, which are more useful for diagnostics than as indicators of structural function.

Edge (aka Floor)	These residues lie along the edge of the adenine residue and extend parallel to it toward the ribose moiety, frequently making contact with C2A, N3A, and C4A of the cofactor. In NADP-binding structures, these often do not bind the phosphate, but can be mutated to make interactions with the O2' hydroxyl.	
Bidentate	In addition to making contact with the O2' hydroxyl or phosphate, these residues also interact with the O3' hydroxyl.	

Face (aka R-chain, Offset face, Ring-binder)	These residues make contact with the plane of the adenine moiety, and often also interact with the O2' moiety.	
Pyrophosphate	These residues, in addition to making contact with the O2' moiety, also interact with the pyrophosphate moiety that bridges the adenosine and nicotinamide ends of the molecule. As such, these residues are not generally included for mutagenesis because they play a significant role in the binding energetics of the cofactor as a whole	
Motif	These residues are the Xs in the GX(X)GX(X)[G/A] motif common to Rossmann folds. The glycines of this motif provide hydrogen bonds to the pyrophosphate, but the other amino acids are sometimes involved in binding other elements of the NAD(P) molecule.	

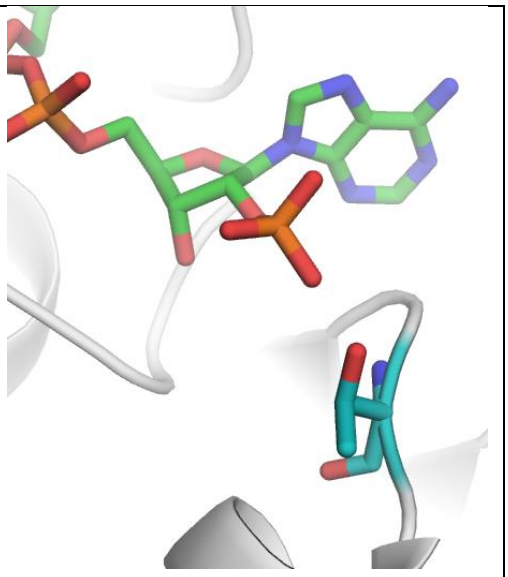
Other (includes Simple, Nonsimple, Simple*, and Peripheral	All other residues which do not fit into these categories are given one of these other designations, which together contain a plurality of all specificity-determining residues. In our analysis, no specific constellations of residues exist within these classes capable of being defined as a distinct class.	

Table 2: Summary of the residue classifications used by CSR-SALAD.

4.2.2.1. Residue identification

Identification of specificity-determining residues works by first using an expansive distance-based definition to identify potential residues and then paring back residues which fail to meet certain geometric criteria.

Potential phosphate-binding residues are those which have atoms within 4.2 Å of certain cofactor atoms. For NADP, these are the atoms of the phosphate, {O2B, P2B, O1X, O2X, O3X}. For NAD, this set is modified slightly to find residues that might be mutated to bind the phosphate but which currently do not interact with the 2'-hydroxyl. That set is {O2B, O3B, C4A, N3A}.

Atoms within this distance are considered individually before being considered as members of an amino acid. Backbone residues are not considered, since they do not change upon mutation. Also, any atom not belonging to an amino acid is excluded, since they similarly cannot be mutated to alter selectivity. Hydrogen atoms are also not considered, because they are not present in most structures and the geometry definitions have not been developed with them in mind.

Because the list of NAD atoms from which potential residues are found contains residues on the adenine ring, it is important to only consider residues that can be mutated to interact with the 2' moiety. Specifically, residues must be on the same face of the adenine as the hydroxyls of the ribose (see figure 2). For that reason, atoms are excluded if they are 0.75 Å (or more) closer to O4B than to O2B.

After this point, the atoms are considered on the basis of the residues to which they belong. Glycines in GX(X)GX(X)[G/A] motifs are excluded, as are the other amino acids if that option has been selected. Other glycines are excluded if their N is closer to the cofactor atom than their CA, which would indicate that mutations would produce side chains facing away from the cofactor.

To check whether other residues point towards the cofactor, it is necessary to introduce the concept of the side-chain pseudocenter, which is adapted from Bahar and Jernigan, 1996. This is the geometric centroid, in 3D space, of the locations of a set of key atoms in the side-chain. These key atoms are listed in Table 3.

Gly	{CA}
Ala	{CB}
Val	{CG1, CG2}
Ile	{CG1, CG2, CD1}
Leu	{CD1, CD2, CG, CB}
Ser	{OG}
Thr	{OG1, CG2}
Asp	{OD1, OD2}
Asn	{OD1, ND2}
Glu	{OE1, OE2, CG}
Gln	{OE1, NE2}
Lys	{NZ, CE}
Arg	{NE, NH1, NH2}
Cys	{SG}
Met/Mse	{CE, SD/SE, CG}
Phe	{CG, CD1, CD2, CE1, CE2, CZ}
Tyr	{CE1, CE2, CZ, OH}
Trp	{CG, CD1, CD2, NE1, CE2, CE3, CZ2, CZ3}
His	{CG, ND1, CD2, CE1, NE2}
Pro	{CB, CG, CD}

Table 3: Key atoms of residue side chains used for pseudocenter calculation.

Once the side-chain pseudocenter is determined the angle is measured from P2B for NADP or C2B for NAD to the CA to the pseudocenter (Figure 3). If this angle exceeds 100° for NAD or 90° for NADP, the residue is excluded.

If two or fewer residues are identified for a structure and peripheral residues are not manually excluded, this process is repeated with an expanded cutoff distance of 5.2 Å, and additional residues are added to the library design process.

4.2.2.2. Residue classification

Multiple criteria are considered when assigning a residue to one of the structural classes. The workflow of this classification is described below.

The first step of residue classification is to, for each residue, get the ‘recontacts’. These are defined as atoms of the cofactor within 4.3 Å (or 5.3 Å for residues selected in the second round) of any of the pseudocenter-determining atoms shown in Table 3.

For structures with multiple chains, only the equivalent residue with an atom closest to the nearest cofactor atom is considered for structure determination. However, if

recontacts to certain atoms of the cofactor (O1A, O2A, O1N, O2N, O3, C5A,O3B) are present for any equivalent amino acid, they will be included even if the closest residue does not possess these recontacts. This allows for certain residue classifications to over-ride the rotamer of the closest equivalent amino acid if they occur anywhere in the crystal structure.

The residue classifications are made on the basis of the following dichotomous key:

1a) Residue is in Rossmann fingerprint motif……..Motif
1b) Else……..Go to 2
2a) Residue has no recontacts……Go to 3
2b) Else……..Go to 4
3a) Residue side chain contains oxygen or nitrogen……..…Peripheral
3b) Else…….Exclude residue
4a) Residue has no recontacts except 2' phosphate or hydroxyl……...Go to 5
4b) Else……..Go to 7
5a) Residue is arginine and has all side-chain atoms…….Go to 8
5b) Else……...Simple
6a) Arginine CD atom within 4.2 Å of cofactor C4A and/or C5A……Face
6b) Else……...Simple
7a) Residue recontacts C5A and at least one of {C4A, N3A, C2A, N1A, C6A, C8A, N7A}…….Face
7b) Else……..Go to 8
8a) Residue recontacts at least one of {N9A, C4A, N3A, C2A, N1A, C6A, C8A, N7A}……Go to 9
8b) Else……Go to 19
9) Compute $\langle i,j,k \rangle$ for residue (see below)…….....Go to 10
10a) $|k| > 2.5$ Å and $(|i| + |j|) < 2.5$…….....Face
10b) Else……..Go to 11
11a) $|i| < 3$ Å and -2 Å $< j < 3$ Å and $\|\langle i,j,k \rangle\| < 5.35$ Å……Face
11b) Else……..Go to 12
12a) Residue is not glycine and $|k| < 3.2$ Å and $|k| < |j|$ and $|i| < 5$ Å……..Go to 13
12b) Else……..Go to 15
13) Compute $\langle l,m,n \rangle$ for residue (see below)…….Go to 14
14a) $|n| < \max(|l|,|m|)$ and $m < 0$……Edge
14b) Else……..Go to 15
15a) Residue is in [Arg, Gln, Tyr, Phe, Glu, Trp, His, Pro]……Go to 16
15b) Else……..Go to 19
16) Compute normal vector $\boldsymbol{v}$ to residue sidechain (see below)……..Go to 17
17) Compute angle θ between $\boldsymbol{v}$ and normal vector to adenine……...Go to 18
18a) $\theta < 20°$ or $\theta > 160°$.……Face
18b) Else……..Go to 19
19a) Residue recontacts oxygen atoms of the diphosphate……Pyrophosphate
19b) Else……..Go to 20
20a) Residue recontacts O3B……Bidentate
20b) Else……..Go to 21
21a) Cofactor is NAD……..Go to 22
21b) Else……..Go to 23

22a) O2B not within 4.4 Å of any pseudocenter-determining atom…….Peripheral
22b) Else……..Go to 23
23) All other residues……Simple

Three of the steps above (9, 13, and 16) require further explication. These steps are involved in determining the position of the residue not merely in terms of its distance from atoms of the cofactor but in terms of its orientation in 3-D space. Because the cofactors are arbitrarily oriented in *xyz*-space, we must first establish a uniform coordinate space in which to work. To do this, we create a new basis set <*x'*,*y'*,*z'*> composed of three vectors based on adenine atoms (Figure 2). The *x'* component is defined by the vector from C4A to N3A, and *y'* is defined by the vector from C4A to C5A. The *z'* coordinate is orthogonal to these, defined by the cross-product *x'* × *y'*.

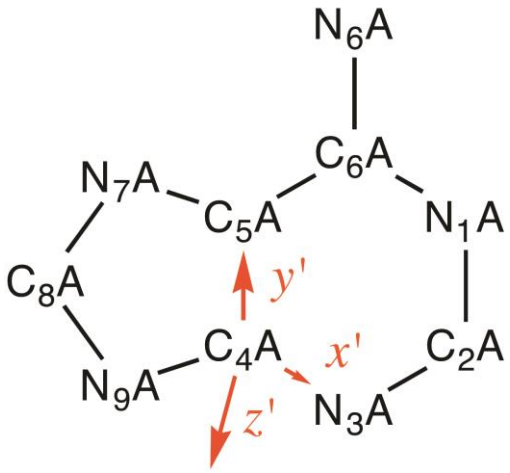


Fig 2. The components of *x'y'z'*-space.

This set of vectors forms an alternate coordinate space in which the positions of atoms or pseudocenters can be discussed, particularly relative to the adenine moiety itself. Two such positions are required for residue classification. The first vector, <*i*,*j*,*k*>, is the vector going from the cofactor C4A atom to the residue pseudocenter. The second, <*l*,*m*,*n*> is the vector from C4A to the alpha carbon of the residue. Both of these vectors are expressed in the adenine-centered *x'y'z'*-space.

For step 17, a vector ***v*** normal to a residues' side-chain must be calculated. Unlike <*i*,*j*,*k*> and <*l*,*m*,*n*>, ***v*** is in *xyz*-space, and is compared to the *z'* vector calculated above to determine if the residue side chain is parallel to the plane of the adenine. The vector ***v*** is calculated as the cross product of two other vectors, ***t*** and ***u***, as defined by Table 4.

Residue	***t***	***u***
Arg	NE - NH1	NE - NH2
Gln	OE1 - NE2	OE1 - CD
Tyr	CE1 - CE2	CE1 - CZ
Phe	CG - CD1	CG - CD2
Glu	OE1 - OE2	OE1 - CG
Trp	CG - CD1	CG - CD2
His	CG - ND1	CG - ND2
Pro	CB - CD	CB - N

Table 4: Atom pairs used to calculate vectors ***t*** and ***u*** and thereby $\boldsymbol{v} \equiv \boldsymbol{t} \times \boldsymbol{u}$.

If one or more face residues are identified but no edge residue is, the two amino acids prior to the face residue(s) in primary sequence are checked as potential edge residues using a loosed criteria, due to the frequency of this sequence-structure motif in Rossmann fold proteins. For each of these residues that are not glycine, three criteria are checked:

1) The residue pseudocenter is closer to cofactor atom O2B than the residue alpha carbon.
2) The angle between the vector ***r***, running from the residue alpha carbon to the pseudocenter, and the bottom edge of the adenine (N3A to N9A) is less than 45°.
3) The angle between the ***r*** and *z'*, as defined above, is between 70° and 110°.

If the residue meets all three criteria, its classification is replaced with 'edge' if it had previously been classified as anything other than face, or it is added as an edge residue if it had escaped inclusion in 4.2.2.1.

4.2.3. Library design

4.2.3.1. Amino acid selection

Once residues have been classified, the next step is to select the desirable amino acids for mutation and screening. For each residue and each classification, we have selected degenerate codons in a range of sizes that possess both the wild-type amino acid and the amino acids we believe are most likely to provide productive interactions with the switched cofactor. These amino acids are selected and prioritized subjectively and heuristically on the basis of three factors: (1) which mutations to residues like these have previously been shown to be useful in cofactor switching, (2) which residues does structural intuition suggest could provide useful contacts in stabilizing the desired cofactor, and (3) which amino acids exist at corresponding positions in the proteins that naturally bind the desired amino acid.

4.2.3.2. Codon selection

The degenerate codons used by CSR-SALAD are provided in Tables 5 and 6. Codons are optimized for expression in *E. coli* using the Automated Nucleotide Tool (ANT). When two codons of equal size but which coded for different numbers of amino acids were available, the codon which coded for greater diversity at the amino acid level was preferred.

	Edge	**Bidentate/ Motif/ Pyrophosphate**	**Face**	**Simple/ Peripheral**
Phe	KWC:DFVY KWK:DEFLVY*	KWC:DFVY KWK:DEFLVY* BWK:DEFHLQVY	---:F TWC:FY KWC:DFVY YWK:FHLQY	KWC:DFVY KWK:DEFLVY*
Leu	SWA:ELQV	SWA:ELQV SWK:DEHLQV	YWC:FHLY WYK:FILMST	SWA:ELQV SWK:DEHLQV
Ile	RWA:EIKV	RWA:EIKV	WWC:FINY HWC:FHILNY	RWA:EIKV
Met	RWG:EKMV	RWG:EKMV	WTK:FILM HYK:FILMPST	RWG:EKMV
Val	GWK:DEV SWA:ELQV RWK:DEIKMNV	GWK:DEV	SWC:DHLV VWC:DHILNV	GWK:DEV
Ser	ADC:INS RDC:DGINSV	RRC:DGNS RVC:ADGNST	RDC:DGINSV	RRC:DGNS VDC:DGHILNRSV
Pro	---:P SMC:ADHP	SMC:ADHP	MYA:ILPT	SMC:ADHP
Thr	RMC:ADNT RMK:ADEKNT	RMC:ADNT RMK:ADEKNT	AHC:INT MMC:HNPT ANK:IKMNRST	RMC:ADNT RMK:ADEKNT
Ala	GMC:AD RMC:ADNT	KCA:AS RVC:ADGNST	VCA:APT VYA:AILPTV	KCA:AS RVC:ADGNST
Tyr	WWC:FINY HWC:FHILNVY	WWC:FINY DWC:DFINVY	TWC:FY YWC:FHLY DHC:ADFINSTVY	DAK:DEKNY* DWK:DEFIKLMNVY*
His	VAC:DHN VAK:DEHKNQ	VAC:DHN VAK:DEHKNQ	MMC:HNPT	VAC:DHN VAK:DEHKNQ
Gln	---:Q SAA:EQ	---:Q SAA:EQ	CHA:LPQ	---:Q SAA:EQ
Lys	RAA:EK VAA:EKQ RWA:EIKV	RRK:DEGKNRS	MWG:KLMQ MHG:KLMPQT	RAK:DEKN RWK:DEIKMNV DWK:DEFIKLMNVY*

Asp	---:D GAK:DE SAK:DEHQ	---:D GAK:DE SAK:DEHQ	---:D GAK:DE SAK:DEHQ	---:D GAK:DE SAK:DEHQ
Glu	---:E VAA:EKQ	---:E VAA:EKQ	---:E VAA:EKQ	---:E VAA:EKQ
Cys	KRC:CDGY KDC:CDFGVY	KRC:CDGY	KKC:CFGV KDC:CDFGVY	KRC:CDGY
Trp	WKG:LMRW YDG:LQRW*	YRG:QRW* YRK:CHQRWY*	TKK:CFLW	YRG:QRW* YRK:CHQRWY*
Arg	SRA:EQR SDA:EGLQRV	SRK:DEGHQR	YVC:CHPRSY YVC:CHNPRSTY HNC:CFHILNPRSTY	SRK:DEGHQR RRK:DEGKNRS VRK:DEGHKNQRS
Gly	GRC:DG GDC:DGV	GRC:DG GNC:ADGV	RDC:DGINSV	GRC:DG GNC:ADGV RRK:DEGKNRS
Asn	---:N RAK:DEKN	RAC:DN VAK:DEHKNQ	WWC:FINY	RAC:DN RMC:ADNT RMK:ADEKNT

Table 5: Codons used for cofactor switching NADP-dependent proteins to NAD.

	Edge	**Bidentate/ Motif/ Pyrophosphate**	**Face**	**Simple/ Peripheral**
Phe	---:F WWK:FIKLMNY*	TYC:FS YHC:FHLPSY	YKC:CFLR YDC:CFHLRY	WHC:FINSTY
Leu	VWA:EIKLQV	TYA:LS WYA:ILST	CKA:LR CDC:HLR CDK:HLQR	CKA:LR MDA:IKLQR
Ile	AWK:IKMN	ADK:IKMNRS	AKA:IR ANA:IKRT	ADL:IKMNRS
Met	AWG:KM AWK:IKMN	AWG:KM ADK:IKMNRS	AKG:MR	AWG:KM ADK:IKMNRS
Val	RYA:AITV RHA:AEIKTV	RKC:GISV VDC:DGHILNRSV	RKG:GMRV RBG:AGMRTV	RKC:GISV VDC:DGHILNRSV
Ser	KCA:AS KYA:ALSV	RGC:GS RSC:AGST	RGC:GS RSC:AGST	---:S RGC:GS RSC:AGST
Pro	CYA:LP SYA:ALPV	YCA:PS HCA:PST NCA:APST	MSC:PRST	YCA:PS HCA:PST NCA:APST
Thr	RCA:AT RMC:ADNT	DCA:AST	RSA:AGRT VVC:ADGHNPRST	DCA:AST

Ala	---:A KCA:AS KYA:ALSV	RSC:AGST	RSA:AGRT	RSC:AGST
Tyr	WWK:FIKLMNY*	HAC:HNY HAK:HKNQY*	CHRY:YRC	HAC:HNY HAK:HKNQY*
His	MWC:HILN	MRC:HNRS	MRC:HNRS	MMC:HNPT MVC:HNPRST
Gln	MWA:IKLQ	MRA:KQR MVA:KPQRT	MVA:KPQRT	MRK:HKNQRS
Lys	---:K AAK:KN	AAK:KN AMK:KNT AVK:KNRST	ARK:KNRS MRK:HKNQRS	AAK:KN AMK:KNT AVK:KNRST
Asp	RVC:ADGNST RNC:ADGINSTV	RMC:ADNT RVC:ADGNST	RMC:ADNT VMC:ADHNPT	RRC:DGNS RVC:ADGNST
Glu	KMA:AES* KHA:AELSV* DHA:AEIKLSTV*	VRA:EGKQR VRW:DEGHKNQRS	SRA:EGQR VVA:AEGKPQRT	DMA:AEKST* DMK:ADEKNSTY*
Cys	KSC:ACGS KBC:ACFGSV	KSC:ACGS	YRC:CHRY HVC:CHNPRSTY	KSC:ACGS
Trp	WKG:LMRW	WVG:KRSTW*	YRG:QRW* YRK:CHQRWY*	WVG:KRSTW*
Arg	---:R CKA:LR YDC:CFHLRY	ARK:KNRS MRK:HKNQRS	---:R CRA:QR CNA:LPQR	ARK:KNRS MRK:HKNQRS
Gly	---:G GBA:AGV	---:G RSC:AGST	---:G RSA:AGRT	---:G RSC:AGST
Asn	ARC:NS AWK:IKMN	ARC:NS MRK:HKNQRS	AVK:KNRST	ARC:NS MRK:HKNQRS

Table 6: Codons used for cofactor switching NAD-dependent proteins to NADP.

4.2.3.3. Library design

Designing a library from the available sets of codons is dependent on the maximum library size provided by the user. As a first step, CSR-SALAD calculates the minimum possible library size for the residues selected. If this is above the provided maximum, the option to not mutate certain residues is added based on the frequency of those amino acids in the binding pockets of enzymes with the same cofactor preference. That is, CSR-SALAD steps through the following priority lists and adds to all amino acids of the indicated identity a '---' non-mutation option, and amino acids rarely used for NAD(P) binding pockets are more likely to be spared from mutation than those commonly used to make up binding pockets.

NADP: W<C<M<F<P<I<L<E<D<V<Q<H<A<Y<G<N<T<K<R<S
NAD: C<W<M<Y<H<F<T<Q<A<P<G<R<S<N<L<V<E<K<I<D

Once the minimum possible library size is below the selected maximum library size, selection of a specific library can begin. CSR-SALAD starts with the library that has the largest codon at each position, then steps through the residue classifications in the order [Motif, Pyrophosphate, Peripheral, Edge, Face, Simple, Bidentate]. For each residue classification, it moves through the residues of the library in primary sequence order and decreases each to the next largest. After each codon reduction, the library size is checked and if it is below the provided maximum, that library is selected, and the whole process is iterated until that threshold is reached.

4.2.4. Activity recovery

Cofactor switched enzymes – or any enzymes with multiple mutations away from wild type – often lose significant amounts of activity. To recover this, compensatory mutations must be found, and we have empirically identified 'hot-spots' in the protein structure which seem to be enriched in compensatory mutations. CSR-SALAD locates and indicates these hotspots (five classes, sorted into three priority levels) and provides them to users. We recommend performing site-saturation mutagenesis (i.e. with an NNK codon) at each of these positions – either serially or in parallel followed by recombination or re-randomization – and screening for improved activities.

4.2.4.1. Classes of recovery hot-spots

CSR-SALAD identified five classes of recovery hotspots, which are sorted into three priority levels, as laid out in 2.5. In this section, we will delve into each in more detail.

The high priority level is comprised of residues which would have been mutated if the library size was sufficient, but which were excluded from mutagenesis to keep library sizes tractable. These include any residues for which the '---' non-mutation codon was recommended and also residues from the motif, pyrophosphate-binding, or peripheral categories that were excluded on the basis of the advanced options. It may also be valuable to perform site-saturation mutagenesis on residues that were mutated in the library but kept their original identity in the best mutant.

The next priority level contains residues around the adenine moiety of the cofactor, which we have previously shown to exert considerable control over cofactor binding and positioning. We have modified the definition somewhat from that used in Cahn et al., defining these residues as those within 4.1 Å of N7A, C5A, N6A, or N1A (the top edge of the adenine moiety) but not included in the residues identified in 4.2.2.1.

The final set of residues includes those which do not directly interact with the cofactor. The first class of these is residues that hydrogen bond to residues identified in 4.2.2.1. Mutation of the latter would leave these residues with unsatisfied hydrogen bond donors or acceptors, which can be destabilizing or can lead to unfavorable protein conformations. These residues are identified on the basis of having a side-chain oxygen or nitrogen atom within 3.8 Å of a side-chain oxygen or nitrogen of one of the residues used for cofactor switching. The second class includes charged residues around the adenosine binding pocket (within 8.2 Å of the 2' phosphate or hydroxyl) that putatively contribute to the charge balance of the pocket. That is, for NADP-bound structures, these

would be positively-charged lysines, arginines, or histidines, and for NAD-bound structures the negatively-charged aspartates and glutamates. These residues are mutated in several early cofactor switching papers – generally on the basis of missing structural information or incorrect homology models suggesting they played a more direct role – but in some instances were found to be beneficial, and so are identified by CSR-SALAD.

5. References

Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S., and Arnold, F.H. (2016) Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Engineering Design and Selection* **29**(1), 31-38.

Cornish-Bowden, A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Research* **13**, 3021-3030

Bahar, I., and Jernigan, R.L. (1996) Coordination geometry of nonbonded residues in globular proteins. *Folding and Design* **1**, 357-370.

Supporting Material 2 | Citation information for **Supporting Tables 1** and **2**.

Ashida, H., et al. (2004), 'Conversion of cofactor specificities of alanine dehydrogenases by site-directed mutagenesis', *Journal of Molecular Catalysis B: Enzymatic*, 30 (3–4), 173-76.

Banta, S., et al. (2002), 'Alteration of the specificity of the cofactor-binding pocket of Corynebacterium 2,5-diketo-D-gluconic acid reductase A', *Protein Engineering*, 15 (2), 131-40.

Baroni, S., et al. (2012), 'A single tyrosine hydroxyl group almost entirely controls the NADPH specificity of *Plasmodium falciparum* ferredoxin-NADP+ reductase', *Biochemistry*, 51 (18), 3819-26.

Bastian, S., et al. (2011), 'Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*', *Metabolic Engineering*, 13 (3), 345-52.

Bernard, N., et al. (1995), 'D175 Discriminates Between NADH and NADPH in the Coenzyme Binding Site of *Lactobacillus delbrueckii* subsp. Bulgaricus D-Lactate Dehydrogenase', *Biochemical and Biophysical Research Communications*, 208 (3), 895-900.

Bocanegra, J. A., Scrutton, N. S., and Perham, R. N. (1993), 'Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering', *Biochemistry*, 32 (11), 2737-40.

Brinkmann-Chen, S., et al. (2013), 'General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH', *Proceedings of the National Academy of Sciences*, 110 (27), 10946-51.

Bubner, P., Klimacek, M., and Nidetzky, B. (2008), 'Structure-guided engineering of the coenzyme specificity of *Pseudomonas fluorescens* mannitol 2-dehydrogenase to enable efficient utilization of NAD(H) and NADP(H)', *FEBS Letters*, 582 (2), 233-37.

Capone, M., et al. (2011), 'Re-engineering the discrimination between the oxidized coenzymes NAD(+) and NADP(+) in clostridial glutamate dehydrogenase and a thorough reappraisal of the coenzyme specificity of the wild-type enzyme', *FEBS Journal*, 278 (14), 2460-68.

Chen, R., Greer, A., and Dean, A. M. (1995), 'A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity', *Proceedings of the National Academy of Sciences*, 92 (25), 11666-70.

Chen, R., Greer, A., and Dean, A. M. (1996), 'Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase', *Proceedings of the National Academy of Sciences*, 93 (22), 12171-76.

Clermont, S., et al. (1993), 'Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase - Role of the acidic residue in the fingerprint region of the nucleotide-binding fold', *Biochemistry*, 32 (38), 10178-84.

Cui, Dongbing, et al. (2015), 'A computational strategy for altering an enzyme in its cofactor preference to NAD(H) and/or NADP(H)', *FEBS Journal*, 282 (12), 2339-51.

Dambe, Tresfore R., et al. (2006), 'Crystal structure of NADP(H)-dependent 1,5-anhydro-D-fructose reductase from *Sinorhizobium morelense* at 2.2 angstrom resolution: Construction of a NADH-accepting mutant and its application in rare sugar synthesis', *Biochemistry*, 45 (33), 10030-42.

Döhr, Olaf, et al. (2001), 'Engineering of a functional human NADH-dependent cytochrome P450 system', *Proceedings of the National Academy of Sciences*, 98 (1), 81-86.

Ehrensberger, A. H., Elling, R. A., and Wilson, D. K. (2006), 'Structure-Guided Engineering of Xylitol Dehydrogenase Cosubstrate Specificity', *Structure*, 14 (3), 567-75.

Ehsani, M., et al. (2009), 'Reversal of coenzyme specificity of 2,3-butanediol dehydrogenase from *Saccharomyces cerevisae* and *in vivo* functional analysis', *Biotechnology and Bioengineering*, 104 (2), 381-89.

Elmore, C. L. and Porter, T. D. (2002), 'Modification of the nucleotide cofactor-binding site of cytochrome P-450 reductase to enhance turnover with NADH *in vivo*', *Journal of Biological Chemistry*, 277 (50), 48960-64.

Eppink, M. H. M., et al. (1999), 'Switch of coenzyme specificity of p-hydroxybenzoate hydroxylase', *Journal of Molecular Biology*, 292 (1), 87-96.

Fasan, R., et al. (2011), 'Improved product-per-glucose yields in P450-dependent propane biotransformations using engineered *Escherichia coli*', *Biotechnology and Bioengineering*, 108 (3), 500-10.

Feeney, R., Clarke, A. R., and Holbrook, J. J. (1990), 'A singe amino-acid substitution in lactate dehydrogenase imporves the catalytic efficiency wth an alternative coenzyme', *Biochemical and Biophysical Research Communications*, 166 (2), 667-72.

Friesen, J. A., et al. (1996), 'Structural determinants of nucleotide coenzyme specificity in the distinctive dinucleotide binding fold of HMG-CoA reductase from *Pseudomonas mevalonii*', *Biochemistry*, 35 (37), 11945-50.

Galkin, A, et al. (1997), 'Construction of a new leucine dehydrogenase with preferred specificity for $NADP^+$ by site-directed mutagenesis of the strictly NAD^+-specific enzyme', *Protein Engineering*, 10 (6), 687-90.

Gand M, et al. (2016), 'A NADH-accepting imine reductase variant: Immobilization and cofactor regeneration by oxidative deamination', *Journal of Biotechnology*, 230, 11-18.

Gul-Karaguler, N., et al. (2001), 'A single mutation in the NAD-specific formate dehydrogenase from *Candida methylica* allows the enzyme to use NADP', *Biotechnology Letters*, 23 (4), 283-87.

Hoelsch, K., et al. (2013), 'Engineering of formate dehydrogenase: Synergistic effect of mutations affecting cofactor specificity and chemical stability', *Applied Microbiology and Biotechnology*, 97 (6), 2473-81.

Holmberg, N., Ryde, U., and Bülow, L. (1999), 'Redesign of the coenzyme specificity in L-lactate dehydrogenase from *Bacillus stearothermophilus* using site-directed mutagenesis and media engineering', *Protein Engineering*, 12 (10), 851-56.

Hong, S-H., et al. (2016), 'Alternative biotransformation of retinal to retinoic acid or retinol with cofactor switch by aldehyde dehydrogenase from *Bacillus cereus*', *Applied and Environmental Microbiology*, Epub ahead of print.

Hsieh, J.-Y, et al. (2006), 'Determinants of the dual cofactor specificity and substrate cooperativity of the human mitochondrial $NAD(P)^+$-dependent malic enzyme: Functional roles of glutamine 362', *Journal of Biological Chemistry*, 281 (32), 23237-45.

Jensen, C. N., et al. (2013), 'Mutations of an NAD (P) H-dependent flavoprotein monooxygenase that influence cofactor promiscuity and enantioselectivity', *FEBS Open Bio*, 3, 473-78.

Kamerbeek, N. M., Fraaije, M. W., and Janssen, D. B. (2004), 'Identifying determinants of NADPH specificity in Baeyer-Villiger monooxygenases', *European Journal of Biochemistry*, 271 (11), 2107-16.

Katzberg, M., et al. (2010), 'Engineering cofactor preference of ketone reducing biocatalysts: A mutagenesis study on a γ-diketone reductase from the yeast *Saccharomyces cerevisiae* erving as an example', *International Journal of Molecular Sciences*, 11 (4), 1735-58.

Khoury, G. A., et al. (2009), 'Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity', *Protein Science*, 18 (10), 2125-38.

Kristan, K., et al. (2007), 'Rational design of novel mutants of fungal 17 β-hydroxy steroid dehydrogenase', *Journal of Biotechnology*, 129 (1), 123-30.

Liang, L., Zhang, J. Q., and Lin, Z. L. (2007), 'Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing', *Microbial Cell Factories*, 6.

Ma, C., et al. (2010), 'Relaxing the coenzyme specificity of 1,3-propanediol oxidoreductase from *Klebsiella pneumoniae* by rational design', *Journal of Biotechnology*, 146 (4), 173-78.

Maddock, D. J., Patrick, W. M., and Gerth, M. L. (2015), 'Substitutions at the cofactor phosphate-binding site of a clostridial alcohol dehydrogenase lead to unexpected changes in substrate specificity', *Protein Engineering Design and Selection*, 28 (8), 251-8.

Marohnic, C. C., Bewley, M. C., and Barber, M. J. (2003), 'Engineering and characterization of a NADPH-utilizing cytochrome b_5 reductase', *Biochemistry*, 42 (38), 11170-82.

Maurer, S. C., et al. (2005), 'Catalytic hydroxylation in biphasic systems using CYP102A1 mutants', *Advanced Synthesis & Catalysis*, 347 (7-8), 1090-98.

Medina, M., et al. (2001), 'Probing the determinants of coenzyme specificity in ferredoxin-$NADP^+$ reductase by site-directed mutagenesis', *Journal of Biological Chemistry*, 276 (15), 11902-12.

Miller, S. P., Lunzer, M., and Dean, A. M. (2006), 'Direct demonstration of an adaptive constraint', *Science*, 314 (5798), 458-61.

Nakanishi, M., et al. (1997), 'Switch of coenzyme specificity of mouse lung carbonyl reductase by substitution of threonine 38 with aspartic acid', *Journal of Biological Chemistry*, 272 (4), 2218-22.

Nishiyama, M., Birktoft, J. J., and Beppu, T. (1993), 'Alteration of coenzyme specificity of malate dehydrogenase from *Thermus flavus* by site-directed mutagenesis', *Journal of Biological Chemistry*, 268 (7), 4656-60.

Paladini, D. H., et al. (2009), 'Induced fit and equilibrium dynamics for high catalytic efficiency in ferredoxin-NADP(H) reductases', *Biochemistry*, 48 (24), 5760-68.

Pazmiño, D. E. T., et al. (2007), 'Altering the substrate specificity and enantioselectivity of phenylacetone monooxygenase by structure-inspired enzyme redesign', *Advanced Synthesis & Catalysis*, 349 (8-9), 1361-68.

Petschacher, B., et al. (2005), 'The coenzyme specificity of *Candida tenuis* xylose reductase (AKR2B5) explored by site-directed mutagenesis and X-ray crystallography', *Biochemical Journal*, 385, 75-83.

Petschacher, B., et al. (2014), 'Cofactor specificity engineering of *Streptococcus mutans* NADH oxidase 2 for $NAD(P)^+$ regeneration in biocatalytic oxidations', *Computational and Structural Biotechnology Journal*, 9.

Pick, A., et al. (2014), 'Improving the NADH-cofactor specificity of the highly active AdhZ3 and AdhZ2 from *Escherichia coli* K-12', *Journal of Biotechnology*, 189, 157-65.

Rane, M. J. and Calvo, K. C. (1997), 'Reversal of the nucleotide specificity of ketol acid reductoisomerase by site-directed mutagenesis identifies the NADPH binding site', *Archives of Biochemistry and Biophysics*, 338 (1), 83-89.

Rodriguez-Arnedo, A., et al. (2005), 'Complete reversal of coenzyme specificity of isocitrate dehydrogenase from *Haloferax volcanii*', *Protein Journal*, 24 (5), 259-66.

Rosell, A., et al. (2003), 'Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase', *Journal of Biological Chemistry*, 278 (42), 40573-80.

Schepens, I., et al. (2000) 'Inhibition of the thioredoxin-dependent activation of the NADP-malate dehydrogenase and cofactor specificity', *Journal of Biological Chemistry*, 275 (28), 20966-21001.

Scrutton, N. S., Berry, A., and Perham, R. N. (1990), 'Redesign of the coenzyme specificity of a dehydrogenase by protein engineering', *Nature*, 343 (6253), 38-43.

Serov, A. E., et al. (2002), 'Engineering of coenzyme specificity of formate dehydrogenase from *Saccharomyces cerevisiae*', *Biochemical Journal*, 367 (3), 841-47.

Shiraishi, N., et al. (1998), 'Engineering of pyridine nucleotide specificity of nitrate reductase: Mutagenesis of recombinant cytochrome b reductase fragment of *Neurospora crassa* NADPH:nitrate reductase', *Archives of Biochemistry and Biophysics*, 358 (1), 104-15.

Takase, R., et al. (2014), 'Structure-based conversion of the coenzyme requirement of a short-chain dehydrogenase/reductase involved in bacterial alginate metabolism', *Journal of Biological Chemistry*, 289 (48), 33198-214.

Watanabe, S., Kodaki, T., and Makino, K. (2005), 'Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc', *Journal of Biological Chemistry*, 280 (11), 10340-9.

Woodyer, R., van der Donk, W. A., and Zhao, H. M. (2003), 'Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design', *Biochemistry*, 42 (40), 11604-14.

Yaoi, T., et al. (1996), 'Conversion of the coenzyme specificity of isocitrate dehydrogenase by module replacement', *Journal of Biochemistry*, 119 (5), 1014-18.

Zeng, Q.-K., et al. (2009), 'Reversal of coenzyme specificity and improvement of catalytic efficiency of *Pichia stipitis* xylose reductase by rational site-directed mutagenesis', *Biotechnology Letters*, 31 (7), 1025-29.

Zhang, L., et al. (1999), 'Change of nucleotide specificity and enhancement of catalytic efficiency in single point mutants of *Vibrio harveyi* aldehyde dehydrogenase', *Biochemistry*, 38 (35), 11440-47.

Zhang, R. Z., et al. (2009), 'Ser67Asp and His68Asp substitutions in *Candida parapsilosis* carbonyl reductase alter the coenzyme specificity and enantioselectivity of ketone reduction', *Applied and Environmental Microbiology*, 75 (7), 2176-83.

Zheng, H., et al. (2013), 'Converting NAD-specific inositol dehydrogenase to an efficient NADP-selective catalyst, with a surprising twist', *Biochemistry*, 52 (34), 5876-83.

Supporting Material 3 | CSR-SALAD libraries designed for proteins used in this study. Libraries are designed using default settings and a maximum library size of 1000 in the most current version of CSR-SALAD.

***Arabidopsis thaliana* glyoxylate reductase**		
Position	Amino acids to screen	Selected amino acid
Switching Library		
Asn30	Asn	Asn
Arg31	Cys, Phe, His, Ile, Leu, Asn, Pro, Arg, Ser, Thr, Tyr	Leu
Thr32	Ala, Asp, Glu, Lys, Asn, Thr	Lys
Lys35	Asp, Glu, Gly, Lys, Asn, Arg, Ser	Asp
	Library size:	*576*
High- and Medium-Priority Recovery Libraries		
Asn30	All	Asn
Cis68	All	Arg
Ala69	All	Ala
Ser72	All	Ser
Val73	All	Val

***Saccharomyces cerevisae* cinnamyl alchohol dehydrogenase**		
Position	Amino acids to screen	Selected amino acid
Switching Library		
Ser210	Asp, Gly, His, Ile, Leu, Asn, Arg, Ser, Val	Asp
Arg211	Cis, His, Asn, Pro, Arg, Ser, Thr, Tyr	Pro
Lys215	Asp, Glu, Gly, Lys, Asn, Arg, Ser	Glu
	Library size:	*684*
High- and Medium-Priority Recovery Libraries		
Ala251	All	Ala
Ser252	All	Ser
Ser253	All	Pro
Thr255	All	Thr
Asp256	All	Asp

***Talaromyces emersonii* xylose reductase**		
Position	Amino acids to screen	Selected amino acid
Switching Library		
Lys271	Asp, Glu, Gly, Lys, Asn, Arg, Ser	Lys
Ser272	Asp, Gly, Asn, Ser	Gly
Asn273	Asp, Asn	Asp
Leu274	Leu	Leu
Arg277	Cis, His, Pro, Arg, Ser, Tyr	Tyr
	Library size:	*384*
High- and Medium-Priority Recovery Libraries		
Leu274	All	Leu
Phe217	All	Phe
Phe237	All	Phe
Ala254	All	Ala
Gln280	All	Glu
Asn281	All	Asn

***Talaromyces emersonii* xylose reductase**		
Position	Amino acids to screen	Selected amino acid
Switching Library		
Gly36	Asp, Glu, Gly, Lys, Asn, Arg, Ser	Glu
Ser38	Asp, Gly, His, Ile, Leu, Asn, Arg, Ser, Val	Asn
Ser39	Asp, Gly, Ile, Asn, Ser, Val	Gly
	Library size:	*432*
High- and Medium-Priority Recovery Libraries		
Asn43	All	Asn
Thr136	All	Thr
Thr137	All	Thr
Tyr174	All	Tyr
Ser177	All	Ser
Met178	All	Met
Leu182	All	Leu