Engineering, Predicting, and Understanding Nicotinamide Cofactor Specificity

Thesis by
Jackson Kenai Blender Cahn

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California
2016
(Defended January 19, 2016)

ACKNOWLEDGEMENTS

First, I'd like to thank my advisors Frances Arnold and Steve Mayo for their support and guidance over the last four years. In their labs, I've had a great deal of freedom to pursue this work down the various twists and turns it has taken, but Frances and Steve have channeled my natural urge to dabble down the most productive avenues. A particular thank-you to Frances for her guidance in taking my ideas and data and turning them into the narratives that follow.

Sabine Brinkmann-Chen brought me into the Arnold Lab, introduced me to cofactor-switching, and taught me nearly everything I know about experimental biochemistry. She has been my main collaborator on everything in this thesis, and none of this would have been possible without her constant assistance and valuable suggestions. I have also been so lucky to have had a number of excellent students who have assisted me. Ruchi Jahagirdar, Lisa Mears, Armin Baumschlager, Nelson Chou, and Caroline Werlang: thank you. Without your observations, your pipetting prowess, and your knowledge of HTML I would still be far from graduating.

There are too many members of the Arnold and Mayo Labs for me to adequately thank all of them for their assistance. Devin Trudeau and Martin Engqvist deserve particular gratitude for their experimental wisdom; Bernardo Sosa Padilla Araujo, Emzo de los Santos, and Alex Nisthal gave me valuable python pointers; Sheel Dodani, Andy Buller, and Seth Leiblich have been incredibly valuable for helping me work through ideas. Dozens of others, in my labs and beyond, have supported me with their friendship over these years, celebrating my successes and commiserating my setbacks. Thank you to all of you. You have made my time here a joy.

Beyond the lab, I owe gratitude to Pavle Nikolovski and Jens Kaiser of the Molecular observatory for their able assistance in all things X-ray, and to Neil Fromer and Heidi Rusina of the Resnick Sustainability Institute, not only for their assistance in obtaining a graduate fellowship but also for their guidance in presenting my research to a broader audience.

Finally, I would like to thank my family. My grandfather John, who served as a model of scientific success and helped me secure my first research opportunity, my father Andy, who taught me a love of science and of puzzles, my mother June, who taught me to work hard, to write clearly, and the prefix 'sesqui-' (see Chapter 5), and my little brothers Devin and Torey, who taught me to keep fighting when things get tough. Love to you all.

ABSTRACT

Oxidoreductases, the enzymes that catalyze the transfer of electrons between molecules, represent the largest group of enzymes in metabolism, and the vast majority of these enzymes use the functionally-equivalent cofactors nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP) for the storage and transport of the electrons. Understanding the interactions of these proteins with their cofactors is therefore crucial to the engineering of biological pathways and systems that involve these enzymes. In particular, because cells tightly regulate the levels of oxidized and reduced NAD and NADP, it is often valuable to engineer the specificity of enzymes to better integrate them into particular metabolic contexts.

The first section of this thesis focuses on this specificity, taking as a model system the ketol-acid reductoisomerase (KARI) enzyme family. Prior to the work described here, all known members of the KARI enzyme family displayed a strict specificity for NADP over NAD. However, the use of these enzymes in a constructed pathway for the production of medium-chain alcohols created a clear need for NAD-specific KARIs to improve yields. **Chapter 1** briefly summarizes the prior state of the art in nicotinamide cofactor specificity engineering before describing how a previous switching of the KARI from *E. coli* was extended to create a simple recipe for the specificity reversal of any KARI enzyme. **Chapter 2** then uses the insights into cofactor specificity in KARIs afforded by the engineering in Chapter 1 to search databases of KARI sequences and predict naturally NAD-specific KARIs, resulting in the discovery of extremophilic NAD-utilizing KARIs with properties that outstrip those of the best engineered enzymes. **Chapter 3** extends this prediction approach into another enzyme family, xylose reductases, and discusses its strengths and limitations across diverse enzyme folds and families.

The next section of the thesis diverges from the topic of cofactor specificity engineering to briefly explore three structural questions which arise from the study of KARIs. **Chapter 4** covers an analysis of several new KARI crystal structures, including those obtained during the efforts described in Chapters 1 and 2. From a comprehensive comparison of these structures, two topics are addressed: (1) the effect of insertions and deletions in the cofactor specificity loop on the binding geometry of NAD and NADP, and (2) the conformational motions involved in the binding of cofactor, substrate, and metal ions. Based on a pair of structures from Chapter 4, **Chapter 5** experimentally explores the structural evolution of the KARI enzyme family's two distinct

structural classes, replicating in a class I KARI the structural duplication that produced the class II KARI fold, and demonstrating a remarkable retention of enzymatic activity. **Chapter 6** explores a curious sensitivity to mutations observed around the adenine moiety of several KARIs, and extends this observation to a range of other NADP- and NAD-dependent enzymes, with implications both for engineering these proteins and for understanding protein evolution more generally.

The third and final section discusses the development of a general method for the cofactor specificity reversal of any NAD(P)-utilizing enzyme. **Chapter 7** explains the approach used and how it represents a new paradigm in protein engineering, as well as covering the thorough experimental validation to which the method was subjected. This method was developed into a web applet (CSR-SALAD) for public use, and **Chapter 8** discusses the development and use of this tool. This third section represents the culmination of the preceeding work, drawing from an understanding of the sequence and structural determinants of cofactor binding as explored in the preceeding six chapters to create the heuristic picture of specificity which informed the comprehensive engineering approach.

PUBLISHED CONTENT

(* indicates equal contributions by authors)

Brinkmann-Chen S.*, Flock T.*, Cahn J. K. B., Snow C. D., Brustad E. M. McIntosh J. A., Meinhold P., Zhang L., Arnold F. H. (2013). **General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH**, *Proceedings of the National Academy of Sciences* **110**(27), 10946-10951. doi:0.1073/pnas.1306073110

J.K.B.C. participated in the design and performance of research, analysis of data, and writing of the paper.

Brinkmann-Chen S.*, Cahn J. K. B.*, and Arnold F. H. (2014). **Uncovering rare NADH-preferring ketol-acid reductoisomerases**, *Metabolic Engineering* **26**, 17-22. doi:10.1016/j.ymben.2014.08.003

J.K.B.C. participated in the design and performance of research, analysis of data, and writing of the paper.

Cahn J. K. B.*, Brinkmann-Chen S.*, Spatzal T., Wiig J. A., Buller A. R., Einsle O., Hu Y., Ribbe M. W., Arnold F. H. (2015). **Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases**, *Biochemical Journal* **468**(3), 475-484. doi:10.1042/BJ20150183

J.K.B.C participated in the conception of the project, solved and analyzed the crystal structures, prepared the data, and participated in the writing of the manuscript.

Cahn J. K. B.*, Brinkmann-Chen S.*, Buller A. R., Arnold F. H. (2016). **Artificial domain duplication replicates evolutionary history of ketol-acid reductoisomerases**, *Protein Science* Epub ahead of print. Doi:10.1002/pro.2852

J.K.B.C. participated in the conception of the project, designed and cloned the constructs, participated in the solving of the crystal structure, analyzed the crystal structure, and participated in the writing of the manuscript.

Cahn J. K. B.*, Baumschlager A.*, Brinkmann-Chen S., Arnold F. H. (2016). **Artificial domain duplication replicates evolutionary history of ketol-acid reductoisomerases**, *Protein Engineering, Design and Selection* **29**(1), 31-38. doi:10.1093/protein/gzv057

J.K.B.C participated in the design and performance of research, analysis of data, and writing of the paper.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

Section I

Chapter 1: Reversing cofactor preference in the ketol-acid reductoisomerase enzyme family

Chapter 2: Uncovering rare NAD-preferring ketol-acid reductoisomerases

Chapter 3: Discovery of an NAD-preferring xylose reductase

C h a p t e r 1

REVERSING COFACTOR PREFERENCE IN THE KETOL-ACID REDUCTOISOMERASE ENZYME FAMILY

Material from this chapter appears in Brinkmann-Chen S., Flock T., Cahn J. K. B., Snow C. D., Brustad E. M. McIntosh J. A., Meinhold P., Zhang L., Arnold F. H. (2013). **General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH**, *Proceedings of the National Academy of Sciences* **110**(27), 10946-10951, and is reprinted by permission from the National Academy of Sciences.

Abstract

To date, efforts to switch the cofactor specificity of oxidoreductases from nicotinamide adenine dinucleotide phosphate (NADP) to nicotinamide adenine dinucleotide (NAD) have been made on a case-by-case basis with varying degrees of success. Here we present a straightforward recipe for altering the cofactor specificity of one family of NADP-dependent oxidoreductases, the ketol-acid reductoisomerases (KARIs). Combining previous results for an engineered NAD-dependent variant of *Escherichia coli* KARI with available KARI crystal structures and a comprehensive KARI-sequence alignment, we identified key cofactor specificity determinants and used this information to construct five KARIs with reversed cofactor preference. Additional directed evolution generated two enzymes having NAD-dependent catalytic efficiencies that are greater than the wild-type enzymes with NADP. High-resolution structures of a wild-type/variant pair reveal the molecular basis of the cofactor switch.

Introduction

Oxidation and reduction chemistry is involved in every component of cellular metabolism, and in every organism on Earth the primary means for storing and transporting reducing equivalents is the nicotinamide moiety. Evolution has developed two ubiquitous nicotinamide-containing molecules for mediating this chemistry (Figure 1-1): nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). These cofactors differ only in the absence or presence of a phosphate diesterified to the 2’

hydroxyl of the adenosine ribose, generally more than 15 Å from the chemically active carbon of the nicotinamide. This phosphate has no effect on the reduction potential of the molecule, yet the vast majority of enzymes display a strict specificity for one cofactor or the other. This allows the cell to partition metabolic reactions into two groups for separate regulation, but can pose an engineering hurdle.

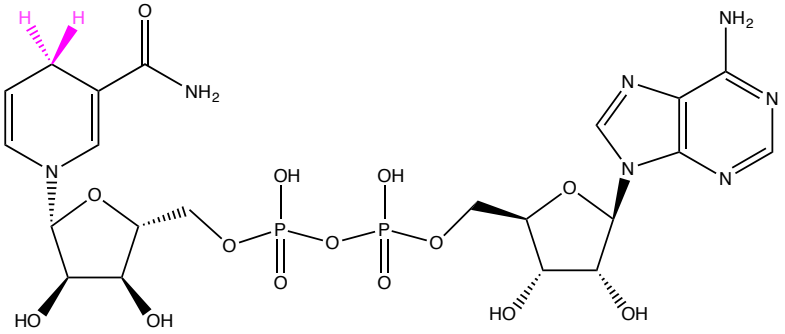
For purposes of *in vitro* catalysis NADPH (the reduced form of NADP) is 16 times more expensive than NADH and has a shorter shelf life.[1,2] Even more importantly, for the purposes of *in vivo* biocatalysis, NADP-dependent enzymes cannot be used under the anaerobic conditions favored for industrial biotechnology because in many organisms NADPH is only produced during aerobic metabolism.[3,4] As bioengineering and metabolic engineering have developed as fields, cofactor specificity has proven a common engineering hurdle. Because of this, NAD(P)-binding enzymes have been the subject of many studies attempting to understand the molecular determinants of cofactor specificity and to engineer specificity reversal (Tables 1-1 and 1-2). However, despite the use of numerous approaches including combinatorial active-site saturation,[5] computational design,[6] and homology guided design,[7] no single method has developed which has been consistently successful in the reversal of cofactor specificity, and many of the published results have been only marginally successful.

Two factors contribute to the challenge of cofactor switching. First, the ubiquity and proposed ancient origin of NAD(P) – potentially dating back to the 'RNA world'[8] – have lead to the evolution of cofactor binding in diverse folds and orientations.[9,10] Even within single enzyme families, the binding of the cofactor molecule can use different geometries between homologues,[11,12] and this structural diversity has limited the development of general methods. Furthermore, a survey of the studies in Tables 1-1 and 1-2 shows that nearly all of these proteins have required multiple simultaneous mutations to achieve reversal of specificity;[3] as a result many traditional protein engineering techniques, which operate by walking up a fitness landscape one beneficial mutation at a time, cannot be used because of the enormously non-additive nature of the specificity fitness landscape.

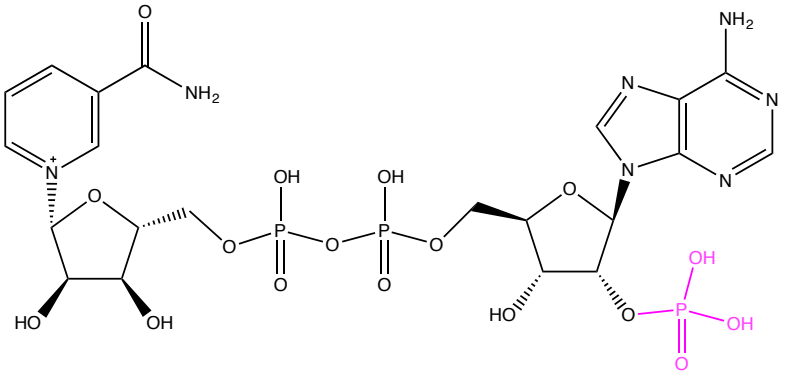
Ketol-acid reductoisomerases (KARI; EC 1.1.1.86) are a family of NADP-dependent oxidoreductases that catalyze an alkyl-migration followed by a ketone reduction,

converting (*S*)-2-acetolactate (S2AL) or 2-aceto-2-hydroxybutyrate to (*R*)-2,3-dihydroxy-isovalerate (RDHIV) and (*R*)-2,3-dihydroxy-3-methylvalerate, respectively,[13] essential intermediates in the biosynthesis of branched-chain amino acids (BCAAs).[14,15] The demand for these essential amino acids, used in the preparation of animal feed, human dietary supplements, and pharmaceuticals, is currently estimated to exceed 1,500 tons per year.[16] In addition, the BCAA pathway has been engineered to produce fine chemicals and biofuels, including 1-butanol and isobutanol.[17,18] Under the anaerobic conditions preferred for large-scale fermentations, biosynthesis of BCAAs and other products that use this pathway is limited by the pathway's cofactor imbalance and reduced cellular production of NADPH.[3,19] One approach to overcoming the cofactor imbalance is to engineer KARI to use NADH generated in glycolysis, thereby enabling anaerobic production of BCAA pathway products.[3,19]

The three prior reports of cofactor-switched KARIs from two different organisms show few commonalities in terms of approach or even residues targeted for engineering.[3,19,20] A general recipe for switching KARI cofactor specificity would allow metabolic engineers to take advantage of the natural sequence diversity of the KARI family and the concomitant diversity in properties such as expression level, pH tolerance, or thermal stability. By combining a systematic analysis of all Swiss-Prot-annotated[21] KARIs, information from previous work in the Arnold Lab on switching the cofactor specificity of the *Escherichia coli* KARI,[3] and available KARI structures, we have identified a subset of residues in the β2αB-loop of the Rossmann fold that distinguish NADP and NAD. In this chapter we provide a nuanced guide to engineering KARI cofactor specificity and apply it to six different KARIs that are representatives of the three canonical KARI β2αB-loop lengths. We also demonstrate that wild-type-like activity using the NAD cofactor, required for industrial applications, can be achieved by directed evolution once cofactor preference has been reversed. High-resolution structures of a wild-type KARI and its cofactor-switched variant with the respective cofactors bound demonstrate how the switch was achieved.



Nicotinamide adenine dinucleotide hydride (NADH)



Nicotinamide adenine dinucleotide phosphate (NADP)

Figure 1-1. The nicotinamide cofactors, nicotinamide adenine dinucleotide (here shown in its reduced (hydride) form (NADH)), and nicotnimamide adenine dinucleotide phosphate (NADP). The differences between the molecules are highlighted in magenta, illustrating the spatial and covalent distance between the 2' recognition element (the phosphate of NADP) and the chemically relevant element (the hydride).

Table 1-1. Previous reports of 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. For citation information, see Supplementary Material S1-1.

Study	Protein	Structure	Method	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Banta et al. 2002)	*C. glutamicum* DKR	1A80	Semi-Rational	R238H	?	?	?
(Baroni et al. 2012)	*P. falciparum* FDNR	2OK7	Rational	Y258F	0.667	$4.7x10^1$	-1.243
(Bastian et al. 2011)	*E. coli* KARI	3ULK	Semi-Rational	A71S, R76D, S78D, Q110V	185.000	$5.4x10^4$	-0.072
This study	*S.* sp. KARI	3ULK (h)	Rational	A71S, R76D, S78D, Q110V	64.000	$1.1x10^5$	-2.148
This study	*S. exigua* KARI	4KQW	Rational	S61D, S63D, I95V	88.000	$7.8x10^3$	-1.561
This study	*M. aeolicus* KARI	4KQW (h)	Rational	G50D, S52D	124.000	$1.2x10^3$	-0.841
This study	*L. lactis* KARI	4TSK (h)	Semi-Rational	V48L, R49P, K52L, S53D, E59K, T182S, E320K	151.064	$2.3x10^4$	0.038
This study	*A. acidocaldarius* KARI	4TSK	Semi-Rational	R48P, S51L, S52D, R84A	110.000	$4.5x10^2$	-1.527
(Chen et al. 1995)	*E. coli* IDH	4AJ3	Homology	C201I, C332Y, K344D, Y345I, V351A, Y391K, R395S	202.469	$1.4x10^6$	-1.457
(Dambe et al. 2006)	*S. morelense* AFDH	2GLX	Homology	A13G, S33D	13.581	?	-0.791
(Döhr et al. 2001)	*H. sapiens* P450R	3QFS	Rational	W676A	0.236	$1.0x10^3$	-0.550
(Elmore and Porter 2002)	*R. norvegicus* P450R	1AMO	Rational	W677A	1.156	$5.3x10^4$	-1.807
(Eppink et al. 1999)	*P. fluorescens* PHBH	1K0J (m)	Rational	R33S, Q34R, P36R, D37A, Y38E	3.111	$5.0x10^4$	-2.017
(Fasan et al. 2011)	*B. megaterium* P450R	4DQL	Semi-Rational	R966N, K972H, Y974F, W1046D	0.615	$4.4x10^2$	-1.072
(Kamerbeek et al. 2004)	*P. fluorescens* HAPMO	2YLR (h)	Semi-Rational	K439F	0.615	$4.3x10^2$	-2.021
(Kamerbeek et al. 2004)	*A.* sp. CHMO	4RG3	Semi-Rational	K326A	0.096	$5.5x10^1$	-2.936
(Katzberg et al. 2010)	*S. cerevisae* DKR	4PVD	Semi-Rational	N9E	0.856	$1.2x10^2$	-2.936
(Khoury et al. 2009)	*C. boidinii* XR	1K8C (h)	Computational	K272G, S273G, N274D	>1	?	-1.071
(Kristan et al. 2007)	*C. lunatus* BHSDH	3QWF	Homology	Y49D	7.770	?	-4.188
(Liang et al. 2007)	*P. stipitis* XR	1K8C (h)	Semi-Rational	K270R, N272D	2.932	9.3	-1.171
(Maddock et al. 2015)	*E. coli* CaADH	1KEV (h)	Homology	G198D, S199V, P201E, Y218A	>1	?	-4.591
(Maurer et al. 2005)	*B. megaterium* P450R	4DQL	Rational	R966D, W1046S	0.372	$2.3x10^2$	-0.401

Table 1-1. continued

Study	Protein	Structure	Method	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Medina et al. 2001)	*A.* PCC7119 FDNR	2BSA	Rational	S223D	0.120	$8.1x10^3$	-5.352
(Nakanishi et al. 1997)	*M. musculus* CR	1CYD	Rational	T38D	30.909	$1.3x10^3$	-0.510
(Paladini et al. 2009)	*P. sativum* FDNR	4AF7 (1QGA)	Rational	Y308S	0.015	$3.3x10^2$	-1.627
(B. Petschacher et al. 2005)	*C. tenuis* XR	1K8C	Rational	K274R, N276D	1.249	$1.9x10^1$	-1.170
(Pick et al. 2014)	*E. coli* AdhZ3	1YQD (h)	Semi-Rational	S199N, S200N, N201D	1.245	$6.9x10^1$	-0.754
(Pick et al. 2014)	*E. coli* AdhZ2	1UUF (1YQD)	Rational	T205D, T206I, S207N	2.188	$3.8x10^1$	0.000
(Rane and Calvo 1997)	*E. coli* KARI	3ULK	Rational	R68D, K69L, K75V, R76D	31.435	$5.8x10^4$	-0.548
(Rodriguez-Arnedo et al. 2005)	*H. volcanii* IDH	1AI2 (h)	Homology	R291S, K343D, Y344I, V350A, Y390P	>1	?	-0.768
(Rosell et al. 2003)	*R. perezi* ADH8	1P0F	Homology	G223D, T224I, H225N	7571.333	$8.1x10^4$	0.551
(Scrutton et al. 1990)	*E. coli* GTR	1GET	Homology	A179G, A183G, V197E, R198M, K199F, H200D, R204P	8.133	$1.8x10^4$	-1.482
(Shiraishi et al. 1998)	*N. crassa* CbR		Rational	S920D, R932S	65.000	$7.2x10^4$	-2.626
(Takase et al. 2014)	*S.* sp. A1-R	3AFN	Loop Substitution	H37N, G38S, R39H, K40V, A41D	>1	?	-3.241
(Yaoi et al. 1996)	*T. thermophilus* IDH	2D1C	Loop Substitution	R231A, K283D, Y284I, N287G, V288I, I290A	67.692	$5.2x10^4$	-1.051
(Zeng et al. 2009)	*P. stipitis* XR	1K8C (h)	Computational	K21A, N272D	>1	?	-0.098
(L. Zhang et al. 1999)	*V. harveyi* ALDH	1EZ0	Semi-Rational	T175E	129.545	$4.8x10^3$	-0.763
(R. Z. Zhang et al. 2009)	*C. parapsilosis* SCR	3CTM (1CYD)	Rational	S67D, P69D	0.313	4.6	-0.180

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 $\mathbf{log}\left(\frac{CE_{mut}^{NAD}}{CE_{WT}^{NADP}}\right)$

Table 1-2. Previous reports of 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. For citation information, see Supplementary Material S1-1.

Study	Protein	Structure	Method	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Ashida et al. 2004)	*S.* sp AlaDH	2VHW (h)	Rational	D198A	13.846	$3.4x10^4$	-0.963
(Bernard et al. 1995)	*L. delbruckii* LDH	1J49	Rational	D175A	1.043	$4.4x10^1$	-0.962
(Bocanegra et al. 1993)	*E.coli* DHLDH	4JQ9 (1GEU)	Homology	E205V, M206R, F207K, D208H, P212R	>1	?	0.668
(Bubner et al. 2008)	*P. fluorescens* M2DH	1M2W	Rational	E68K, D69A	18.667	$4.5x10^3$	0.766
(Capone et al. 2011)	*C. symbiosum* GDH	1BGV (4XGI)	Rational	F238S, P262S	0.323	$5.8x10^1$	-3.323
(Chen et al. 1996)	*T. thermophilus* IMDH	2ZTW	Loop Substitution	N/A	1000.000	$8.7x10^4$	0.187
(Clermont et al. 1993)	*B. stearothermophilus* GAPDH	3CMC	Rational	D32A, L187A, P188S	1.556	?	-1.699
(Cui et al. 2015)	*G. oxydans* Gox2181	3AWD (2WDZ)	Computational	Q20R, D43S	1.457	?	0.073
(Ehrensberger et al. 2006)	*G. oxydans* XDH	1ZEM	Rational	D38S, M39R	>1	?	0.699
(Ehsani et al. 2009)	*S. cerevisae* BDH	2D8A (h)	Homology	E221S, I222R, A223S	>1	?	0.059
(Feeney et al. 1990)	*B. stearothermophilus* LDH	1LDN	Rational	D53S	0.152	3.4	-1.313
(Friesen et al. 1996)	*P. mevalonii* HMG-CoAR	4I4B	Rational	D146A, L148K	0.135	$7.6x10^4$	-3.683
(Galkin et al. 1997)	*T. intermedius* LuDH	1LEH (h,1BW9)	Homology	D203A, I204R, D210R	74.286	?	-1.636
(Gul-Karaguler et al. 2001)	*C. methylica* FDH	2FSS (2NAD)	Rational	D195S	0.024	$6.1x10^3$	-1.479
(Hoelsch et al. 2013)	*M. vaccae* FDH	2GSD (h)	Rational	C145S, A198G, D221Q, C225V	13.254	?	?
(Holmberg et al. 1999)	*B. stearothermophilus* LDH	1LDN	Homology	I51K, D52S	2.200	$4.9x10^1$	-1.582
(Hsieh et al. 2006)	*H. sapiens* m-NAD-ME	1PJ3	Rational	Q362K	3.169	$2.9x10^2$	-0.534

Table 1-2. continued

Study	Protein	Structure	Method	Best Mutant	Final Specificity[a]	Specificity Change[b]	Log Relative Activity[c]
(Jensen et al. 2013)	*S. maltphilia* SMFMO	4A9W (2XLP)	Homology	H194T	0.966	1.5	-0.216
(Ma et al. 2010)	*K. pneumonia* PDOR	3OX4 (h)	Computational	D41G	1.559	?	-0.655
(Marohnic et al. 2003)	*R. norvegicus* CB5R	1IB0	Homology	D239T	10.453	$4.1x10^4$	-0.669
(Miller et al. 2006)	*E. coli* IMDH	1CM7 (2ZTW)	Directed Evolution	K100R, A229T, D236R, L248M, D289K, I290Y, A296V, G337Y	367.000	$5.1x10^4$	-0.228
(Nishiyama et al. 1993)	*T. flavus* MDH	1BMD	Homology	E41G, I42S, P43E, Q44R, A45S, M46F, K47Q	24.000	$5.2x10^2$	-0.470
(Barbara Petschacher et al. 2014)	*S. mutans* NOX	2BC0 (h,2CDU)	Homology	D192A, V193R, V194H, A199R	10.000	$6.4x10^4$	0.559
(Serov et al. 2002)	*S. cerevisae* FDH	2NAD (h)	Rational	D196A, Y197R	2.311	?	-3.796
(Takase et al. 2014)	*S.* sp. A1-R'	4TKM	Loop Substitution	T16S, E17Q, N37H, S38G, H39R, V40K, D41A	84.628	$1.1x10^3$	0.665
(Watanabe et al. 2005)	*P. stipitis* XDH	1PL6 (h)	Homology	D207A, I208R, F209T	2.641	$1.1x10^4$	0.049
(Woodyer et al. 2003)	*P. stutzeri* PDH	4E5K	Semi-Rational	E175A, A176R	2.763	$3.0x10^2$	0.995
(Zheng et al. 2013)	*B. subtilis* InDH	3NT2	Rational	A12K, D35S, V36R	4.750	?	0.032

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 $\mathbf{log}\left(\frac{CE_{mut}^{NAD}}{CE_{WT}^{NADP}}\right)$

Materials and Methods

Cloning and library construction

The genes encoding *S. exigua* SeKARI, *L. lactis* LlKARI, and *Shewanella* sp. ShKARI were obtained from DNA2.0. The genes encoding *M. aeolicus* MaKARI and *A. acidocaldarius* AaKARI were obtained as gBlocks from Integrated DNA Technologies. For each gene, the gBlocks were assembled via PCR, using T7 promoter and terminator primers and Phusion polymerase following the manufacturer's instructions (Thermo Scientific). Site-saturation mutagenesis libraries were made by splicing by overlap extension PCR, as described.[3,22] Error-prone PCR was performed according to a published protocol,[23] using commercial T7 promoter and terminator primers. All KARIs and libraries were cloned into pET22(b)+, using NdeI and XhoI in frame with the C-terminal his-tag for expression in *E. coli*. Heterologous protein expression, high-throughput expression, and purification were conducted as described.[3]

Kinetic assays and high-throughput screening

For the high-throughput assays, *E. coli* cells were lysed with 100 mM potassium phosphate at pH 7, 750 mg/L lysozyme, and 10 mg/L DNaseI. KARI activities were then assayed by monitoring NAD(P)H consumption in the presence of S2AL at 340 nm in a plate reader. The assay buffer contained 100 mM potassium phosphate at pH 7, 1 mM DTT, 200 μM NAD(P)H, 12.5 mM S2AL for LlKARI and 2.5 mM for the other KARIs, and 10 mM $MgCl_2$. The LlKARI error-prone PCR library was screened at 5 mM S2AL. The EcIlvC^{6E6} library was screened at 1 mM S2AL. The SeKARI^{DD} library was screened at 100 μM NADH and 2.5 mM S2AL.

KARI sequence alignment

Manually annotated and reviewed sequence data for ketol-acid reductoisomerases (E.C. 1.1.1.86) were retrieved from the UniProt Database.[24] Clustal Omega[25,26] was used to perform a multiple sequence alignment. The sequence logo plot[27] was created with the WebLogo 3.3 interface.[28]

Crystallization and data collection

N-hydroxy-*N*-isopropyloxamate was prepared as described.[29] High-throughput screening of crystallization conditions for SeKARI and SeKARI^{DDV} was conducted at the Beckman Molecular Observatory at the California Institute of Technology. For SeKARI with NADPH, the best condition was an unbuffered 0.2 M di-ammonium tartrate solution containing 20% (wt/vol) polyethylene glycol (PEG) 3350 as precipitant. For SeKARI^{DDV} with NADH and *N*-hydroxy-*N*-isopropyloxamate as inhibitor, the best condition was an unbuffered 0.1 M potassium thiocyanate solution with 30% (wt/vol) PEG monomethyl ether 2000 as precipitant. The crystals were soaked in Fomblin oil for cryoprotection before flash-freezing in liquid nitrogen. Diffraction data were collected using a Dectris Pilatus 6M detector on beamline 12–2 at the Stanford Synchrotron Radiation Laboratory at 100 K. Diffraction datasets were integrated with XDS[30] and scaled using SCALA.[31]

Structure determination and refinement

For SeKARI, the structure of *Pseudomonas aeruginosa* KARI (PDB code 1NP3[32]) was used as for molecular replacement. A multiblock refinement was applied dividing the model in six subparts according to secondary structure elements (residues 1–202, 203–228, 229–252, 253–278, 279–308, and 309–337) to allow automated standard refinement with Phenix (CCP4 suite). Refinement was conducted by iterating automatic refinement with Refmac5 (CCP4 suite) and manual refinement using Coot.[33] We used the refined wild-type structure as a model for molecular replacement to obtain the structure for SeKARI^{DDV}. After placement of the inhibitor, several iterations of automated refinement with Refmac5 and manual refinement in Coot were performed. The structures were submitted to the protein database as PDB 4KQW (SeKARI) and PDB 4KQX (SeKARI^{DDV}).

Results and Discussion

In previous work, Bastian *et al*. described *E. coli* KARI variant EcIlvC^{6E6} with four mutations (A71S, R76D, S78D, and Q110V) that resulted in a 54,000-fold reversal in cofactor specificity for NAD over NADP.[3] This variant was also highly active when using NAD (85% of wild-type activity using NADP). Structural analysis of wild-type EcIlvC with and without bound cofactor (PDB codes 3ULK and 1YRL[34,35]) showed that three of the four EcIlvC^{6E6} mutations (A71S, R76D, and S78D) are in a loop connecting the β2-

strand and the αB-helix of the Rossmann fold,[36] herein referred to as the β2αB-loop. R76 and S78 are in direct contact with the 2′-phosphate of NADP. The existence of β2αB-loops of varying lengths obscured sequence patterns of NADP specificity in KARI multiple sequence alignments.[13,19,20,37] To find the commonalities among KARI β2αB-loops, we systematically analyzed the loop regions of the entire enzyme class. We generated a multiple sequence alignment of all 643 Swiss-Prot-annotated[21] KARI sequences (Table S1-1) and used structural data to identify and refine the β2αB-loop region in the alignment. An excerpt (Figure 1-2a) of the alignment of a few representative KARIs shows the well-known, conserved GxGxxG motif[38] and, 18 amino acids downstream, the β2αB-loop, which is diverse in both length and amino acid sequence.

On the basis of the sequence alignment alone, the diversity of this region might seem to argue against a conserved function for loop residues (Figure 1-2a). However, analysis of the 643 KARI sequences shows three different loop lengths: six (14%), seven (68%), and 12 residues (18%) (Fig. 1-2b). On subalignment of KARIs according to loop length, common conservation patterns emerge: a positively charged residue (73% arginine and 9% lysine) usually appears at the N-terminal end of the β2αB loop, and small, polar residues such as serine (85%) or threonine (11%) predominate at the C-terminal end. The last two residues in the six-residue loops are conserved (lysine/arginine and serine), and the last three residues follow the pattern SxS in seven-residue and RxS in 12-residue loops.

Comparison of the β2αB-loop segment of all available KARI structures (seven structures from four different organisms[32,34,37,39-41]) provides an explanation for the trends found in the sequence alignment. Three of the seven structures have NADP cocrystallized, and in all those cases, the small polar C-terminal residues interact with the NADP 2′-phosphate. The conserved N-terminal arginine is homologous to R68 in EcIlvC^{6E6}, which has recently been shown to form a cation–pi interaction with the NADP adenine ring[34] and likely plays a dual role in binding the nucleobase, as well as the 2′-phosphate. On the basis of these findings, we postulate that the β2αB-loop is a major determinant of NADP cofactor specificity in wild-type KARIs. This analysis of the β2αB-loop provides the basis for transferring previously reported cofactor-switching mutations in EcIlvC^{6E6} to corresponding residues in KARIs that have similar 12-residue loops, as well as KARIs with six- and seven-residue loops.

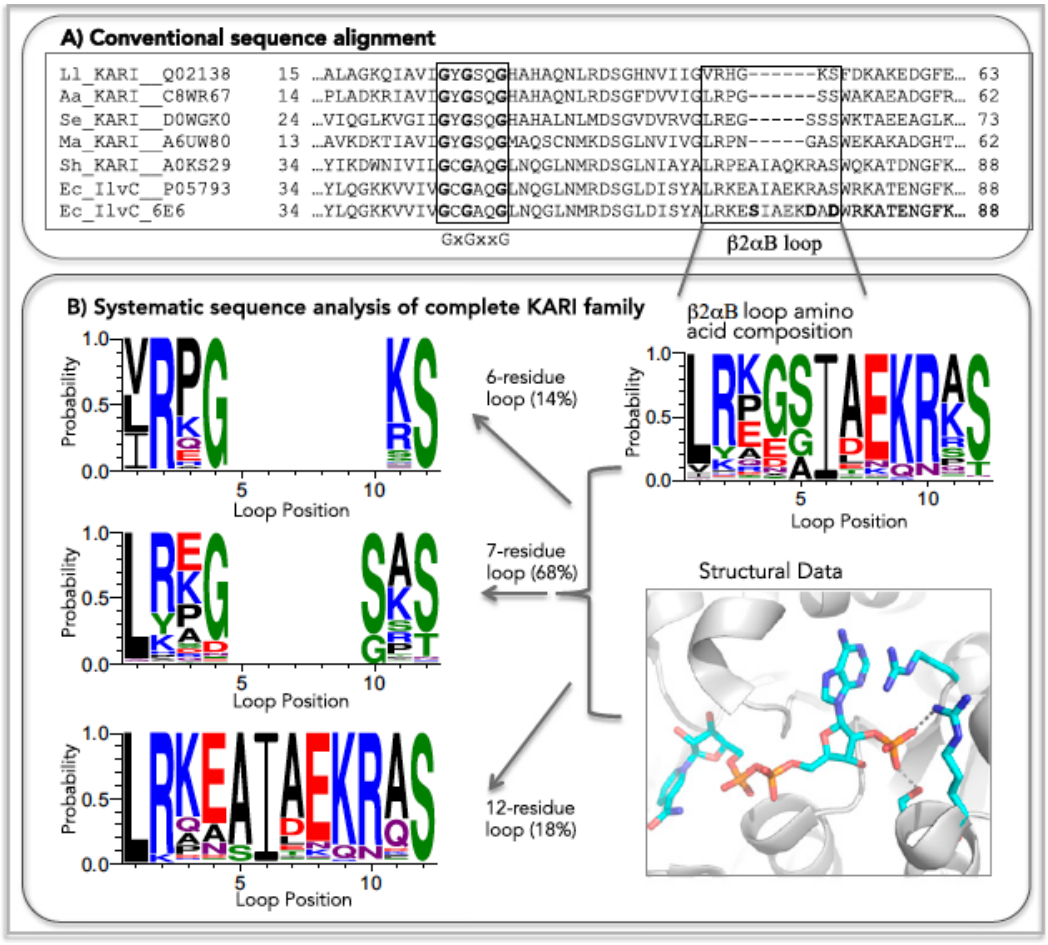


Figure 1-2. Excerpt of a multiple sequence alignment (a) of *L. lactis* KARI (LlKARI, UniProt ID Q02138), *A. acidocaldarius* KARI (AaKARI, C8WR67), *S. exigua* KARI (SeKARI, D0WGK0), *M. aeolicus* KARI (MaKARI, A6UW80), *Shewanella* sp. KARI (ShKARI, A0KS29), *E. coli* KARI (EcIlvC, P05793), and EcIlvC^{6E6}. The GxGxxG motif is highlighted in bold and boxed. The β2αB-loop is boxed. Mutations of variant EcIlvC^{6E6} are shown in bold. Subalignments according to β2αB-loop length (b) uncover three different conservation patterns we suggest are important for 2′-phosphate interaction. Colors in the sequence logo refer to hydrophobic (black), positively charged (blue), negatively charged (red), and polar (green) residues.

Transfer of a 12-residue cofactor switch solution to seven- and six-residue β2αB-loops

We examined the transferability of mutations previously reported for EcIlvC^{6E6}, which has a 12-residue loop, to KARIs with shorter loops using SeKARI from *S. exigua* (seven residues) and LlKARI from *L. lactis* (six residues) as models. On inspection of the sequence alignment in Figure 1-2a, we identified residues S61 and S63 of SeKARI as corresponding to residues R76 and S78 in EcIlvC (R76D and S78D in EcIlvC^{6E6}). EcIlvC mutation A71S had no match in the SeKARI amino acid sequence because of the latter enzyme's shorter β2αB-loop. Grafting both aspartates from EcIlvC^{6E6} into SeKARI resulted in variant SeKARI^{DD} (S61D and S63D) and a 7,800-fold reversal of cofactor specificity from NADPH to NADH. SeKARI^{DD} (Table 1-3) had an eightfold decreased k_{cat} with NADPH (from 0.8 to 0.1 s^{-1}), whereas the k_{cat} with NADH increased from 0.4 to 1.0 s^{-1}. The mutations increased the K_M for NADPH 880-fold, but only 2.5-fold for NADH. Overall, the catalytic efficiency for NADH remained the same, whereas catalytic efficiency for NADPH was reduced 7,300-fold to 0.11 $mM^{-1}s^{-1}$. That switching SeKARI cofactor specificity could be achieved with only two mutations suggests that KARIs with seven-residue and 12-residue β2αB-loops share similar cofactor specificity determinants involving interactions with small, polar residues at the end of the β2αB-loop.

LlKARI is a representative of the 14% of the 643 KARI sequences with a six-residue β2αB-loop. Whereas in 12- and seven-residue β2αB-loop KARIs the antepenultimate and ultimate residues are highly conserved, in six-loop KARIs the ultimate conserved serine is usually preceded by a positively charged residue at the penultimate position. We hypothesized that residues K52 and S53 of LlKARI were equivalent in function to residues R76 and S78 in EcIlvC. LlKARI^{DD} variant with mutations K52D and S53D, however, expressed at an extremely low level and exhibited no measurable activity with either cofactor, which could be a result of the destabilizing effects of the two adjacent aspartate residues. Similar results were obtained with LlKARI^{ED}. These two aspartate mutations are separated by an additional residue in KARIs with the longer β2αB-loops. The failed transfer of *E. coli* cofactor switch mutations to LlKARI suggests that KARIs with six-residue loops require a modified approach.

To develop a recipe for switching cofactor specificity in KARIs having a six-residue β2αB-loop, we generated single-site-saturation mutagenesis libraries at each of the six loop residues in LlKARI (V48, R49, H50, G51, K52, and S53), expecting that the β2αB-loop also is key to specificity in this KARI family. On screening for activity with both cofactors, we found no single mutation that resulted in an NAD-preferring variant (rate of NADH consumption > rate of NADPH consumption at saturating substrate conditions). Mutations were identified at R49 (proline) and V48 (leucine) that increased activity with both cofactors. Saturation mutagenesis at H50, G51, K52, and S53 did not yield variants with improved NADH activity. Convinced that NADP specificity in KARIs with a six-residue β2αB loop is conveyed in a similar manner as in KARIs with seven- and 12-residue loops, we built a dual-site library by saturation mutagenesis at Lys52 and Ser53 while also incorporating the V48L and R49P mutations. Screening this library to ~80% coverage, we identified two variants, LlKARI^{LPLD} (mutations V48L, R49P, K52L, and S53D) and LlKARI^{LPED} (mutations V48L, R49P, K52E, and S53D), with 46-fold and 54-fold specificity for NADH over NADPH (Table 1-3), representing specificity shifts of 6,600 (LlKARI^{LPLD}) and 7,700 (LlKARI^{LPED}).

These cofactor-switched KARIs with six-residue loops contained mutations to four of the six loop residues, suggesting that KARIs with six-residue loops are slightly more difficult templates for engineering the cofactor switch. Although mutation of the conserved arginine at the beginning of the β2αB-loop was not required to switch cofactor specificity in seven- and 12-loop KARIs, variant LlKARI^{LRLD} with proline reverted to arginine showed that the arginine contributes to NADP cofactor preference in six-residue loop KARIs. Reversion of P49 to arginine not only reduced the cofactor K_M values (twofold for NADH and threefold for NADPH, Table 1-3) but also decreased the k_{cat} on NADH (fourfold), thereby reducing the 46-fold preference for NADH over NADPH to only sixfold. Reversion of L48 to valine (LlKARI^{VPLD}) lowered the catalytic efficiency on NADH threefold (because of a reduction in k_{cat}), whereas the NADPH K_M was reduced twofold. Although polar residues at the C-terminal end of the loop and charged residues at the N-terminal end of the loop continue to support NADP specificity, removing the polar residues at the C-terminal end of the loop might not be sufficient to switch the specificity of six-residue loop KARIs in favor of NAD. In the shorter six-residue loop KARIs, the N-terminal, positively charged arginine residue appears to form stronger interactions with the

2′-phosphate than in the seven- and 12-residue loop KARIs, possibly because of its closer packing and proximity to the cofactor. The six-amino acid loop KARI structure with cofactor that we describe in Chapter 4 will help explain these differences.

Recovering catalytic activity of cofactor-switched enzymes

Shifting cofactor preference often decreased the overall activity (i.e., catalytic efficiency using NADH) relative to the wild-type enzyme using NADPH in our KARIs, which is also true for other cofactor-switched enzymes.[20,42-44] To demonstrate that the activity of the cofactor-switched variants can be improved to match or even surpass the wild-type enzyme, we randomly mutated SeKARI^{DD}, LlKARI^{LPLD}, and EcIlvC^{6E6} and screened for higher total activity in cell lysate. For SeKARI, we identified variant SeKARI^{DDV} with mutation I95V. Interestingly, this mutation corresponds to Q110V, which was previously found in EcIlvC^{6E6}, and is speculated to confer general activation by optimizing cofactor orientation for catalysis. With LlKARI, we isolated variant LlKARI^{2G6}, which contained three new mutations: E59K, T182S, and E320K. These mutations effectively restored the enzyme to wild-type levels of activity; the catalytic efficiency of LlKARI^{2G6} with NADH was ~10% higher than wild-type LlKARI with NADPH (71 vs. 65 $mM^{-1}s^{-1}$). The EcIlvC^{6E6} random mutant library yielded variant EcIlvC$^{P2D1\text{-}A1}$, also with three additional mutations (D146G, G185R, and K433E) and an approximately twofold greater catalytic efficiency on NAD than the wild-type on NADP (Table 1-3). The random mutations in EcIlvC$^{P2D1\text{-}A1}$ and LlKARI^{2G6} are surface mutations, and their effects are difficult to rationalize. These three examples demonstrate that the activity of a cofactor-switched enzyme can be improved, even to levels that exceed wild-type activity, and that activating mutations can be found outside of the specificity-determining β2αB-loop.

Table 1-3. Biochemical properties of EcIlvc (12 residues), ShKARI (12 residues), SeKARI (seven residues), MaKARI (seven residues), LlKARI (six residues), AaKARI (six residues), and variants. T_{50} values are given for some enzymes.

Enzyme	Mutations	T_{50} [°][a]	K_M for cofactors [μM]		k_{cat} for cofactors [s^{-1}]		k_{cat}/K_M [$mM^{-1}s^{-1}$]		NADH/NADPH of k_{cat}/K_M
			NADH	NADPH	NADH	NADPH	NADH	NADPH	
EcIlvC	--	44.0 ± 0.1	1,075 ± 370	41 ± 3	0.3 ± 0.0	3.6 ± 0.4	0.3 ± 0.1	88 ± 11	0.003 ± 0.001
EcIlvC^{6E6}	**A71S**, **R76D**, **S78D**, Q110V	43.7 ± 0.2	30 ± 6	650 ± 80	2.3 ± 0.2	0.20 ± 0.02	74 ± 15	0.40 ± 0.05	185 ± 50
EcIlvC$^{P2D1\text{-}A1}$	**A71S**, **R76D**, **S78D**, Q110V, D146G, G185R, K433E	41.3 ± 0.2	26 ± 1	> 1,400	4.3 ± 0.3	0.54 ± 0.20	165 ± 22	< 0.4	> 412 ± 162
ShKARI	--		415 ± 44	1.0 ± 0.1	1.1 ± 0.1	4.5 ± 0.1	2.6 ± 0.3	4,500 ± 450	0.0006 ± 0.0001
ShKARI^{DD}	**R76D**, **S78D**		90 ± 24	> 1,000	1.30 ± 0.01	0.10 ± 0.02	14 ± 4	0.12 ± 0.02	> 119 ± 46
ShKARI^{6E6}	**A71S**, **R76D**, **S78D**, Q110V		75 ± 10	600 ± 130	2.40 ± 0.01	0.30 ± 0.05	32 ± 4	0.5 ± 0.1	73 ± 23
SeKARI	--	51.3 ± 0.2	45 ± 10	1.0 ± 0.1	0.41 ± 0.02	0.8 ± 0.1	9 ± 2	800 ± 100	0.01 ± 0.00
SeKARI^{DD}	**S61D**, **S63D**	44 ± 1	113 ± 4	880 ± 523	0.97 ± 0.01	0.10 ± 0.01	9 ± 1	0.11 ± 0.07	78 ± 49
SeKARI^{DDV}	**S61D**, **S63D**, I95V		47 ± 15	> 1000	1.01 ± 0.01	0.25 ± 0.04	22 ± 7	0.25 ± 0.04	87 ± 32
[b]MaKARI	--		59 ± 4	17.3 ± 0.2	0.3 ± 0.0	0.7 ± 0.1	4.6 ± 0.3	43 ± 6	0.12 ± 0.02
[b]MaKARI^{DD}	**G50D**, **S52D**		26 ± 2	80 ± 9	0.15 ± 0.01	0.004 ± 0.001	6.2 ± 0.6	0.05 ± 0.01	124 ± 24
LlKARI	--	50 ± 1	285 ± 30	13 ± 1	0.10 ± 0.01	0.8 ± 0.1	0.43 ± 0.06	65 ± 11	0.007 ± 0.002
LlKARI^{LPLD}	**V48L**, **R49P**, **K52L**, **S53D**	46.5 ± 0.3	108 ± 9	1,000 ± 100	0.40 ± 0.01	0.08 ± 0.01	3.7 ± 0.3	0.08 ± 0.01	46 ± 8
LlKARI^{LRLD}	**V48L**, **K52L**, **S53D**		60 ± 8	306 ± 70	0.09 ± 0.00	0.09 ± 0.00	1.7 ± 0.2	0.30 ± 0.07	6 ± 1
LlKARI^{VPLD}	**R49P**, **K52L**, **S53D**		105 ± 7	447 ± 91	0.13 ± 0.01	0.07 ± 0.00	1.2 ± 0.1	0.15 ± 0.03	8 ± 2
LlKARI^{LPED}	**V48L**, **R49P**, **K52E**, **S53D**		128 ± 9	1,180 ± 280	0.35 ± 0.12	0.06 ± 0.01	2.7 ± 0.9	0.05 ± 0.01	54 ± 26
LlKARI^{2G6}	**V48L**, **R49P**, **K52L**, **S53D**, E59K, T182S, E320K		15 ± 4	749 ± 95	1.01 ± 0.03	0.35 ± 0.08	70 ± 19	0.47 ± 0.12	153 ± 57
[b]AaKARI	--		28 ± 2	18.0 ± 0.2	0.26 ± 0.01	0.66 ± 0.01	9 ± 1	37 ±1	0.24 ± 0.02
[b]AaKARI^{PLD}	**R48P**, **S51L**, **S52D**		43 ± 6	> 1,000	0.03 ± 0.00	0.013 ± 0.0004	0.7 ± 0.1	< 0.013	> 54 ± 20
[b]AaKARI^{PLDA}	**R48P**, **S51L**, **S52D**, R84A		27 ± 3	> 1,000	0.03 ± 0.00	0.01 ± 0.00	1.1 ± 0.2	< 0.01	> 110 ± 20
[b]AaKARI^{LS}	Loop switch (LS) of LlKARI^{LPLD} to AaKARI: **R48P**, **P49H**, **S51L**, **S52D**		46 ± 6	> 1,000	0.03 ± 0.00	0.009 ± 0.006	0.6 ± 0.1	< 0.01	> 55 ± 8

Mutations are given relative to each wild-type sequence. Each value represents the average of three independent measurements. Mutations located within the β2αB-loop are highlighted in bold.

[a]Half-denaturation temperature (T_{50}) determination: 30 μL aliquots of purified enzyme were transferred to PCR tubes. Each tube was assigned a specific incubation temperature on the block of an Eppendorf master cycler PCR machine. The measurements were conducted in duplicates. The tubes were incubated in their slots for 15 min, and the reactions were quenched on ice. Residual activity was determined with the activity assay. T_{50} is defined as the termperature at which 50% of the initial activity is retained after 15 min incubation.

[a]MaKARI, AaKARI, and their variants show cooperative behavior, and their kinetics follow the Hill equation with a Hill coefficient of 2.0 instead of the Michaelis-Menten equation: affinity is described as K_H and catalytic efficiency as k_{cat}/K_H to compare them to the other KARIs. No change in Hill coefficient was observed upon mutations.

Cofactor switch guide for the KARI enzyme family

We propose the following guide for switching KARI cofactor specificity, which does not require *a priori* knowledge of the KARI structure (Figure 1-3). The first step is the identification of the β2αB-loop and its length via sequence alignment against the KARIs reported in this work or a multiple sequence alignment of KARIs. If the target KARI has a 12- or seven-residue loop, replacement of the last and third-to-last residue of the loop with aspartates is likely to achieve a switch in cofactor specificity. In the case of the six-residue loop, a modified approach is required. The last polar loop residue should be mutated to aspartate, and the conserved charged residue near the N-terminus of the loop should be mutated to proline. Simultaneously, the penultimate loop residue should be targeted for site-saturation mutagenesis. This approach led to reversed cofactor preference in both test cases. Last, to achieve wild-type-like activity for NAD, additional mutations that fine-tune cofactor orientation, as exemplified by Q110V or I95V, may be introduced, as will be discussed further in Chapter 6. Additional enhancement of activity can be achieved by random mutagenesis and screening.

Application of the cofactor switch guide

We tested the proposed protocol on three additional KARIs representing the β2αB-loop lengths (twelve, seven, and six residues) and composition. With the addition of these three KARIs, we covered the different phylogenetic subbranches of the KARI enzyme family. Representative of 12-residue β2αB-loop KARIs was ShKARI from *Shewanella* sp. *Methanococcus aeolicus* MaKARI[45] exemplifies the seven-residue type, and AaKARI from *Alicyclobacillus acidocaldarius* has a six-residue loop (Table 1-3). We also used this last enzyme to test the transferability of the Q110/I95 position by making and screening a site-saturation library at position R84.

By introducing a customized set of two to four mutations based on the guide in Figure 1-3, we obtained variants with the desired cofactor specificity for KARI family members sharing as little as 20% sequence identity (Table S1-2). Low catalytic efficiency in cofactor-switched variants can be remedied by directed evolution, as demonstrated for EcIlvC$^{P2D1-A1}$ and LlKARI^{2G6}. Mutations (alanine or valine) at positions corresponding to EcIlvC's Q110 improved activity in three different KARIs. Overall, catalytic efficiency

ratios (NADH/NADPH) of more than 400-fold (Figure 1-4) with catalytic efficiencies up to 188% of those of the corresponding wild-type enzymes using NADPH were achieved, corresponding to shifts in cofactor specificity of 200–200,000-fold.

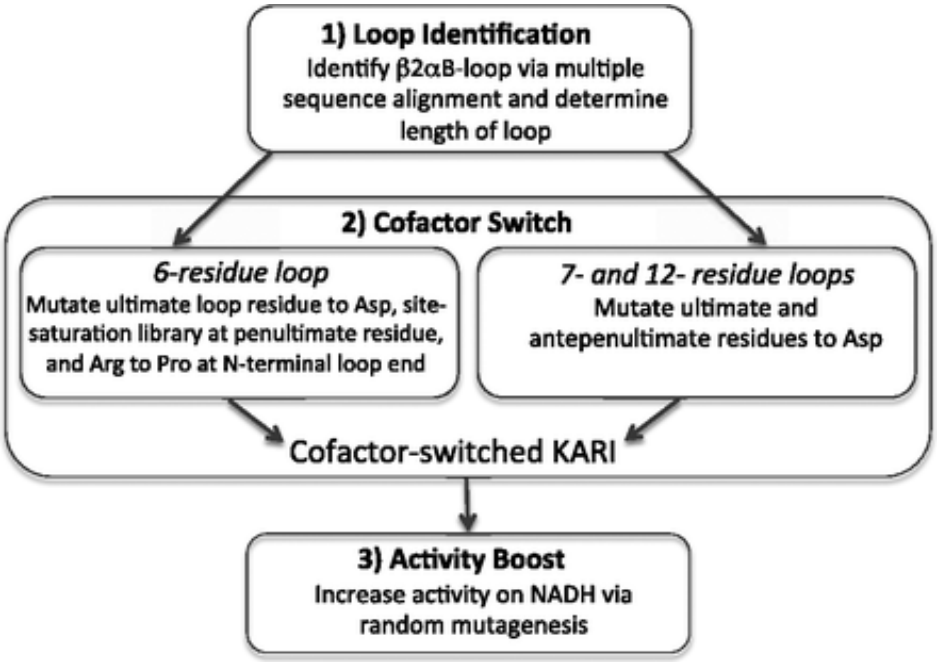


Figure 1-3. Cofactor switch guide for the KARI enzyme family. Steps include (1) identification of the loop, (2) determination of the β2αB-loop length and mutation based on loop length, and (3) improvement of overall activity on NAD via random mutagenesis.

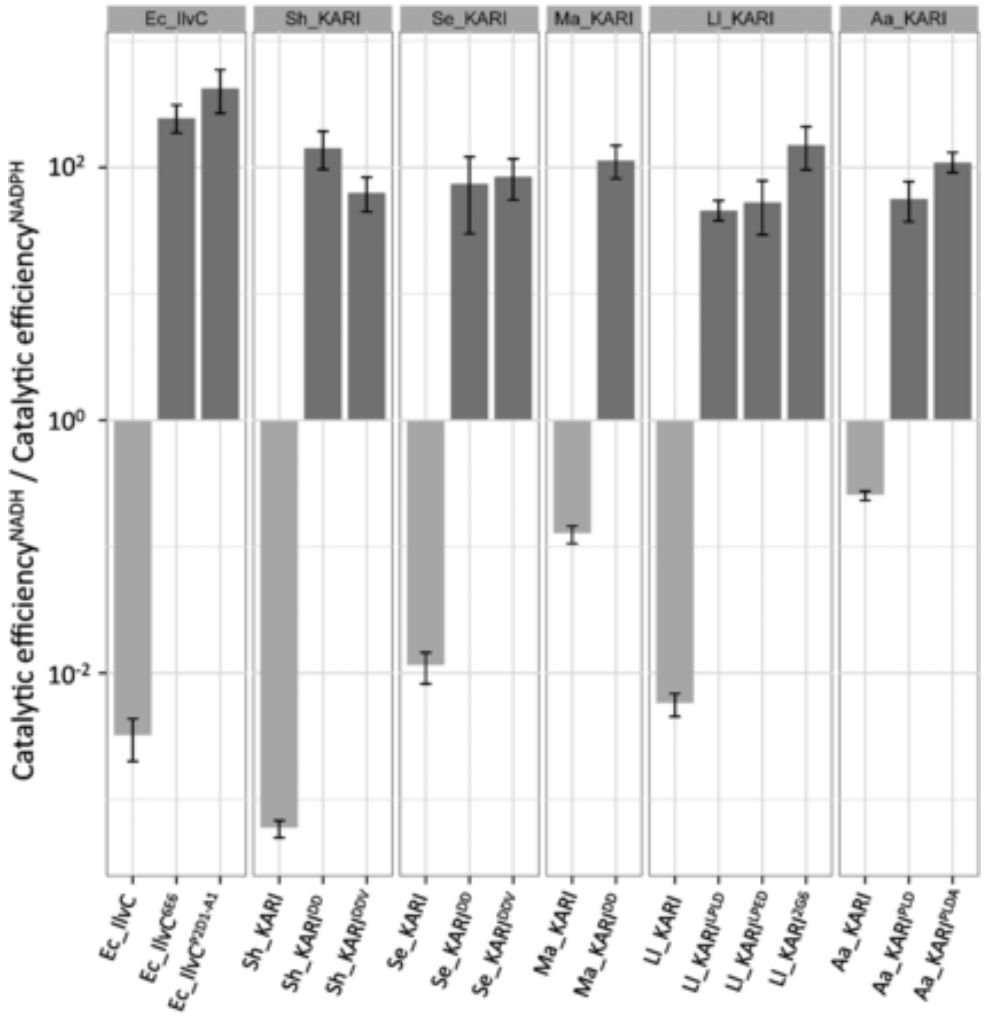


Figure 1-4. Catalytic efficiency NADH/catalytic efficiency NADPH, on log scale, for six wild-type KARIs (light gray) and their cofactor-switched variants (dark gray).

Molecular determinants of cofactor specificity in KARIs

We solved the crystal structures of SeKARI wild-type enzyme (1.39 Å, PDB 4KQW) and variant SeKARI^{DDV} (1.8 Å, PDB 4KQX) with their respective cofactors. The crystallographic parameters are summarized in Table 1-4. These structures confirm that only the β2αB-loop is involved in interactions with the respective 2′-moiety, and thus is responsible for specificity (Figure 1-5). In the wild-type structure, three residues form direct interactions with the 2′-phosphate of NADP: R58, S61, and S63. The structures support the suggested dual role of R58: the positively charged guanidinium moiety is 3.5 Å from the adenine moiety of the cofactor in both structures, forming cation–pi stacking interactions,[46] as reported for *E. coli* KARI.[34] At the same time, this side chain could form a salt bridge to the negatively charged 2′-phosphate of NADP (possibly also involving the oxygen of the phosphoester bond) and a hydrogen bond to the 2′-hydroxyl of NAD.

The residues that were mutated to alter cofactor preference, S61 and S63, are in a position to hydrogen bond directly to at least a single oxygen atom of the phosphate. The high-resolution structures revealed water molecules surrounding the 2′-phosphate group, enabling additional, indirect interactions with the side chains of R58, S61, and S63 (Figure 1-5). The side chain of S62 stabilizes this network of water molecules, as does the R58 backbone. In the mutant structure, the β2αB-loop is moved slightly closer toward the cofactor. Mutations S61D and S63D would interrupt the serine hydrogen bonding interactions and also result in electrostatic repulsion to the 2′-phosphate of NADP. The carboxyl groups of the two aspartates compensate for the missing 2′-phosphate by filling the pocket and substituting its negative charge. In addition, the side-chain carboxylate moiety of S61D is at an ideal distance for hydrogen bonding to the 2′-hydroxyl group of NAD. As in the wild-type structure, water molecules link the 2′-hydroxyl moiety of the ribose sugar with R58 and D63 by hydrogen bonds. All other interactions of SeKARI^{DDV} with the cofactor, for instance, involving the GxGxxG motif, remain the same. The remaining loop residues L57, E59, and G60 are not involved in binding the cofactor.

Whereas these mutations shifted the KARI cofactor preference, improved catalytic activity was achieved by substituting an additional residue that is not part of the loop, I95, with valine. This mutation retained the hydrophobic nature while allowing the adenine

moiety to shift about 1 Å toward the side chain of residue 95. This compensates exactly the distance the β2αB-loop is reoriented inward in the mutant structure at position R58 and preserves the cation–pi stacking of the adenine moiety and the side chain of R58, which is in the same rotamer conformation in the wild-type and in the cofactor-switched mutant. We propose that this movement compensates for the slightly different conformations of NADP and NAD and readjusts the catalytically active nicotinamide moiety of NAD to take on a more favorable position for electron transfer. A similar activating mechanism is speculated for EcIlvC^{6E6}'s Q110V mutation.[3]

Conclusions: General cofactor binding principles for KARIs

We identified a common motif for cofactor specificity in KARIs by using structural knowledge to identify the critical role of the β2αB-loop and deconvolute its three key variations in multiple sequence alignments, thereby defining a limited set of mutations that generate NAD specificity. Applying this to KARIs with different loop lengths has enabled us to develop a robust guide to switching the cofactor preference from NADP to NAD of any enzyme in this family. This approach opens the door to exploration of a wealth of different KARI properties in the context of valuable BCAA pathways under anaerobic conditions.

Table 1-4. Crystallographic parameters of wild-type SeKARI and Variant SeKARI^{DDV}.

	SeKARI	SeKARI^{DDV}
Protein Data bank code	4KQW	4KQX
Space group	$P\ 12_11$	$P\ 2_12_12_1$
Monomers per asymmetric unit	2	2
Water molecules per monomer, *n*	443	140
Cocrystallized cofactor	NADPH	NADH
Cocrystallized compounds	L-(+)-tartarate	*N*-hydroxy-*N*-isopropyloxamate
Unit-cell parameters (Å)	a = 52.205	a = 49.985
	b = 118.934	b = 105.346
	c = 62.158	c = 122.357
	α = 90.00	α = 90.00
	β = 101.17	β = 90.00
	γ = 90.00	γ = 90.00
Resolution range (Å)	38.81-1.39	79.83-1.80
Number of unique reflections	130,511	78,635
R_{work}	0.1595	0.2014
R_{free}	0.1866	0.2520
Refinement parameters		
RMSD from ideal		
Bond (Å)	0.0293	0.0210
Angle (°)	2.7444	2.1084
Chirality	0.1858	0.1499
Ramachandran parameters		
Favored	95.85%	94.33%
Allowed	4.15%	5.51%
Outliers	0.00%	0.15%

Crystallographic parameters of solved structures of wild-type SeKARI and variant SeKARI^{DDV}.

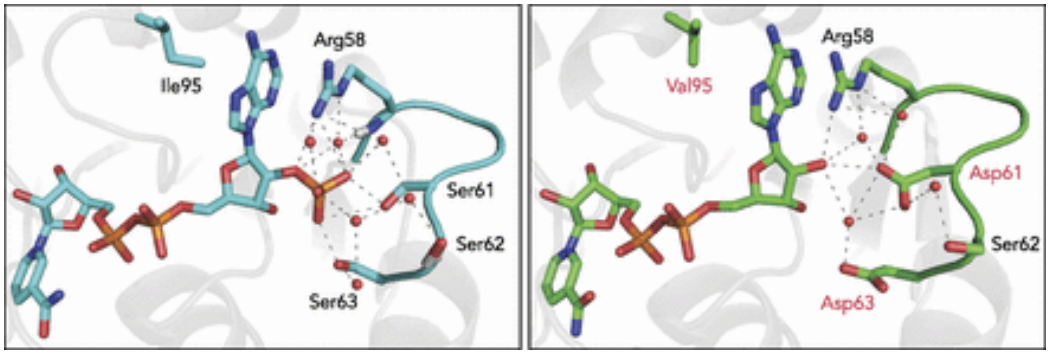


Figure 1-5. Crystal structures of SeKARI wild-type enzyme with cocrystallized NADP (*left*, cyan) and variant SeKARI^{DDV} with cocrystallized NAD (*right*, green). The β2αB-loop is highlighted, and side chains involved in defining cofactor-specificity are shown as sticks. Introduced mutations (S61D, S63D, and I95V) are shown with red labels. In SeKARI^{DDV}, S61D and S63D compensate for the missing 2′-phosphate and electrostatically and sterically exclude NADP. Mutation I95V allows the adenine moiety to shift 1 Å inward. The backbone of R58 follows this movement, so that in both structures the cation–pi interaction with the adenine moiety is preserved.

References

1. Flock, T. (2012) *Investigation of Cofactor Specificity-Determining Amino Acid Residues of Ketol-Acid Reductoisomerases: Structure-Guided Evolution to Swith the Nicotinamide Cofactor Preference.* M.S., California Institute of Technology
2. 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**, 11604-11614
3. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metabolic Engineering* **13**, 345-352
4. Verho, R., Londesborough, J., Penttila, M., and Richard, P. (2003) Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*. *Appied and Environmental Microbiology* **69**, 5892-5897
5. 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**
6. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* **18**, 2125-2138
7. Rosell, A., Valencia, E., Ochoa, W. F., Fita, I., Pares, X., and Farres, J. (2003) Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase. *Journal of Biological Chemistry* **278**, 40573-40580
8. Denessiouk, K. A., Rantanen, V. V., and Johnson, M. S. (2001) Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **44**, 282-291
9. Carugo, O., and Argos, P. (1997) NADP-dependent enzymes. 1: Conserved stereochemistry of cofactor binding. *Proteins: Structure, Function, and Genetics* **28**, 10-28
10. Bellamacina, C. R. (1996) The nicotinamide dinucleotide binding motif: A comparison of nucleotide binding proteins. *FASEB Journal* **10**, 1257-1269
11. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W., and Arnold, F. H. (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochemical Journal* **468**, 475-484
12. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951

13. Dumas, R., Biou, V., Halgand, F., Douce, R., and Duggleby, R. G. (2001) Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase. *Accounts of Chemical Research* **34**, 399-408
14. Arfin, S. M., and Umbarger, H. E. (1969) Purification and properties of acetohydroxy acid isomeroreductase of *Salmonella typhimurium*. *Journal of Biological Chemistry* **244**, 1118-1127
15. Umbarger, H. E., and Davis, B. (1960). *The Bacteria* **3**, 167
16. Park, J. H., and Lee, S. Y. (2010) Fermentative production of branched chain amino acids: A focus on metabolic engineering. *Applied Microbiology and Biotechnology* **85**, 491-506
17. Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, K. M., Brynildsen, M. P., Chou, K. J. Y., Hanai, T., and Liao, J. C. (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic Engineering* **10**, 305-311
18. Atsumi, S., Hanai, T., and Liao, J. C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89
19. Hasegawa, S., Uematsu, K., Natsuma, Y., Suda, M., Hiraga, K., Jojima, T., Inui, M., and Yukawa, H. (2012) Improvement of the redox balance increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions. *Applied and Environmental Microbiology* **78**, 865-875
20. 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**, 83-89
21. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research* **31**, 365-370
22. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods in Enzymology* **154**, 367-382
23. Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences* **103**, 5869-5874
24. Un Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **40**, D71-D75
25. Goujon, M., McWilliam, H., Li, W. Z., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* **38**, W695-W699
26. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**

27. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos – A new way to display consensus sequences. *Nucleic Acids Research* **18**, 6097-6100
28. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190
29. Aulabaugh, A., and Schloss, J. V. (1990) Oxalyl hydroxamates as reaction-intermediate analogs for ketol-acid reductoisomerase. *Biochemistry* **29**, 2824-2830
30. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D - Biological Crystallography* **66**, 125-132
31. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D - Biological Crystallography* **62**, 72-82
32. Ahn, H. J., Eom, S. J., Yoon, H. J., Lee, B. I., Cho, H. J., and Suh, S. W. (2003) Crystal structure of class I acetohydroxy acid isomeroreductase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **328**, 505-515
33. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D - Biological Crystallography* **60**, 2126-2132
34. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G., and Guddat, L. W. (2012) Bacterial and plant ketol-acid reductoisomerases have different mechanisms of induced fit during the catalytic cycle. *Journal of Molecular Biology* **424**, 168-179
35. Tyagi, R., Duquerroy, S., Navaza, J., Guddat, L. W., and Duggleby, R. G. (2005) The crystal structure of a bacterial class II ketol-acid reductolsomerase: Domain conservation and evolution. *Protein Science* **14**, 3089-3100
36. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**, 194-199
37. Dumas, R., Curien, G., Derose, R. T., and Douce, R. (1993) Branched-chain-amino-acid biosynthesis in plants – Molecular cloning and characterization of the gene encoding acetohydroxy acid isomeroreductase (ketol-acid reductoisomerase) from *Arabidopsis thaliana* (thale cress). *Biochemical Journal* **294**, 821-828
38. Wierenga, R. K., De Maeyer, M. C. H., and Hol, W. G. J. (1985) Interaction of pyrophosphate moieties with α-helixes in dinucleotide-binding proteins. *Biochemistry* **24**, 1346-1357
39. Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., and Pebay-Peyroula, E. (1997) The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 angstrom resolution. *EMBO Journal* **16**, 3405-3415
40. Leung, E. W. W., and Guddat, L. W. (2009) Conformational changes in a plant ketol-acid reductoisomerase upon Mg^{2+} and NADPH binding as revealed by two crystal structures. *Journal of Molecular Biology* **389**, 167-182
41. Thomazeau, K., Dumas, R., Halgand, F., Forest, E., Douce, R., and Biou, V. (2000) Structure of spinach acetohydroxyacid isomeroreductase complexed with its reaction product dihydroxymethylvalerate, manganese and (phospho)-ADP-ribose. *Acta Crystallographica Section D - Biological Crystallography* **56**, 389-397

42. Kristan, K., Stojan, J., Adamski, J., and Rizner, T. L. (2007) Rational design of novel mutants of fungal 17 β-hydroxy steroid dehydrogenase. *Journal of Biotechnology* **129**, 123-130

43. Petschacher, B., Leitgeb, S., Kavanagh, K. L., Wilson, D. K., and Nidetzky, B. (2005) The coenzyme specificity of *Candida tenuis* xylose reductase (AKR2B5) explored by site-directed mutagenesis and X-ray crystallography. *Biochemical Journal* **385**, 75-83

44. Shiraishi, N., Croy, C., Kaur, J., and Campbell, W. H. (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**, 104-115

45. Xing, R. Y., and Whitman, W. B. (1991) Characterization of enzymes of the branched-chain amino-acid biosynthetic-pathway in *Methanococcus* spp. *Journal of Bacteriology* **173**, 2086-2092

46. Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology. *Proceedings of the National Academy of Sciences* **96**, 9459-9464

C h a p t e r 2

UNCOVERING RARE NAD-PREFERRING KETOL-ACID REDUCTOISOMERASES

Material from this chapter appears in Brinkmann-Chen S., Cahn J. K. B., and Arnold F. H. (2014). **Uncovering rare NADH-preferring ketol-acid reductoisomerases**, *Metabolic Engineering* **26**, 17-22, and is reprinted by permission from Elsevier, Inc.

Abstract

All naturally-occurring members of the ketol-acid reductoisomerase (KARI) enzyme family characterized to date have been shown to prefer the nicotinamide adenine dinucleotide phosphate (NADP) cofactor to nicotinamide adenine dinucleotide (NAD). However, KARIs with the reversed cofactor preference are desirable for industrial applications, including anaerobic fermentation to produce branched-chain amino acids. By applying insights gained from structural and engineering studies of this enzyme family to a comprehensive multiple sequence alignment of KARIs, we identified putative NAD-utilizing KARIs and characterized eight whose catalytic efficiencies using NAD were equal to or greater than NADP. These are the first naturally NAD-preferring KARIs reported and demonstrate that this property has evolved independently multiple times, using strategies unlike those used previously in the laboratory to engineer a KARI cofactor switch.

Introduction

With burgeoning genomic databases and increasing ease of gene synthesis, metabolic engineers can now readily mine nature's rich collection of enzymes. However, finding a sequence with specific desired properties can be difficult, particularly when only a few members of a protein family have been characterized and a detailed understanding of the structure-function relationship is lacking. The ketol-acid reductoisomerase (KARI, EC 1.1.1.86, also known as acetohydroxyacid isomeroreductase (AHAIR)) enzymes have attracted much interest for production of amino acids and biofuels.[1-5] These oxidoreductases catalyze the second step in the branched chain amino-acid (BCAA) biosynthesis pathway,[6] conversion of (*S*)-2-acetolactate (S2AL) to (*R*)-2,3-

dihydroxyisovalerate (RDHIV) via a methyl shift coupled to a reduction with concomitant oxidation of a nicotinamide adenine dinucleotide cofactor. The BCAA pathway is present in many organisms but not in mammals. Because of this, microbial production of branched-chain amino acids for animal feed or human supplements is a multimillion-dollar business.[7,8] The BCAA pathway has also been engineered to produce isobutanol, a potential source of renewable chemicals and fuels.[9]

All wild-type KARIs characterized and described in the literature have displayed a strong preference for nicotinamide adenine dinucleotide phosphate (NADP) over nicotinamide adenine dinucleotide (NAD).[2] Because intracellular levels of NAD are much higher than NADP, particularly under fermentative conditions, NAD-dependent oxidoreductases are strongly preferred in pathways for large-scale biocatalytic processes.[2] In the engineered isobutanol production pathway, replacement of the natural *E. coli* KARI (EcIlvC) and the alcohol dehydrogenase (ADH) with NAD-preferring engineered proteins increased the yield to nearly 100% of theoretical and improved titer and specific productivity.[2]

In the previous chapter, we used available KARI structure data to identify the amino acid residues in the β2αB-loop of the Rossmann fold that distinguish between the two cofactors.[3] A sequence alignment of Swiss-Prot-annotated KARI sequences allowed us to divide this diverse enzyme family into three groups based on β2αB-loop length (6-, 7-, and 12-residue loops) and to develop a simple recipe for switching the cofactor specificity of each major KARI enzyme subfamily from NADP to NAD. This engineering work provided valuable information on the determinants of cofactor binding and also led us to question whether nature might have already undertaken a similar engineering task to create an NAD-preferring KARI. Despite recent advances in bioinformatic cofactor specificity prediction,[10] few attempts have been made to find alternate cofactor utilization profiles within large enzyme families.[11] Because no method existed to predict the cofactor specificity of uncharacterized KARIs based on their primary sequences, we used knowledge gained from our previous work to exhaustively search known KARI sequences for KARIs with β2αB-loops predicted to improve utilization of NAD. Here, we report the discovery of the first known bispecific and naturally NAD-preferring KARIs. The catalytic efficiencies of two of the naturally NAD-preferring KARIs reported here exceed those of

any previously engineered variants. We suggest that rare proteins such as these, with properties desirable for metabolic engineering, are available in nature and may be uncovered using knowledge gleaned from structural and mutational studies.

Materials and Methods

Biological media were purchased from Research Products International (Mt. Prospect, IL, USA), NAD(P)H from Codexis, Inc. (Redwood City, CA, USA), oligonucleotides and gBlocks from Integrated DNA Technologies (San Diego, CA, USA), DNA polymerases, restriction enzymes, and T4 ligase from New England Biolabs (Ipswich, MA, USA). (*S*)-2-acetolactate (S2AL) was provided by Gevo, Inc. (Denver, CO, USA). DNA sequencing was performed by Laragen (Los Angeles, CA, USA). Standard molecular biology methods were taken from Maniatis *et al*.[12]

Cloning, variant construction, expression, and kinetic assays

The genes encoding KARIs were obtained as gBlocks, codon-optimized for *E. coli*. For each gene, the gBlocks were assembled either via PCR using T7 promoter and terminator primers and Phusion polymerase following the manufacturer's instructions or via Gibson cloning.[13] All KARIs were cloned into pET22(b)+ between restriction sites *Nde*I and *Xho*I in frame with the C-terminal His_6-tag for expression in *E. coli* BL-21(DE3) 'E. cloni Express' cells from Lucigen (Middleton, WI, USA). Heterologous protein expression and purification were conducted as described.[2] Protein concentration was determined via the Bradford assay (Bio-Rad, Hercules, CA, USA).

KARI activities were assayed by monitoring NAD(P)H consumption in the presence of (*S*)-2-acetolactate (S2AL) at 340 nm in a plate reader (Tecan Infinite M200, San Jose, CA, USA). The assay buffer contained 100 mM potassium phosphate pH 7, 1 mM DTT, 200 mM NAD(P)H, 2.5 mM S2AL, and 10 mM $MgCl_2$. For kinetic assays, we monitored the consumption of the cofactors via their fluorescence emission at 440 nm with excitation at 340 nm.

Sequence alignment

All sequences annotated as EC 1.1.1.86 and non-fragmentary were downloaded from UniProtKB[14] on 11/1/2013, a total of 8,043 sequences. The sequences were aligned using MAFFT version 7.[15] Duplicate sequences were removed, as were those with less than 20% identity to the *Slackia exigua* KARI. This left 3,383 unique KARI sequences, with an alignment length of 933 positions. The β2αB-loop was identified based on the sequences with solved structures deposited in the RCSB PDB[16] (*E. coli, S. exigua, S. oleracea, O. sativa, P. aeruginosa*, and *A. acidocaldarius*[3,17-23]). Further analysis was completed using a custom-written Python script (Supplementary Material 2-1).

Results and Discussion

To develop a recipe for changing the cofactor specificity of any KARI from NADP to NAD in the previous chapter, we used an alignment of 643 Swiss-Prot-annotated KARI sequences.[3] In this chapter, we use an expanded alignment that includes un-reviewed UniProt Knowledgebase (UniProtKB)[14] sequences and searched this expanded sequence space for β2αB-loops deviating from the classic NADP-preferring KARI motifs. To do so, all 8,043 sequences annotated as EC 1.1.1.86 from UniProtKB were aligned using MAFFT; duplicate and low-identity sequences were removed as described in the Methods section. The β2αB-loop and its context were then identified based on KARI crystal structures.[3,17-23] The distribution of amino acids at each position, including positions present in only a very small number of sequences, is shown in Figure 2-1. The four phosphate-binding residues are highly conserved, but there is considerable variation at most of the other positions of the loop as well as in the length of the loop. Figure 2-2 shows these positions in the NADP co-crystal structure of *Slackia exigua* KARI obtained in the previous chapter. SeKARI is a 7-residue loop KARI whose β2αB-loop is near the consensus sequence.

As previously reported,[3] KARIs can be divided into three sub-families based on the length of the β2αB-loop. Of the 3,383 unique KARIs in our expanded alignment, 2,262 (66.8%) had 7-residue loops, 593 (17.5%) had 6-residue loops, and 512 (15.1%) had 12-residue loops, which is in good agreement with previous alignment results. Eleven sequences were missing the β2αB-loops (these were unlabeled fragmentary sequences), leaving just six KARIs with other loop lengths.

Based on previous structural studies and mutational analysis,[4,9,24] we have proposed that cofactor specificity is controlled by up to four amino acid residues in the loop as outlined in Chapter 1: the first two positions, the antepenultimate position, and the ultimate position, as indicated by red arrows in Figure 2-1 and labeled for the *S. exigua* KARI in Figure 2-2. We searched the expanded sequence alignment for proteins having an acidic residue at any of these four positions. In our previous work, introduction of acidic residues into the β2αB-loop was crucial for switching the cofactor specificity;[2,3] this has also been shown to be a major determinant of cofactor specificity in nature[25-28] and in engineered proteins.[29] For the six KARI sequences with non-standard β2αB-loop length we could not predict which residues would interact directly with the cofactor. We therefore visually inspected those loops for the presence of multiple acidic residues in their N-terminal or C-terminal ends. Ultimately, from the alignment of 3,383 KARI sequences, 58 were identified by this method as being possibly NAD-preferring, including four of the six KARIs with nonstandard β2αB-loop lengths. To avoid over-representing KARIs from closely related organisms, the full sequences of these KARIs were clustered using the UniProt alignment tool, and for each of the 26 clusters one representative sequence was selected to create the list of putative NAD-dependent KARIs shown in Figure 2-3. Most of the clusters were singletons or contained only a handful of sequences; the two largest clusters are represented by *Metallosphaera sedula* and *Sulfolobus islandicus*.

From this list, we successfully expressed, purified, and assayed the activity toward both cofactors of eight of the putative NAD-utilizing KARIs, highlighted in Figure 2-3. The kinetic values in Table 2-1 show that none of these KARIs has the typical preference for NADP. Four of the eight – *Hydrogenobaculum* sp. KARI (HsKARI), *Ignisphaera aggregans* KARI (IaKARI), *Metallosphaera sedula* KARI (MsKARI), and *Syntrophomonas wolfei* KARI (SwKARI) – showed roughly equal catalytic efficiency using NADPH and NADH (catalytic efficiency ratio of NADH/NADPH ~ 1), while four – *Archaeoglobus fulgidus* KARI (AfKARI), *Desulfococcus oleovorans* KARI (DoKARI), *Thermacetogenium phaeum* KARI (TpKARI), and an uncultured archaeon KARI (UaKARI) – were NAD-preferring, with catalytic efficiency ratios (NADH/NADPH) ranging from 13 to 152 (Table 2-1). The NADH K_M or K_H values of these KARIs were equal to or lower than the NADPH values of the NADPH-preferring KARI from *E. coli*, EcIlvC. IaKARI in particular stood out, with very low K_M values for both cofactors.

Two of the naturally NAD-preferring KARIs, TpKARI and UaKARI, have catalytic efficiencies on NADH that exceeded the best engineered NAD-preferring KARI, EcIlvC$^{P2D1-A1}$, by 2.8-fold and 1.2-fold, respectively. Both TpKARI and UaKARI have very low (~1 mM) K_H for NAD (Table 2-1). These features render these two excellent candidates for anaerobic BCAA or isobutanol fermentations. IaKARI may also be of value due to its ability to utilize both cofactors and its good activity at low cofactor concentrations.

In the NAD-preferring KARIs engineered in the previous chapter,[2,3] most, if not all, NADP-dependent activity was abolished by the mutations. In contrast, nature achieved NAD cofactor utilization with NADPH K_M or K_H values equivalent to NADP-preferring KARIs, leading to lower specificities than many of our engineered KARIs.[3] We also demonstrated in the previous chapter that NADP-preferring KARIs can have their specificity reversed by mutation of the ultimate and antepenultimate residues of the β2αB-loop to aspartate in the case of the 7- or 12-residue loops, or, in the case of the 6-residue loop, mutation of the ultimate position to aspartate and of the second to proline. Interestingly, none of the naturally NAD-utilizing KARIs we found follow this recipe. Aspartate is present in only four out of 26 ultimate positions and one out of 21 antepenultimate positions, although SwKARI has a glutamate at both positions. Of all 26 sequences, only SwKARI has acidic residues at more than one of the four key positions. Only two of the sequences have proline at the second position, and neither of those KARIs has a 6-residue β2αB-loop or an aspartate at the ultimate position. These differences may reflect different selective pressures (e.g., specificity reversal in the engineered enzymes versus utilization of NAD in the natural enzymes). They could also arise from the very different evolutionary paths taken.

Eight of the 26 putative NAD-utilizing KARIs possess acidic residues at the first position of the loop, six at the second, three at the antepenultimate position, and five at the ultimate position. It is interesting that the plurality of these putative NADH-utilizing KARIs have acidic residues the first position of the β2αB-loop. Although none of our previous cofactor-specificity reversals required an acidic residue at this position, Hasegawa *et al*. introduced a mutation to glutamate at this position, suggested by comparison with the NAD-dependent dihydrolipoamide dehydrogenases, in an attempt to reverse the cofactor

specificity of the *C. glutamicum* KARI.[4] The resulting mutant enzyme was specific for NAD, but had less activity with NADH than the wild-type enzyme. It is also worth noting that all five KARIs with 6-membered β2αB-loops have an aspartate at the antepenultimate position, but that the structure of the KARI from *Alicyclobacillus acidocaldarius*,[23] an NADP-preferring enzyme with a 6-membered β2αB-loop which will be discussed in Chapter 4, shows this position to be structurally homologous with the non-interacting pre-antepenultimate position of the 7-membered loop KARIs. Thus it is unclear whether and how this aspartate would interact with the cofactor. Outside the β2αB-loop, the 58 putative NAD-utilizing KARIs have no unique conserved residues.

All of the KARIs described here are members of the class I (short chain) KARI subfamily, which is believed to be evolutionarily older than the class II (long chain) KARIs such as *E. coli* KARI,[17] and all originate from microbial hosts adapted to extreme conditions (thermophiles, acidophiles, and halophiles).[30-35] Additionally, with the exception of the archaeote *M. sedula*, all hosts are anaerobes. However, adaptation to anoxic environments does not automatically imply the presence of an NAD-preferring KARI, as among the thousands of sequences there are many NADP-preferring KARIs from anaerobe hosts. For each of the eight characterized KARIs, we examined the three KARIs with the highest overall sequence identity and predicted their cofactor specificity using the sequence of their cofactor binding loops (Table 2-2). For six of the eight newly characterized KARIs, the majority of these nearest neighbors, which generally came from closely related organisms, have loops sequences that suggest a conventional NADP binding mode. For the remaining two KARIs, IaKARI and MsKARI, which are 64% identical to each other, the nearest neighbors have very similar sequences and were filtered out during the clustering step. That these eight enzymes have used differing sets of substitutions for cofactor utilization and that each is phylogenetically isolated from the others but close to non-reversed sequences suggests that NAD-utilization has arisen independently in KARIs of several different organisms.

In many protein families, the few characterized representatives have come from commonly studied mesophilic organisms. The properties of these enzymes are often assumed to be representative of the enzyme family as a whole. However, a rational, knowledge-guided search of all available sequences can uncover novel properties in more

distantly related organisms, many of which can be valuable tools for metabolic engineers. In this study, we have shown that the presence of acidic residues at conserved phosphate-binding positions can be used to identify candidate genes encoding NAD-preferring proteins in the industrially important KARI enzyme family. With minor modifications, we expect this approach to be useful in finding enzymes with differing cofactor requirements in other oxidoreductase families.

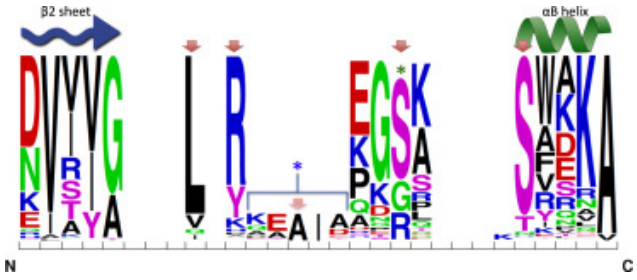


Figure 2-1. The distribution of amino acids at each position of the β2αB-loop and flanking structures, as aligned by MAFFT. Gaps show positions that have amino acids in only a small number of the 3,383 sequences considered. The β2-sheet and αB-helix are indicated. Red arrows indicate the positions known to interact with the NADP phosphate in published crystal structures. The green asterisk marks the residue present in the 7-residue-loop KARIs but absent in the 6-residue loops. The blue asterisk shows the residues present in the 12-residue loop but absent in the 7- and 6-residue loops. Logo was made using Berkeley WebLogo.[36]

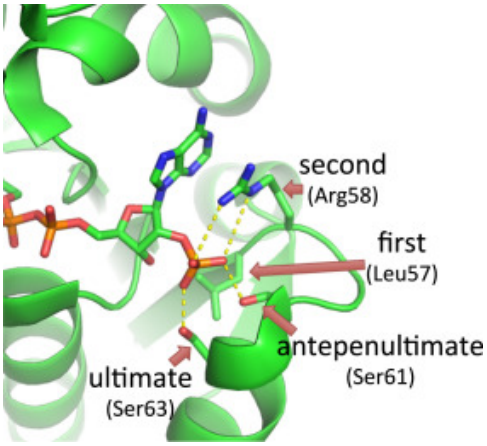
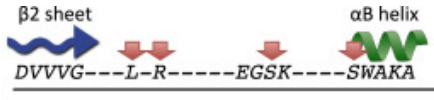


Figure 2-2. The structure of the cofactor-binding site of *Slackia exigua* KARI with bound NADPH (PDB 4KQW[3]), showing the canonical NADP-binding motif of four key residues. The *S. exigua* KARI sequence is close to the consensus sequence.




```
DVVVG---L-R-----EGSK----SWAKA                    Consensus of KARIs

 PVRVG---N-I-----DD-A----YRQRA   H6RL58_BLASD Blastococcus saxobsidens
 AVRVG---N-I-----DD-R----YFELA   A3NGR4_BURP6 Burkholderia pseudomallei
 NVVVG---D-L-----LGNE----YSKRA   G4RLE5_THETK Thermoproteus tenax
 NVIIG---N-I-----ED-E----YANRA   E1QUK3_VULDI Vulcanisaeta distributa
 EVIVG---N-V-----ND-E----YRRRA   D9PYY3_ACIS3 Acidilobus saccharovorans
 NVIIG---N-V-----KD-R----YHDLA   C4KG48_SULIK Sulfolobus islandicus
 NVIVG---QSK-----QFQR----DWDRA   B1ZV88_OPITP Opitutus terrae
*EVIIG---Q-L-----EGDA----YWEKA   A8ZTR0_DESOH Desulfococcus oleovorans
 NVIIG-AGN-K-----DRYP----DWENA   D8GTE6_CLOLD Clostridium ljungdahlii
 NVTVG---L-E-----RQGK----SWEKA   Q6L044_PICTO Picrophilus torridus
 NVKLG---L-E-----REGN----SWKKA   S0AM07_FERAC Ferroplasma acidarmanu
*NVVVG---L-E-----RQGD----SWRRA   E0SRA9_IGNAA Ignisphaera aggregans
 NVIVG---L-E-----RKGN----SWILA   L0A7Y1_CALLD Caldisphaera lagunensis
*QVTVG---L-E-----REGK----SWEQA   ILVC_METS5   Metallosphaera sedula
*NVVIG---E-TEILGGNKNP----SWEKA   D1JA89_9ARCH uncultured archaeon
 DVVVG---L-P-----EGNE----DRPVA   C7NTA6_HALUD Halorhabdus utahensis
 NVIIS---E-K-----EGTA----NYKLA   K1X9E2_9BACT uncultured bacterium
 DVIAA---D-L-----KDSP----AWKRA   ILVC_METTP   Methanosaeta thermophila
 DVIVS---E-L-----EGTE----NYELA   D9QTJ4_ACEAZ Acetohalobium arabaticum
*NVVVG---L-P-----EWDKA---TWERA   ILVC_ARCFU   Archaeoglobus fulgidus
*RVIVA---L-D-----DKSP----HRKTA   ILVC_HYDS0   Hydrogenobaculum sp.
 NVIIG---L-R-----RDLP----DWDKA   D5XAE9_THEPJ Thermincola potens
*KVVVA---D-I-----PSSE----NWKKA   K4LVZ1_THEPS Thermacetogenium phaeum
*NVIVG---L-RKPFDEASEK----EWNAV   ILVC_SYNWW   Syntrophomonas wolfei subsp. wolfei
 DVLVS---E-I-----PNTP----NYELA   F2NIE5_DESAR Desulfobacca acetoxidans
 EVVVS---D-R-----AGSE----NFEKA   A0LMK6_SYNFM Syntrophobacter fumaroxidans
```

Figure 2-3. An alignment of the 26 KARI sequences identified as putatively NAD-utilizing, as compared to the consensus sequence of 3,383 KARIs (top). Sequences shown in bold and with asterisks were experimentally characterized; blue indicates a bispecific KARI and red a NAD-preferring KARI. For each KARI, the UniProt identifier is shown along with the host name. The secondary structural elements are shown at top, along with the four key positions for phosphate binding (red arrows). Some of the sequences have been manually realigned.

Table 2-1. Biochemical characterization of bispecific and naturally NAD-specific KARIs and comparison to wild-type *E. coli* KARI (EcIlvc) and engineered EcIlvc^{6E6} and EcIlvc$^{P2D1\text{-}A1}$.[2]

Enzyme	K_M or K_H (μM)		k_{cat} (s^{-1})		k_{cat}/K_M or K_H (mM^{-1} s^{-1})		Ratio
	NADH	NADPH	NADH	NADPH	NADH	NADPH	
EcIlvc	1075	41	0.3	3.6	0.3	88	0.0003
EcIlvc^{6E6}	30	650	2.3	0.20	74	0.40	185
EcIlvc$^{P2D1\text{-}A1}$	26	> 1400	4.3	0.54	165	< 0.4	> 412
HsKARI	39	46	0.12	0.12	3.2	2.7	1.2
IaKARI	< 1	< 1	0.02	0.03	> 20	> 25	~ 0.8
MsKARI	24	31	0.06	0.07	2.5	2.1	1.2
SwKARI	57	44	0.28	0.22	5.0	5.0	1.0
AfKARI	5	26	0.1	0.04	20	1.5	13
DoKARI	32	n.a.	0.25	n.a.	8.0	n.a.	-
TpKARI	< 1	40	0.46	0.25	460	6.0	7.4
UaKARI	1.1	38	0.22	0.05	200	1.3	152

All enzymes were His_6-tagged and purified prior to characterization. Enzyme activities were determined in 100 mM potassium phosphate pH 7 with 1 mM DTT, 200 μM NADPH or NADH, 2.5mM *S*2AL, and 10 mM $MgCl_2$. Concentrations of the purified enzymes were determined using the Bradford assay. The K_M or K_H values for the cofactors, corresponding to the concentration of the half-maximum activity, were measured with appropriate dilutions of NADPH and NADH in the presence of saturating concentrations of substrate S2AL. Mutations located within the cofactor-binding pocket of EcIlvC^{6E6}: A71S, R76D, S78D, and Q110V. Additional mutations in EcIlvC$^{P2D1\text{-}A1}$: D146G, G185R, and K433E. For standard errors and Hill coefficients, refer to Table S2-1. n.a. = not active.

Table 2-2

Nearest neighbors of the eight KARIs characterized in this work. Results are limited to the top three hits, and the cofactor binding loop and predicted cofactor specificity are shown.

KARI (β2αb-loop sequence)	% ID	Organism	Oxygen requirements[a]	β2αb-loop sequence	Predicted cofactor specificity
DoKARI (QLEGDAY)	86	Uncultured *Desulfobacterium*	Anaerobic	QMEGDAY	NADH
	80	*Desulfococcus multivorans*	Anaerobic	QREGGAS	NADPH
	78	*Desulfatibacillum alkenivorans*	Anaerobic	QRDGGKS	NADPH
SwKARI (LRKPFDEASEKE)	73	*Syntrophothermus lipocalidus*	Anaerobic	LRKPEDDFTTAE	Bispecific
	72	*Desulfitobacterium dehalogenans*	Anaerobic	LRADSRR	NADPH
	71	*Syntrophobotulus glycolicus*	Anaerobic	LRKDSSR	NADPH
AfKARI (LPEWDKAT)	77	*Archaeoglobus veneficus*	Anaerobic	LYKGSRS	NADPH
	77	*Archaeoglobus profundus*	Anaerobic	LYKGSKS	NADPH
	76	*Archaeoglobus sulfaticallidus*	Anaerobic	DKKGTRN	unclear
TpKARI (DIPSSEN)	72	*Caldicellulosiruptor saccharolyticus*	Anaerobic	LYHGSKS	NADPH
	72	*Caldicellulosiruptor hydrothermalis*	Anaerobic	LYHGSKS	NADPH
	71	*Caldicellulosiruptor obsidiansis*	Anaerobic	LYQGSKS	NADPH
UaKARI (ETEILGGNKNPS)	98	Uncultured archaeon	Anaerobic	ETEILGGNKNPS	NADH
	64	*Akkermansia muciniphila*	Anaerobic	VRPGKS	NADPH
	52	*Paenibacillus lactis*	Facul. anaerobic	LREGKS	NADPH
IaKARI (LERQGDS)	73	*Sulfolobus solfataricus*	Facul. anaerobic	LEREGKS	Bispecific
	73	*Sulfolobus islandicus*	Facul. anaerobic	LEREGKS	Bispecific
	69	*Acidianus hospitalis*	Aerobic	LEREGNS	Bispecific
MsKARI (LEREGKS)	90	*Metallosphaera cuprina*	Aerobic	LEREGKS	Bispecific
	88	*Metallosphaera yellowstonensis*	Aerobic	LEREGKS	Bispecific
	80	*Acidianus hospitalis*	Aerobic	LEREGNS	Bispecific
HsKARI (LDDKSPH)	79	*Thermocrinis albus*	Microaerophilic	LPAGSKS	NADPH
	77	*Hydrogenobacter thermophilus*	Aerobic	LPEGSKS	NADPH
	77	*Hydrogenivirga* sp.	Microaerophilic	LHEKSRS	NADPH

% ID: percent identity of the nearest neighbor to its respective KARI.

[a]Oxygen requirements of the host organism were obtained from various microbiological databases.

References

1. Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, K. M., Brynildsen, M. P., Chou, K. J. Y., Hanai, T., and Liao, J. C. (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic Engineering* **10**, 305-311
2. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metabolic Engineering* **13**, 345-352
3. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951
4. Hasegawa, S., Uematsu, K., Natsuma, Y., Suda, M., Hiraga, K., Jojima, T., Inui, M., and Yukawa, H. (2012) Improvement of the redox balance increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions. *Applied and Environmental Microbiology* **78**, 865-875
5. Liu, X. H., Chen, P. Q., Wang, B. L., Dong, W. L., Li, Y. H., Xie, X. Q., and Li, Z. M. (2010) High throughput receptor-based virtual screening under ZINC database, synthesis, and biological evaluation of ketol-acid reductoisomerase inhibitors. *Chemical Biology and Drug Design* **75**, 228-232
6. Chunduru, S. K., Mrachko, G. T., and Calvo, K. C. (1989) Mechanism of ketol acid reductoisomerase - Steady-state analysis and metal-ion requirement. *Biochemistry* **28**, 486-493
7. Becker, J., and Wittmann, C. (2012) Systems and synthetic metabolic engineering for amino acid production - The heartbeat of industrial strain development. *Current Opinion in Biotechnology* **23**, 718-726
8. Vogt, M., Haas, S., Klaffl, S., Polen, T., Eggeling, L., van Ooyen, J., and Bott, M. (2014) Pushing product formation to its limit: Metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metabolic Engineering* **22**, 40-52
9. Atsumi, S., Hanai, T., and Liao, J. C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89
10. Geertz-Hansen, H. M., Blom, N., Feist, A. M., Brunak, S., and Petersen, T. N. (2014) Cofactory: Sequence-based prediction of cofactor specificity of Rossmann folds. *Proteins: Structure, Function, and Bioinformatics* **82**, 1819-1828
11. Di Luccio, E., Elling, Robert A., and Wilson, David K. (2006) Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochemical Journal* **400**, 105-114
12. Sambrook, J., Frisch, E. F., Maniatis, T. (1989) *Molecular Cloning: A laboratory manual.*, Cold Spring Harbor Laboratory Press, New York

13. Gibson, D. G. (2011) Chapter Fifteen - Enzymatic assembly of overlapping DNA fragments. in *Methods in Enzymology* (Christopher, V. ed.), Academic Press. pp 349-361
14. UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Research* **41**, D43-D47
15. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780
16. Bairoch, A., and Apweiler, R. (1999) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Research* **27**, 49-54
17. Ahn, H. J., Eom, S. J., Yoon, H. J., Lee, B. I., Cho, H. J., and Suh, S. W. (2003) Crystal structure of class I acetohydroxy acid isomeroreductase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **328**, 505-515
18. Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., and Pebay-Peyroula, E. (1997) The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 angstrom resolution. *EMBO Journal* **16**, 3405-3415
19. Leung, E. W. W., and Guddat, L. W. (2009) Conformational changes in a plant ketol-acid reductoisomerase upon Mg^{2+} and NADPH binding as revealed by two crystal structures. *Journal of Molecular Biology* **389**, 167-182
20. Thomazeau, K., Dumas, R., Halgand, F., Forest, E., Douce, R., and Biou, V. (2000) Structure of spinach acetohydroxyacid isomeroreductase complexed with its reaction product dihydroxymethylvalerate, manganese and (phospho)-ADP-ribose. *Acta Crystallographica Section D - Biological Crystallography* **56**, 389-397
21. Tyagi, R., Duquerroy, S., Navaza, J., Guddat, L. W., and Duggleby, R. G. (2005) The crystal structure of a bacterial class II ketol-acid reductolsomerase: Domain conservation and evolution. *Protein Science* **14**, 3089-3100
22. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G., and Guddat, L. W. (2012) Bacterial and plant ketol-acid reductoisomerases have different mechanisms of induced fit during the catalytic cycle. *Journal of Molecular Biology* **424**, 168-179
23. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W., and Arnold, F. H. (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochemical Journal.* **468**, 475-484
24. 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**, 83-89
25. Baker, P. J., Britton, K. L., Rice, D. W., Rob, A., and Stillman, T. J. (1992) Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold: Implications for nucleotide specificity. *Journal of Molecular Biology* **228**, 662-671

26. Carugo, O., and Argos, P. (1997) NADP-Dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins: Structure, Function, and Bioinformatics* **28**, 10-28
27. Pletnev, V. Z., Weeks, C. M., and Duax, W. L. (2004) Rational proteomics II: Electrostatic nature of cofactor preference in the short-chain oxidoreductase (SCOR) enzyme family. *Proteins: Structure, Function, and Bioinformatics* **57**, 294-301
28. Van Petegem, F., De Vos, D., Savvides, S., Vergauwen, B., and Van Beelumen, J. (2007) Understanding nicotinamide dinucleotide cofactor and substrate specificity in class I flavoprotein disulfide oxidoreductases: Crystallographic analysis of a glutathione amide reductase. *Journal of Molecular Biology* **374**, 883-889
29. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* **18**, 2125-2138
30. Auernik, K. S., Maezato, Y., Blum, P. H., and Kelly, R. M. (2008) The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon *Metallosphaera sedula* provides insights into bioleaching-associated metabolism. *Applied and Environmental Microbiology* **74**, 682-692
31. Beeder, J., Nilsen, R. K., Rosnes, J. T., Torsvik, T., and Lien, T. (1994) *Archaeoglobus fulgidus* isolated from hot North Sea oil field waters. *Applied and Environmental Microbiology* **60**, 1227-1231
32. Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000) *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *International Journal of Systematic and Evolutionary Microbiology* **50**, 1601-1609
33. McInerney, M. J., Bryant, M. P., Hespell, R. B., and Costerton, J. W. (1981) *Syntrophomonas wolfei* gen. nov. sp. nov., an anaerobic, syntrophic, fatty acid-oxidizing bacterium. *Applied and Environmental Microbiology* **41**, 1029-1039
34. Niederberger, T. D., Götz, D. K., McDonald, I. R., Ronimus, R. S., and Morgan, H. W. (2006) *Ignisphaera aggregans* gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *International Journal of Systematic and Evolutionary Microbiology* **56**, 965-971
35. Romano, C., D'Imperio, S., Woyke, T., Mavromatis, K., Lasken, R., Shock, E. L., and McDermott, T. R. (2013) Comparative genomic analysis of phylogenetically closely related *Hydrogenobaculum* sp. isolates from Yellowstone National Park. *Applied and Environmental Microbiology* **79**, 2932-2943
36. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190

Chapter 3

DISCOVERY OF AN NAD-PREFERRING XYLOSE REDUCTASE

Abstract

The inherent cofactor imbalance in the xylose uptake pathway has made D-xylose reductase a prime target of cofactor specificity engineering. In the previous chapter we demonstrated that an understanding of the structure-function relationship in the NADP cofactor specificity-determining loop of ketol-acid reductoisomerases enabled discovery of rare homologs that prefer NAD. Here we used this knowledge-based approach to identify a naturally-occurring NAD-dependent xylose reductase, which is of interest for engineering xylose metabolism and efficient microbial utilization of plant biomass.

Introduction

Enzymes that catalyze oxidation and reduction reactions primarily use nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) to transport and store hydride reducing equivalents. Though the chemically active nicotinamide group is the same in both molecules, phosphorylation of the adenine ribose moiety allows enzymes to distinguish between the two pools of reducing equivalents, which is valuable for regulation of metabolism and partitioning of pathways.

Rectifying imbalances in the production and depletion of these cofactors, however, can be a significant hurdle in the engineering of metabolic pathways. This has led to a significant body of protein engineering work focused on adjusting the cofactor utilization of enzymes to the reducing equivalents available within the cell. Engineering is required when the natural homologs of the target enzyme all have the same cofactor preference, as is widely believed to be the case for many important enzyme classes. In the previous chapter, we used knowledge gained from enzyme structural and engineering studies, to uncover naturally NAD-utilizing ketol-acid reductoisomerases (KARIs, EC 1.1.1.86), an enzyme class previously believed to be uniformly NADP-preferring.[1] To do this, we identified the residues involved in binding the NADP phosphate based on the structures of enzymes representing each of three recurrent motifs. Because acidic residues are enriched in the adenine ribose-binding pockets of natural NAD-binding proteins and depleted in those of

natural NADP-binding proteins, and because nearly all engineered NADP-to-NAD cofactor specificity switches have involved the addition of acidic residues, we looked for natural sequences having acidic residues at the positions normally used for phosphate binding. Of 58 such KARIs identified, we experimentally characterized eight, all of which used NAD as well as or better than NADP.

The enzyme D-xylose reductase (EC 1.1.1.307, XR) has received more attention than any other enzyme as a target for switching cofactor preference. XR catalyzes the first step in the isomerization of xylose to xylulose, which is required for the entry of the sugar into metabolism. However, this isomerization pathway represents a major bottleneck in engineered xylose utilization because XR prefers NADP whereas the next enzyme, xylitol dehydrogenase (XDH), has a strict preference for NAD (Figure 3-1a). This was recognized as a problem as early as 1983,[2] and many efforts have focused on reversing the cofactor specificity of one of the enzymes in order to produce a cofactor-closed loop, where each enzyme regenerates the cofactor required by the other.[3-21] These engineering efforts were universally motivated by the belief that "using known natural enzymes, the assembly of a chimeric pathway in which XR and XDH show exactly comparable utilization of NADP(H) and NAD(H) appears to be currently out of reach."[16] Only one naturally NAD-preferring XR, from *Candida parapsilosis*, discovered by screening of fungal strains for xylitol production, has been reported.[22,23] To our knowledge, however, this XR has not been used in metabolic engineering, highlighting the need for more candidate NAD-utilizing enzymes. Because XRs can come from a diverse assortment of organisms, experimental screening poses significant challenges. The structure-based approach we developed for the KARI enzymes, however, can easily evaluate sequences from a broad range of organisms and identify candidates for experimental verification.

Materials and Methods

Genes encoding the selected XRs were codon-optimized for *E. coli* expression using GeneDesigner software (DNA 2.0) or the IDT Codon Optimization Tool (Integrated DNA Technologies) and synthesized as gBlocks (Integrated DNA Technologies). Each was inserted into a pET22b(+) vector using standard Gibson ligation and the resulting plasmids were used to transform BL-21(DE3) *E. coli*. Initial characterization of these putative XRs was performed in lysate.

For characterization, at least four 700 μL Luria Broth cultures in 96-well blocks were inoculated with saturated overnight culture and grown at 37 °C for 3 hours before induction with IPTG to a final concentration of 500 μM. After inoculation, the cultures were held at 20 °C for 24 hours prior to centrifugal harvesting. The cell pellets were frozen at -20 °C until assayed. After thawing, the cell pellets were resuspended in 200 μL of a buffer containing 50 mM sodium phosphate pH 7.0, 750mg/L hen egg white lysozyme, and 10mg/L DNase I. XR activity was assayed by monitoring the depletion of NAD(P)H using A340 on a Tecan plate reader. The assay buffer contained 400 mM D-xylose (Alfa Aesar) and 300 μM NADH (Sigma Aldrich) or NADPH (Codexis).

For expression, 500 mL of Terrific Broth medium was inoculated with 2 mL of saturated overnight culture and grown at 37 °C for four hours before expression and harvest as before. After thawing, the cell pellet was resuspended in 25 mL of a buffer containing 20 mM Tris, 100 mM NaCl, and 20 mM imidazole and lysed by sonication. The lysate was clarified on the centrifuge, and MbmXR was purified from the supernatant using an ÄKTA FPLC (GE Healthcare) and a His-Trap column (GE Healthcare). The protein was buffer-exchanged twice into half a volume of 20 mM Tris buffer pH 7 using an Centricon filter (Amicon), with a final yield of approximately 0.9 mg protein/mL culture. XR activity was assayed by monitoring the depletion of NAD(P)H using A340 on a Tecan plate reader. For determining cofactor K_M values, where low concentrations could not be measured using absorbance, the fluorescence of the reduced nicotinamide was used, exciting at 340 nm and measuring emission at 440 nm, as described.[1]

Results and Discussion

Compared to the highly diverse KARI family, the key phosphate-binding loop of xylose reductase is more conserved, with no major deviations in length (Figure 3-1b). As exemplified by the extensively studied XR of *Candida tenuis* (Figure 3-1c), four largely conserved residues determine specificity for NADP. From an initial set of 721 XR enzymes downloaded from UniProt and aligned with Clustal Omega,[24] we searched for sequences possessing acidic residues at these positions, finding five (Figure 3-1d). Because three shared the same four cofactor binding residues (differing only by an N6D substitution relative to the consensus sequence), we selected only one (BspXR) for experimental characterization. We also selected the two unique sequences, MbmXR and RmcXR. For

comparison, we also expressed three previously-characterized XRs, the highly NADP-preferring TemXR,[19] the "bispecific" CteXR,[25] and the one XR known to be NAD-specific, CpaXR.[23]

Of the new enzymes tested, only MbmXR shows faster consumption of NADH than of NADPH (Table 3-1). The other enzymes, BspXR and RmcXR, are more active with NADP, despite having acidic amino acids at the putative specificity-determining positions. The MbmXR enzyme has the N6D substitution (numbered based on the alignment in Figure 3-1b) common to three of the others, but also possesses an R10M substitution that eliminates two potential hydrogen bonds to the NADP phosphate. This R10M substitution seems to be particularly important to specificity because BspXR, which lacks it, shows a clear preference for NADP. Additionally, RmcXR, which also is two substitutions away from the consensus sequence (Figure 3-1d) but still possesses R10, is bispecific with a preference for NADP. However, all of the prior attempts to engineer xylose reductases have focused on the positions corresponding to K4 and N6; only Liang and coworkers included the R10 position in their protein engineering strategy.[13] Their resulting XR (with R10F) had the greatest specificity towards NADH of any engineered XR, albeit with poor catalytic efficiency. Out of 134 XRs for which the homologous position can be identified, 81% have R and only MbmXR has M. However, 9.7% have other hydrophobic amino acids.

We scaled up the expression of MbmXR and purified the enzyme for further studies. As was observed in lysate, MbmXR has a pronounced preference for NAD over NADP, with a ratio of k_{cat}/K_M(NADH) to k_{cat}/K_M(NADPH) of approximately 4.6 (Table 3-2). This distinction comes entirely from a difference in K_M, as the k_{cat} values are similar for both cofactors. The xylose K_M is approximately 400 mM, which is higher than that reported for other XRs.[6,13,19-21] However, the catalytic efficiency of MbmXR towards NAD exceeds by an order of magnitude that of any engineered NAD-preferring XR (Table 3-3) for which data are available.

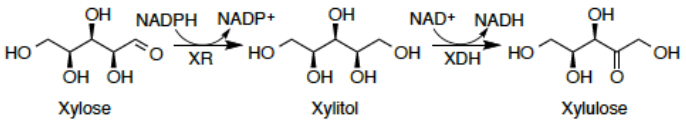
The catalytic efficiency of MbmXR towards NAD is roughly equivalent to that of the NAD-dependent CpaXR;[23] however, its specificity for NAD is less. Interestingly, CpaXR is not predicted to be NAD-utilizing by our approach because it lacks negatively-charged amino acids in the structurally relevant positions. Instead, it has only a single,

relatively conservative K4R substitution from consensus. This mutation has come up in multiple efforts to engineer specificity,[5,20] but in a crystal structure of the mutant CteXR enzyme the residue points away from the cofactor, with no clear structural role in conferring specificity.[6]

In a previous study, Wilson and co-workers attempted to predict naturally NAD-preferring members of the aldo-keto reductase (AKR) enzyme superfamily, which contains XR.[26] In that study, they hypothesize that a glutamate residue on a helix near the cofactor is responsible for the bispecificity of CteXR based on structures of that protein bound to both cofactors.[25] On this basis, they select nine proteins for expression and discover a new NAD-dependent methylglyoxal reductase. However, in our set of XRs, 95% have this glutamate residue, with an additional 4% having aspartate. Since nearly all XRs are NADP-preferring, this method is of negligible utility in the prediction of cofactor specificity of XRs. Clearly, any attempt to assign specificity to a single residue or set of residues necessarily oversimplifies the complicated nature of specificity determination within the dynamic and complex binding pockets of proteins.

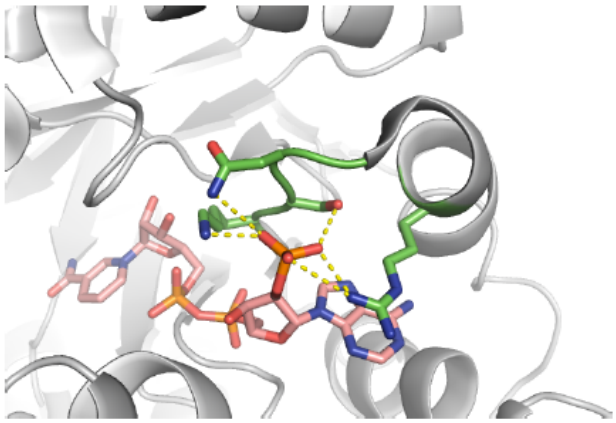
Unlike in KARIs, where 58 putative NAD-utilizing homologs could be identified, we were only able to find one bispecific XR and one NAD-utilizing XR in this set of 721. This may be due to the smaller number of sequences available for search (about 10-fold fewer), but is more likely due to the related issue of phylogenetic distribution. KARIs are present in plants, bacteria, archaea, and fungi, while XR is known only in fungi. Therefore, XRs may not have had sufficient opportunities to evolve NAD utilization, as we postulated occurred multiple times in KARIs. Furthermore, the NAD-utilizing KARIs primarily came from organisms under heightened selective pressures,[1] i.e., extremophiles, which are rare among fungi. *Marssonina brunnea*, in particular, is neither; instead it is a fungus which specifically infects the leaves of poplar trees and is the causative pathogen of the Marssonina leaf spot disease.[27] It is unclear what selective advantage, if any, NAD-dependent xylose reduction has in this context. Nevertheless, MbmXR serves as a useful demonstration of the fact that the diversity of nature can often provide engineers with access to useful enzyme properties, and that a proper understanding of the sequence-structure-function relationships obtained with the help of protein engineering can be used to uncover them.

(a)



(b)

(c)



(d)

UNP AC	Species	Name	Loop
	Consensus		VIPKSNNPERLAQ
*K1WWE5	*M. brunnea m.*	MbmXR	VVPKSDTWEMLVE
*V5F9I4	*B. spectablis*	BspXR	VIPKSDKAHRLAE
G3APT8	*S. passalidarum*	SpsXR	IIPKSDTPERLVE
S7RXD8	*G. trabeum*	GtrXR	VIPKSDDRLRMEE
*D6R259	*R. mucilaginosa*	RmcXR	IIPESVTPSRIES
*Q6YOZ3	*C. parapsilosis*	CpaXR	VIPRSNSPDRMAQ
*O74237	*C. tenuis*	CteXR	VIPKSNLPERLVQ
*C5J3R6	*T. emersonii*	TemXR	VIPKSNNPGRLAQ

Figure 3-1. The two-step enzymatic conversion of xylose to xylulose by xylose reductase (XR) and xylitol dehydrogenase (XDH) (a), showing an imbalance of cofactor utilization. The amino acid distribution in and around the cofactor specificity-determining loop of an alignment of 724 xylose reductases (b). Orange bars indicate the phosphate-binding amino acids. The phosphate binding mode of CteXR (c), showing the four canonical interactions with the phosphate, from PDB 1K8C.[28] An alignment of the sequences of the cofactor specificity-determining loops of the xylose reductases discussed in this study (d). Asterisks mark enzymes experimentally characterized. UniProt accession codes are provided for each. Phosphate-binding amino acids (or homologous positions) are shown in orange; deviations from conserved consensus residues are highlighted with grey.

Table 3-1. Relative cofactor consumption rates by xylose reductases, as measured in *E. coli* lysate.

	MbmXR	BspXR	RmcXR	CpaXR	TemXR	CteXR
NADH consumption rate as percent of NADPH consumption rate	227 ± 18%	14 ± 5 %	82 ± 6 %	186 ± 10%	5 ± 4 %	93 ± 31 %

Table 3-2. Kinetic parameters for purified MbmXR.

k_{cat} (s^{-1}) [NADH]	K_M (μM) [NADH]	k_{cat} / K_M (s^{-1} mM^{-1}) [NADH]	k_{cat} (s^{-1}) [NADPH]	K_M (μM) [NADPH]	k_{cat} / K_M (s^{-1} mM^{-1}) [NADPH]	K_M (mM) D-xylose
575 ± 26	49 ± 18	11800 ± 4080	648 ± 49	253 ± 90	2560 ± 926	402 ± 60

Table 3-3. Catalytic efficiency of MbmXR compared to engineered NAD-preferring XRs and their wild type sources (WT). NADP-preferring engineered enzymes are not shown, nor are enzymes where Michaelis-Menten kinetics are not available.

	k_{cat} / K_M (s^{-1} mM^{-1}) [NADH]	k_{cat} / K_M (s^{-1} mM^{-1}) [NADPH]	Ratio (NADH / NADPH)	Citation
CteXR WT	289	4330	0.07	(Petschacher et al., 2005)[6]
CteXR K274R/N276D	293	234	1.25	
PstXR WT	654	1340	0.49	(Zeng et al., 2009)[21]
PstXR K21A/N276D	1270	N/A	N/A	
PstXR WT	1460	4650	0.31	(Liang et al., 2007)[13]
PstXR K270S/S271G/N272P/R276F	82	6	13.24	
PstXR K270R/N272D	313	107	2.93	
MbmXR WT	**11800**	**2560**	**4.61**	**This study**

N/A = Not available.

References

1. Brinkmann-Chen, S., Cahn, J. K. B., and Arnold, F. H. (2014) Uncovering rare NADH-preferring ketol-acid reductoisomerases. *Metabolic Engineering* **26** 17-22
2. Bruinenberg, P. M., de Bot, P. M., van Dijken, J. P., and Scheffers, W. A. (1983) The role of redox balances in the anaerobic fermentation of xylose by yeasts. *European Journal of Applied Microbiology and Biotechnology* **18**, 287-292
3. Bruinenberg, P. M., de Bot, P. H. M., van Dijken, J. P., and Scheffers, W. A. (1984) NADH-linked aldose reductase: the key to anaerobic alcoholic fermentation of xylose by yeasts. *Applied Microbiology and Biotechnology* **19**, 256-260
4. Metzger, M. H., and Hollenberg, C. P. (1995) Amino acid substitutions in the yeast *Pichia stipitis* xylitol dehydrogenase coenzyme-binding domain affect the coenzyme specificity. *European Journal of Biochemistry* **228**, 50-54
5. Leitgeb, S., Petschacher, B., Wilson, D. K., and Nidetzky, B. (2005) Fine tuning of coenzyme specificity in family 2 aldo-keto reductases revealed by crystal structures of the Lys-274 → Arg mutant of *Candida tenuis* xylose reductase (AKR2B5) bound to NAD^+ and $NADP^+$. *FEBS Letters* **579**, 763-767
6. Petschacher, B., Leitgeb, S., Kavanagh, K. L., Wilson, D. K., and Nidetzky, B. (2005) The coenzyme specificity of *Candida tenuis* xylose reductase (AKR2B5) explored by site-directed mutagenesis and X-ray crystallography. *Biochemical Journal* **385**, 75-83
7. Watanabe, S., Abu Saleh, A., Pack, S. P., Annaluru, N., Kodaki, T., and Makino, K. (2007) Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from *Pichia stipitis*. *Microbiology* **153**, 3044-3054
8. 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**, 10340-10349
9. Watanabe, S., Pack, S. P., Abu Saleh, A., Annaluru, N., Kodaki, T., and Makino, K. (2007) The positive effect of the decreased NADPH-preferring activity of xylose reductase from *Pichia stipitis* on ethanol production using xylose-fermenting recombinant Saccharomyces cerevisiae. *Bioscience, Biotechnology, and Biochemistry* **71**, 1365-1369
10. Ehrensberger, A. H., Elling, R. A., and Wilson, D. K. (2006) Structure-guided engineering of xylitol dehydrogenase cosubstrate specificity. *Structure* **14**, 567-575
11. Jeppsson, M., Bengtsson, O., Franke, K., Lee, H., Hahn-Hagerdal, R., and Gorwa-Grauslund, M. F. (2006) The expression of a *Pichia stipitis* xylose reductase mutant with higher K_M for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* **93**, 665-673
12. Hou, J., Shen, Y., Li, X. P., and Bao, X. M. (2007) Effect of the reversal of coenzyme specificity by expression of mutated *Pichia stipitis* xylitol

dehydrogenase in recombinant *Saccharomyces cerevisiae*. *Letters in Applied Microbiology* **45**, 184-189

13. 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**
14. Matsushika, A., Watanabe, S., Kodaki, T., Makino, K., Inoue, H., Murakami, K., Takimura, O., and Sawayama, S. (2008) Expression of protein engineered $NADP^+$-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* **81**, 243-255
15. Matsushika, A., Watanabe, S., Kodaki, T., Makino, K., and Sawayama, S. (2008) Bioethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing xylose reductase, $NADP^+$-dependent xylitol dehydrogenase, and xylulokinase. *Journal of Bioscience and Bioengineering* **105**, 296-299
16. Petschacher, B., and Nidetzky, B. (2008) Altering the coenzyme preference of xylose reductase to favor utilization of NADH enhances ethanol yield from xylose in a metabolically engineered strain of *Saccharomyces cerevisiae*. *Microbial Cell Factories* **7**
17. Chin, J. W., Khankal, R., Monroe, C. A., Maranas, C. D., and Cirino, P. C. (2009) Analysis of NADPH supply during xylitol production by engineered *Escherichia coli*. *Biotechnology and Bioengineering* **102**, 209-220
18. Bengtsson, O., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. F. (2009) Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Biotechnology for Biofuels* **2**
19. Fernandes, S., Tuohy, M. G., and Murray, P. G. (2009) Xylose reductase from the thermophilic fungus *Talaromyces emersonii*: Cloning and heterologous expression of the native gene (Texr) and a double mutant ($Texr^{K271R+N273D}$) with altered coenzyme specificity. *Journal of Biosciences* **34**, 881-890
20. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* **18**, 2125-2138
21. Zeng, Q.-K., Du, H.-L., Wang, J.-F., Wei, D.-Q., Wang, X.-N., Li, Y.-X., and Lin, Y. (2009) Reversal of coenzyme specificity and improvement of catalytic efficiency of *Pichia stipitis* xylose reductase by rational site-directed mutagenesis. *Biotechnology Letters* **31**, 1025-1029
22. Kim, S.-Y., Oh, D.-K., and Kim, J.-H. (1999) Evaluation of xylitol production from corn cob hemicellulose hydrolysate by *Candida parapsilosis*. *Biotechnology Letters* **21**, 891-895
23. Lee, J. K., Koo, B. S., and Kim, S. Y. (2003) Cloning and characterization of the xyl1 gene, encoding an NADH-preferring xylose reductase from *Candida*

parapsilosis, and its functional expression in *Candida tropicalis*. *Applied and Environmental Microbiology* **69**, 6179-6188

24. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**
25. Kavanagh, K. L., Klimacek, M., Nidetzky, B., and Wilson, D. K. (2003) Structure of xylose reductase bound to NAD^+ and the basis for single and dual co-substrate specificity in family 2 aldo-keto reductases. *Biochemical Journal* **373**, 319-326
26. Di Luccio, E., Elling, Robert A., and Wilson, David K. (2006) Identification of a novel NADH-specific aldo-keto reductase using sequence and structural homologies. *Biochemical Journal* **400**, 105-114
27. Zhu, S., Cao, Y. Z., Jiang, C., Tan, B. Y., Wang, Z., Feng, S., Zhang, L., Su, X. H., Brejova, B., Vinar, T., Xu, M., Wang, M. X., Zhang, S. G., Huang, M. R., Wu, R., and Zhou, Y. (2012) Sequencing the genome of *Marssonina brunnea* reveals fungus-poplar co-evolution. *BMC genomics* **13**, 382
28. Kavanagh, K. L., Klimacek, M., Nidetzky, B., and Wilson, D. K. (2002) The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry* **41**, 8785-8795

Section II

Chapter 4: Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases

Chapter 5: Artificial domain duplication replicates the evolutionary history of ketol-acid reductoisomerases

Chapter 6: Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)-dependent enzymes

Chapter 4

COFACTOR SPECIFICITY MOTIFS AND THE INDUCED FIT MECHANISM IN CLASS I KETOL-ACID REDUCTOISOMERASES

Material from this chapter appears in Cahn J. K. B., Brinkmann-Chen S., Spatzal T., Wiig J. A., Buller A. R., Einsle O., Hu Y., Ribbe M. W., Arnold F. H. (2015). **Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases**, *Biochemical Journal* **468**(3), 475-484, and is reprinted by permission from Portland Press.

Abstract

Although most sequenced members of the industrially important ketol-acid reductoisomerase (KARI) family are class I enzymes, structural studies to date have focused primarily on the class II KARIs, which arose through domain duplication. In the present study, we present five new crystal structures of class I KARIs. These include the first structure of a KARI with a six-residue β2αB- (cofactor specificity determining) loop and an NADP phosphate-binding geometry distinct from that of the seven- and 12-residue loops. We also present the first structures of naturally occurring KARIs that utilize NAD as their preferred cofactor. These results show insertions in the specificity loops that confounded previous attempts to classify them according to loop length. Lastly, we explore the conformational changes that occur in class I KARIs upon binding of cofactor and metal ions. The class I KARI structures indicate that the active sites close upon binding NAD(P), similar to what is observed in the class II KARIs of rice and spinach and different from the opening of the active site observed in the class II KARI of *Escherichia coli*. This conformational change involves a decrease in the bending of the helix that runs between the domains and a rearrangement of the nicotinamide-binding site.

Introduction

The bifunctional enzyme ketol-acid reductoisomerase (KARI, EC 1.1.1.86, also known as acetohydroxyacid isomeroreductase (AHIR or AHAIR)) catalyses the second step in the biosynthesis of the branched-chain amino acids (BCAAs) valine, leucine and isoleucine.[1] The enzyme converts 2-(*S*)-acetolactate (2SAL) into (*R*)-dihydroxyisovalerate

(RDHIV) in an ordered two-step reaction, wherein an Mg^{2+}-dependent alkyl migration is followed by reduction using a hydride from NAD(P) (Figure 4-1a).[1] This unusual reaction has inspired much work to characterize the order of binding, metal ion requirements, and NAD(P) specificity.[1-7]

Because animals lack the BCAA pathway, there is a sizable market for these amino acids produced in microorganisms for human and animal dietary supplements[8-11] and inhibitors of KARIs such as *N*-hydroxy-*N*-isopropyloxamate (IpOHA) are an active area of research as potential herbicides and antibiotics.[12-16] Beyond this, the BCAA pathway has been re-engineered in microbes for production of isobutanol, a promising second-generation biofuel and chemical feedstock.[17]

KARIs are divided into two classes based on their length and oligomerization state (see Figure 5-1). Class I KARIs include all fungal KARIs and are ~340 amino acid residues in length. Class II KARIs include all plant KARIs and are ~490 residues long. Bacterial KARIs can be either class I or class II.[18] KARIs are composed of two types of domains: an N-terminal Rossmann domain and one or two C-terminal knotted domains. Two intertwined knotted domains are required for function, and in the short-chain or class I KARIs, each polypeptide chain has one knotted domain. As a result, dimerization of two monomers forms two complete KARI active sites (Figure 4-2). In the long-chain or class II KARIs, a duplication of the knotted domain has occurred and, as a result, the protein does not require dimerization to complete its active site. As first proposed by Ahn *et al.*,[18] this domain duplication suggests that class II KARIs evolved from a primordial class I KARI. This pair of knotted domains is extensively intertwined to form a figure-of-eight knot. Dimerization of the class I KARIs also forms this knot, albeit with an additional disconnection due to being composed of two chains. The class II KARI from spinach was the first deeply knotted protein identified[19] and remains the most deeply embedded knot observed in a protein crystal structure.[20]

Class I and class II KARIs are also distinguished by the length of the Rossmann fold's β2αB-loop, which binds the adenine ribose of NAD(P) and has been shown to be the primary determinant of whether an enzyme prefers NAD or NADP.[2,21] From a set of 558 class II KARI sequences downloaded from Pfam,[22] all but 20 have a 12-residue β2αB-loop.

In contrast, only five of 2,860 class I KARIs do. Class I KARIs instead favor six- and seven-residue loops (Table 4-1). Prior to the present study, no KARI from the significant proportion of class I homologues having a six-residue β2αB-loop had been crystallized, nor had any of the class I KARIs with 12-residue β2αB-loops. This loop, which we term the 'specificity loop', was the target of our efforts to engineer and predict the cofactor specificity of KARIs in Chapters 1 and 2.[21,23] The lack of crystal structures for KARIs having six-residue loops, however, prevented a thorough understanding of the loop structure–function relationship.

The other major gap in our knowledge of KARIs is how the class I enzymes change conformation upon binding the Mg^{2+} ions, substrate and cofactor. Two crystallographic studies by Guddat and co-workers on substrate-induced rearrangements of class II KARIs reported different mechanisms of induced fit in plant KARIs[24] and *Escherichia coli* KARI.[25] Though both enzymes show an ordered binding in which Mg^{2+} and NADP bind before substrate,[1,26] the structural effects of binding are dramatically different. In the plant enzymes, the hinge between the Rossmann domain and the knotted domain moves freely in the apo state, an observation supported by H-D exchange mass spectrometry.[5] Binding of Mg^{2+} and NADP draws the domains together, closing the active site and rearranges the α3+1-helix to pre-organize the enzyme for substrate binding. In contrast, in the bacterial enzyme the active site is closed in the apo crystal structure, but binding of Mg^{2+} and NADP opens the interface between the domains and allows the substrate to bind. Nothing, however, is known about the conformational changes in class I KARIs, which may be complicated by the dimeric nature of the protein.

In the present study, we present five new class I KARI structures from four previously un-crystallized enzymes. Three of these structures give insight into the unexpected diversity of the specificity loop. With the other two structures, we demonstrate that the mechanism of induced fit in the bacterial class I KARIs we studied involves closure of the interdomain hinge, coupled with a rearrangement in the nicotinamide amide-binding portion of the active site. This behaviour is opposite to that observed in the bacterial class II KARI previously described by Wong *et al*.[25] and more closely resembles the behavior of the plant KARIs.[24] Taken together, these structures enhance our understanding of the structural and functional diversity of this enzyme class.

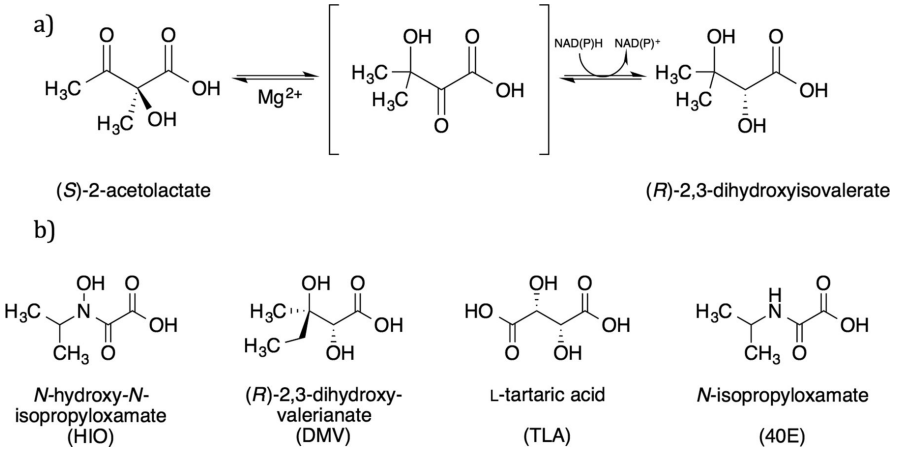


Figure 4-1. The two step reaction catalyzed by KARI (a). The four substrate analogues (b) that have been crystallized in the active sites of KARIs, including in the present study. PDB ligand IDs are shown in parentheses.

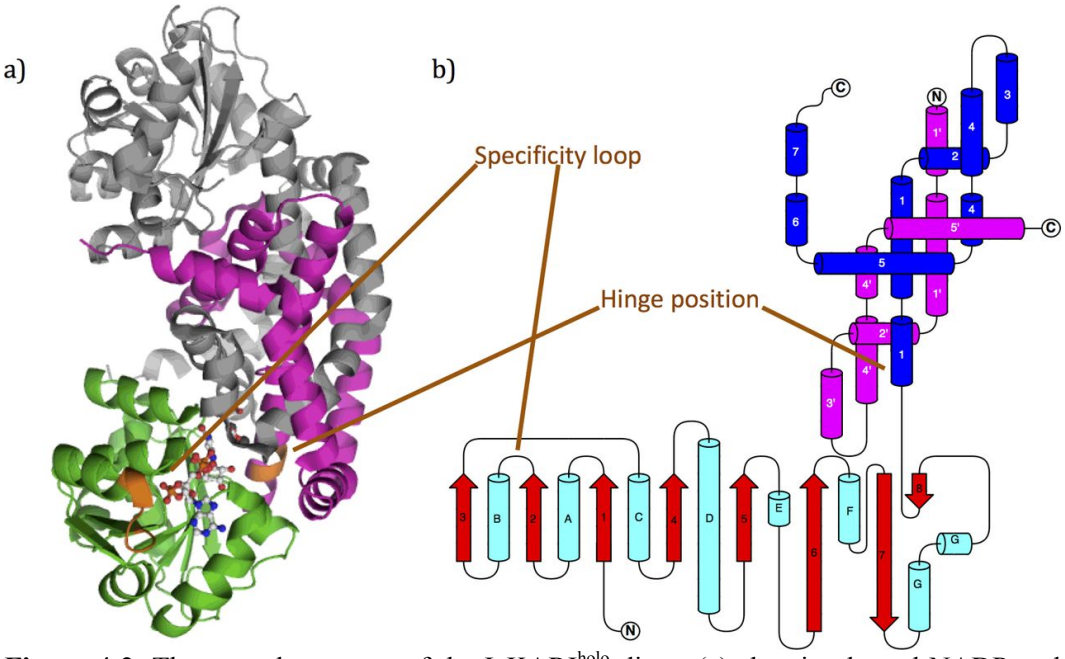


Figure 4-2. The crystal structure of the IaKARI^{holo} dimer (a) showing bound NADP and inhibitor (white) and Mg^{2+} ions (cyan). One monomer is shown in grey; the other is shown in green (Rossmann domain) and magenta (knotted domain). The topology of class I KARIs (b) is illustrated schematically with cyan helices and red β-strands in the Rossmann domain; the knotted domain has blue helices. Also shown is the knotted domain of the dimeric partner (magenta), which has been truncated for simplicity. The figure was constructed using TopDraw.[50]

Table 4-1. Specificity loop length and KARI classes of 3,422 unique KARIs retrieved from Pfam. Four additional class I KARIs had other specificity loops. Numbers in parentheses indicate the number of structures available before this study.

Loop length	Sources	6	7	12
Class I	Bacteria, archea, fungi	595 (0)	2260 (3)	5 (0)
Class II	Bacteria, plants	0 (0)	20 (4)	538 (2)

Materials and Methods

Cloning, expression, and purification of KARIs

Cloning, expression and purification of AaKARI (*Alicyclobacillus acidocaldarius*), UaKARI (uncultured archaeon) and IaKARI (*Ignisphaera aggregans*) were described previously.[2,21] AvKARI was isolated from *Azotobacter vinelandii* strain AvOP (ATCC BAA-1303). Cell growth and AvKARI purification by anion exchange and gel filtration chromatography were performed as published earlier.[27]

Crystallization and data collection of AaKARI, UaKARI, and IaKARI

IpOHA was prepared as described.[21] High-throughput screening of crystallization conditions was conducted at the Beckman Molecular Observatory at the California Institute of Technology. For AaKARI in the presence of NADPH, the best condition was 1M sodium potassium tartrate, 200 mM sodium chloride, 100 mM imidazole, pH 8. For IaKARI in the presence of NADPH and with IpOHA as inhibitor, the best condition was a 0.1 M Bis-Tris, pH 6, solution with 22% polyethylene glycol (PEG) monomethylether 5000. The apo crystal of IaKARI was obtained from the same conditions, but with 200 μM NADH instead of NADPH. For UaKARI in the presence of NADH and with IpOHA as inhibitor, the best condition was a 0.1 M Bis-Tris, pH 5, solution with 20% w/v PEG 1500 as precipitant. For cryoprotection, the crystals were soaked in a mother liquor containing 25% glycerol prior to flash-freezing in liquid nitrogen. Diffraction data were collected using a Dectris Pilatus 6M detector on beamline 12-2 at the Stanford Synchrotron Radiation Laboratory SSRL at 100 K. Diffraction datasets were integrated with XDS[28] and scaled using SCALA.[29]

Crystallization and data collection of AvKARI

AvKARI was crystallized under strictly anaerobic conditions using 0.25 M NaCl, 28% PEG 3350 and 0.1 M bis-tris, pH 5.5. Partial dehydration of the observed crystals was achieved by equilibrating the crystals against 40% PEG 3350 for 12 h, followed by subsequent flash-freezing in liquid nitrogen. Diffraction experiments were carried out at 100 K at the Swiss-Light-Source PXI beamline X06SA (Paul-Scherrer-Institute). Data were integrated using iMOSFLM[30] and scaled in SCALA.[29]

Structure solution

For AaKARI^{holo} the structure of *Slackia exigua* KARI (SeKARI^{holo}, PDB code 4KQW) was used for molecular replacement with MOLREP;[31] IaKARI^{holo} and UaKARI^{holo} used the structure of the DDV mutant of that enzyme (SeKARI^{DDV}, PDB code 4KQX) and molecular replacement was performed using Phaser.[32] For IaKARI^{apo}, the structure of IaKARI^{holo} was used for molecular replacement with Phaser; the structure was split into two PDB files containing the Rossmann and knotted domains, which were fit as separate ensembles. Molecular replacement for AvKARI^{met} used PaKARI^{apo} (PDB code 1NP3[18]) for molecular replacement with MOLREP, with determination and refinement of iron-sites performed using SHELX.[33] After molecular replacement, several iterations of automated refinement with Refmac5[34] (CCP4 suite[35]) and manual refinement in Coot[36] were performed.

Structure analysis

For the whole-enzyme alignments (i.e., Figure 4-4), the alignment was performed using PyMOL;[37] for the domain-specific alignments, LSQ superposition was performed using Coot.[36] For analysis of helix bending, the HELANAL-Plus web server[38] was used, with the helices of each structure assigned by STRIDE. From the first half of the α1-helix (i.e., before the break in helicity) the l, m, n direction cosines of the first and last i/i-3 pair were extracted and the angle between them was computed.

IaKARI thermostability determination

T_{50} is defined as the temperature at which 50% of the initial activity is retained after 1 h incubation. Thirty-microlitre aliquots of purified enzyme were transferred to PCR tubes. Each tube was assigned a specific incubation temperature on the block of an Eppendorf Mastercycler. The measurements were conducted in duplicates. The tubes were incubated in their slots for 1 h and then quenched on ice. Residual activity was determined with the activity assay described previously.[2]

Results and Discussion

We present five crystal structures representing four class I KARIs, with resolutions of 1.15–2.50 Å (Table 4-2). We previously described three of these KARIs in our work on cofactor specificity engineering. In Chapter 1, the cofactor specificity of AaKARI from *A. acidocaldarius* was inverted by the introduction of three mutations in its specificity loop.[21] In Chapter 2, UaKARI from an uncultured organism and IaKARI from *I. aggregans* were correctly predicted to utilize NAD, based on their loop sequences.[2] The fourth KARI, from *A. vinelandii*, crystallized serendipitously as an impurity of another protein purified from *A. vinelandii* lysate. Table 4-3 places these enzymes and the structures reported in the present study in the context of the previously available KARI structural data. For some proteins, including IaKARI in the present study, multiple structures are available with different bound cofactors, metals, and substrate analogues. Refinement statistics of the structures reported in the present study are provided in Table 4-2. All of these KARIs have the same fold as previously solved class I KARIs (Figure 4-2). Their active sites, particularly the metal-binding sites, are also in agreement with those of previous structures (see, e.g., Biou *et al*.[39]).

Various nomenclatures have been used to discuss the structures of KARIs. In this thesis, we refer to the N-terminal domain as the Rossmann domain and the C-terminal domain as the knotted domain. The topology of the class I enzyme is shown and labelled in Figure 4-2b. Secondary structure elements are numbered starting from the first β-strand that makes up the canonical Rossmann fold; secondary structure prior to this strand is not conserved. In the Rossmann domain, β-strands are numbered 1–8 and α-helices are assigned the letters from A–G. In the knotted domain, which is entirely α-helical, helices

are assigned numbers from 1 to 7. Loops between secondary structural elements are named based on the flanking secondary structural elements, as in β2αB-loop or α3α4-loop. We refer to the β2αB-loop of the Rossmann domain as the specificity loop. Figure 4-3, which has an alignment of the KARI sequences discussed herein, is also annotated with the secondary structure labels used. In class I KARIs, the secondary structural elements of the dimeric partner are denoted with a prime as in α3′-helix. In class II KARIs, the duplicated elements are denoted with '+1', as in α3+1-helix. Lastly, we refer to the proteins themselves using a two-letter code plus 'KARI' (i.e., AaKARI) and the structures with the addition of a superscript (AaKARI^{holo}) indicating their complexation state.

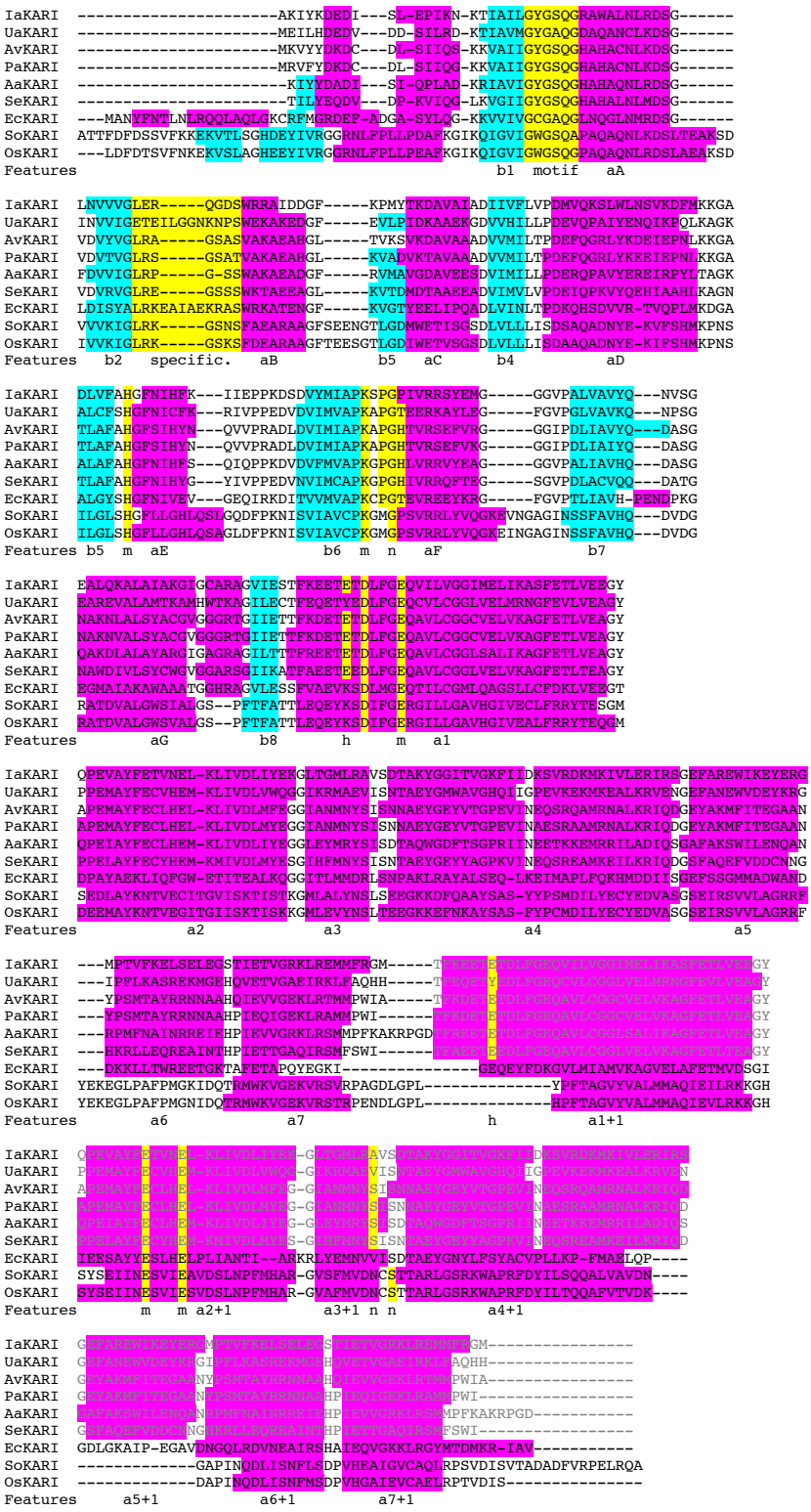
Table 4-2. Data collection and refinement statistics for structures presented in this chapter. Values in parentheses represent statistics from the highest resolution shell.

	AaKARI^{holo} 4TSK	**UaKARI^{holo} 4XDY**	**IaKARI^{holo} 4XDZ**	**IaKARI^{apo} 4XEH**	**AvKARI^{met} 4XIY**
			Data Collection		
Space Group	*P* 2 3	*P* 2_1 2_1 2	*P* 2_1	*P* 2_1 2_1 2	*P* 2_1 3
Cell Dimensions					
***a*,*b*,*c* (Å)**	124.10, 124.10, 124.10	141.75, 148.60, 49.70	54.58, 90.75, 69.53	67.48, 112.35, 46.07	185.05, 185.05, 185.05
***α*,*β*,*γ* (°)**	90, 90, 90	90, 90, 90	90, 100.3, 90	90, 90, 90	90, 90, 90
Resolution (Å)	124.10-2.50 (2.57-2.50)	102.57-1.54 (1.58-1.54)	68.42-1.15 (1.21-1.15)	57.85-1.39 (1.43-1.39)	82.76-2.50 (2.64-2.50)
$R_{p.i.m}$ (%)	5.3 (61.8)	4.4 (49.8)	5.4 (64.3)	2.2 (107.9)	5.7 (31.3)
Mn(I)/sd	8.9 (0.8)	11.1 (1.5)	6.9 (0.9)	13.6 (0.6)	10.9 (2.6)
Completeness (%)	97.7 (89.1)	97.3 (71.8)	95.0 (82.4)	96.8 (76.0)	99.5 (99.9)
Redundancy	3.4 (2.9)	4.5 (4.5)	3.0 (2.5)	3.6 (3.9)	3.5 (3.6)
			Refinement		
No. reflections	22,066	145,320	224,341	65,375	72,509
***R_{work}*/*R_{free}* (%)**	15.6/18.2 (21.2/27.2)	17.9/20.9 (39.1/38.4)	15.8/19.0 (33.2/34.1)	18.0/23.5 (38.2/38.0)	19.4/24.4 (28.7/32.1)
No. atoms					
Protein	2,585	5,296	5,139	2,561	10,017
Ligand/ion	61	126	178	0	48
Water	46	359	595	101	40
RMSD					
Bond lengths (Å)	0.018	0.023	0.024	0.019	0.015
Bond angles (°)	1.998	2.038	2.461	1.930	1.719
Ramachandran map analysis					
Favored	313	653	589	309	1,194
Allowed	17	20	29	13	99
Outliers	1	0	0	0	23

Table 4-3. Published KARI structures (above the line) and the new ones presented below. Each is given a name for discussion in this paper. For each, the metals and bound ligands are given. Cofactor IDs are as follows: NDP = NADPH, NAP = $NADP^+$, NAD = NAD^+, APX = 2'-monophosphate adenosine diphosphate ribose; see Figure 4-1b for structures of substrate analogues. KARI class and β2αB-loop (specificity loop) length are also given for each enzyme.

PDB ID	Name	Host	Resn. (Å)	Metal	Cofactor	Analogue	Class	β2αB-loop
1YVE (39)	$SoKARI^{holo}$	*S. oleracea*	1.65	$2xMg^{2+}$	NDP	HIO	II	7
1QMG (40)	$SoKARI^{prod}$	*S. oleracea*	1.60	$2xMn^{2+}$	APX	DMV	II	7
1NP3 (18)	$PaKARI^{apo}$	*P. aeruginosa*	2.00	--	--	--	I	7
1YRL (49)	$EcKARI^{apo}$	*E. coli*	2.60	--	--	--	II	12
3FR7 (24)	$OsKARI^{met}$	*O. sativa*	1.55	$2xMg^{2+}$	--	--	II	7
3FR8 (24)	$OsKARI^{holo}$	*O. sativa*	2.80	$1xMg^{2+}$	NDP	--	II	7
3ULK (25)	$EcKARI^{holo}$	*E. coli*	2.30	$2xMg^{2+}$	NDP	--	II	12
4KQW (21)	$SeKARI^{holo}$	*S. exigua*	1.38	$2xMg^{2+}$	NAP	TLA	I	7
4KQX (21)	$SeKARI^{DDV}$	*S. exigua*	1.80	$2xMg^{2+}$	NAD	HIO	I	7
4TSK	$AaKARI^{holo}$	*A. acidocaldarius*	2.50	$2xMg^{2+}$	NAP	TLA	I	6
4XDY	$UaKARI^{holo}$	Uncultured archaeon	1.54	$2xMg^{2+}$	NAD	HIO	I	12
4XDZ	$IaKARI^{holo}$	*I. aggregans*	1.15	$2xMg^{2+}$	NDP	40E	I	7
4XEH	$IaKARI^{apo}$	*I. aggregans*	1.39	--	--	--	I	7
4XIY	$AvKARI^{met}$	*A. vinelandii*	2.50	$1xMg^{2+}$, $1xFe^{2+}$	--	--	I	7

Figure 4-3. (next page) Sequence alignment of crystallized KARIs. Helices are colored in magenta and β-strands in cyan. Yellow indicates special features, labeled with the following abbreviations: motif = diphosphate binding motif; specific. = specificity loop; m = metal-binding residue; n = nicotinamide amide-binding residue; h = hinge point (when unambiguously definable). Metal-binding and nicotinamide amide-binding residues are shown only for one active site in the class I proteins. The knotted domain of the class I KARIs has been duplicated for clarity; duplicate residues are shown with grey text.


IaKARI
UaKARI
AvKARI
PaKARI
AaKARI
SeKARI
EcKARI
SoKARI
OsKARI
Features
b1 motif aA
b2 specific. aB b5 aC b4 aD
b5 m aE b6 m n aF b7
aG b8 h m a1
a2 a3 a4 a5
a6 a7 h a1+1
m m a2+1 a3+1 n n a4+1
a5+1 a6+1 a7+1

Structural diversity of the specificity loop: Structure of AaKARI's six-residue specificity loop

Until the work in Chapter 2, KARI family enzymes were believed to be exclusively NADP-dependent.[2] Because NAD utilization is advantageous in the industrial production of amino acids and biofuels,[21,23] engineering KARI cofactor specificity has been of considerable interest. Previous studies by our group aimed at reversing KARI cofactor specificity focused extensively on the specificity loop of the Rossmann domain due to its key role in binding the adenine ribose of the cofactor.[2,21,23] Structures of enzymes with a 12-residue specificity loop[25] and multiple structures with seven-residue specificity loops[21,39,40] with bound NADP were available. However, no structure of a KARI having a six-residue specificity loop was available, with or without NADP. In Chapter 1, we found that whereas similar mutations reversed the specificity of KARIs with seven- and 12-residue specificity loops, a modified approach was required for the six-residue specificity loop enzymes.[21] Ambiguities in the alignments of the different length of the loop sequences limited our ability to propose a structural model to explain the different sets of mutations (see the alignment in Figure 4-4). For that reason, we selected for structural analysis the KARI from *A. acidocaldarius* (AaKARI), a mildly thermophilic enzyme that was engineered for NAD preference in Chapter 1.[21] AaKARI crystallized readily, and we obtained a crystal that diffracted to 2.5 Å.

The overall structure of AaKARI^{holo} is very similar to that of the other class I KARIs that have been published, with RMSDs of 0.48 and 1.56 Å against SeKARI^{holo} and PaKARI^{apo},[18,21] respectively. Figure 4-5 shows the specificity loop of AaKARI^{holo} (part c) as well as other KARIs, including ones that will be discussed later in this chapter. Although the side chains of three residues of the seven- and 12-member specificity loops form hydrogen bonds to the NADP phosphate (e.g., R58, S61, and S63 in SeKARI^{holo}), only the side-chains of two residues do so for the six-member specificity loop of AaKARI^{holo}, R48 and S52. However, the rearrangement of the loop brings the αB-helix closer to the adenine ring, allowing the terminal serine residue's side chain to hydrogen bond not only with the phosphate but also with the O2′ and O3′ of the ribose, and creating a hydrogen bond between the backbone N-H group of S52 and the phosphate oxygen. This interaction

between the phosphate and the dipole of the αB-helix has not been observed previously in a KARI.

In Chapter 1, we found that reversing the cofactor specificity of KARIs having seven- and 12-residue specificity loops from NADP to NAD required installation of aspartate residues at the ultimate and antepenultimate loop positions.[21] Existing structures showed that these two residues provide the primary hydrogen bonding interactions with the NADP phosphate.[21,25] The six-residue specificity loops, in contrast, only required a single aspartate along with mutation of the arginine to proline at the second position of the loop to switch cofactor preference to NAD. Having obtained the first structure of a KARI with a six-residue specificity loop, we can now explain why. Only the ultimate residue of AaKARI^{holo}'s specificity loop forms a hydrogen bond with the phosphate. There is no residue corresponding to the interaction provided by the antepenultimate position. Presumably to compensate for this missing interaction, the contact between the arginine at the second position of the loop and the phosphate is more extensive, such that mutation of this residue is also critical in order to abolish phosphate binding and switch cofactor specificity.

Name	β2	specificity loop	αB	length	class	kingdom	cofactor
Ec_KARI	ISYA	L---R--KEAIAEKRAS	WRKA	12	II	Bacteria	NADP
So_KARI	VKIG	L---R-------KGSNS	FAEA	7	II	Plantae	NADP
Se_KARI	VRVG	L---R-------EGSSS	WKTA	7	I	Bacteria	NADP
Aa_KARI	VVIG	L---R-------PG-SS	WAKA	6	I	Bacteria	NADP
Ua_KARI	VVIG	ETEILGG-----NKNPS	WEKA	12	I	Bacteria	NAD
Ia_KARI	VVVG	LE--R-------QG-DS	WRRA	7	I	Bacteria	NAD/P

Figure 4-4. A structure-guided alignment of the specificity loops (including four residues on either side) of some of the KARIs discussed in this chapter, annotated with the length of the loop, the class of the KARI, the kingdom of origin, and the cofactor preference. Residues contacting the phosphate are highlighted in orange.

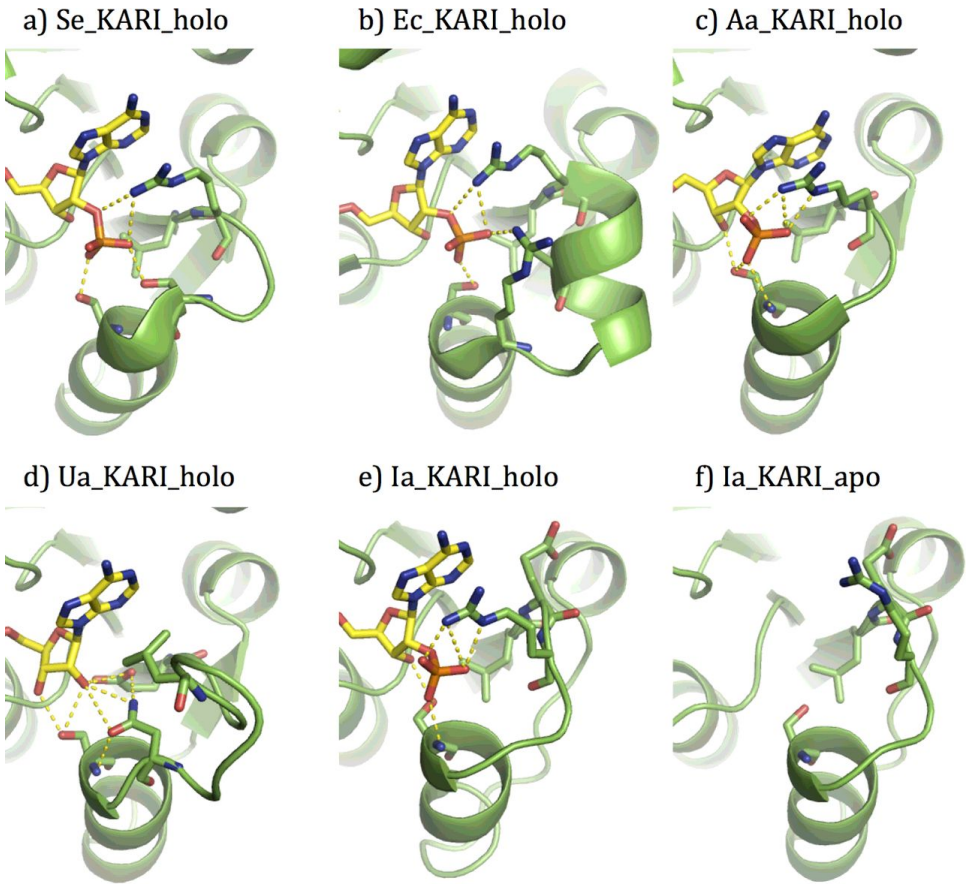


Figure 4-5. The β2αB-loops of SeKARI^{holo} ((a); PDB 4KQW[21]), EcKARI^{holo} ((b); PDB 3ULK[25]), AaKARI^{holo} (c), UaKARI^{holo} (d), IaKARI^{holo} (e) and ^{apo} (f). Polar interactions (i.e., hydrogen bonds) of 3.5 Å or less to the 2' or 3' moieties of the ribose are shown with dashed lines, as are selected interactions between amino acid residues.

Structural diversity of the specificity loop: Structure of the naturally NAD-preferring UaKARI

In the past, it had been assumed that all KARIs were specific for NADP.[23] However, in Chapter 2 we showed that some native KARIs display strong preference for NAD or at least exhibit equal activity with NADP and NAD.[2] The KARI with the greatest preference for NAD (in terms of catalytic efficiency, k_{cat}/K_M) came from an uncultured archaeon (UaKARI)[2] and was also one of the only five known class I KARIs to possess a 12-residue specificity loop (Table 4-1). However, we could not predict the arrangement of the residues around the NAD based on sequence alone because the loop sequence was so different from any previously crystallized KARI (Figure 4-4). It also did not allow us to effectively compare it to the recipe we proposed previously for engineering naturally NADP-preferring KARIs to use NAD.[21] For the present study, we were able to obtain a crystal of UaKARI bound to NAD, which diffracted to 1.54 Å.

UaKARI^{holo} possesses the same overall KARI fold as other class I KARIs, with a 0.62 Å RMSD to SeKARI^{holo}. However, the specificity loop of UaKARI^{holo} differs significantly from any other crystallized KARI (Figure 4-5d). Glu46, the first residue of the β2αB-loop, lies along the N3 edge of the adenine moiety and forms a bidentate interaction with the O2′ hydroxy of the ribose. This residue is structurally homologous to the conserved leucine that makes a similar interaction with the adenine in other KARIs (see e.g., Figures 1-2 and 2-1). The position of this glutamate is stabilized by a hydrogen bond interaction with N55.

In most KARIs, an arginine in the position immediately following the conserved leucine forms a cation–pi interaction with the adenine and a salt bridge with the phosphate of the cofactor.[18] In UaKARI^{holo}, the arginine is replaced by L49, which removes the positive charge. Additionally, the specificity loop is extended by five residues that accommodate the structural differences between arginine and leucine. In *E. coli* KARI, which also has a 12-residue loop, the five additional residues form a short helix inserted after the arginine. In UaKARI^{holo}, by contrast, there are three residues before Leu49 and two after. This leads to a reorientation of the specificity loop (Figure 4-5d). As with the canonical NADP-binding seven- and 12-residue specificity loops, UaKARI^{holo}'s specificity

loop has two additional residues that provide contacts to the (phospho-) ribose. S57 at the C-terminal end of the specificity loop is highly conserved and, as in AaKARI^{holo}, forms hydrogen bonds with both O2′ and O3′ of the ribose. Though there is no phosphate to hydrogen bond to its backbone amine, this backbone nitrogen hydrogen-bonds to the side chain of another residue, the antepenultimate N55. The increased size of N55 over the consensus serine at this position allows it to make contact with O2′ rather than the absent phosphate and the nitrogen forms a hydrogen bond with E46, as noted above.

The lack of a short helix within the loop suggests that the ancestral UaKARI did not have a canonical 12-residue specificity loop. UaKARI may have evolved from a KARI possessing a shorter loop. This is supported by the fact that the KARIs with the sequences most similar to UaKARI (other than a 98% identical one from another uncultured archaeon) both had six-residue specificity loops, as discussed in Chapter 2.[2]

Structural diversity of the specificity loop: The unusual β2αB-loop of the bispecific IaKARI

The *I. aggregans* KARI identified in Chapter 2 (IaKARI) displayed sub-micromolar K_M values for both NAD and NADP[2] and measurements carried out for the present study confirmed that it is highly thermostable, with half-maximal residual activity (T_{50}) after an hour at 95°C. We were unable to obtain a structure of IaKARI bound to NAD, but did succeed in cocrystallizing it with NADP and the transition state analogue *N*-isopropyloxamate (Figure 4-1b). The substrate analogue's electron density indicated that it had been reduced at the *N* position (cause unknown) and a water molecule had taken the place of the missing -OH group (Figure 4-6).

Because IaKARI^{holo} has a seven-residue specificity loop, we had expected the second loop residue, a glutamate, to make contact with the face of the adenine ring and the antepenultimate and ultimate residues to bind the phosphate. Surprisingly, the binding mode for NADP closely resembles that of the six-residue loop from AaKARI^{holo}, with a single serine residue at the ultimate position of the loop hydrogen bonding with both the O3′ hydroxy and the phosphate (Figure 4-5e). Furthermore, the glutamate at the second position of the loop faces away from the cofactor entirely, whereas the arginine at the third position forms the typical packing interaction against the adenine ring. Without an NAD co-crystal structure, it is difficult to speculate what functional role this glutamate serves or

how it contributes to the enzyme's cofactor bispecificity. However, it is possible that a rearrangement of the loop might allow it to occupy the position of either of the conserved residues flanking it. Such rearrangements of adenosine ribose-binding loops have been observed in bispecific glucose-6-phosphate dehydrogenase,[41] xylose reductase,[42] glyceraldehyde-3-phosphate dehydrogenase,[43] and coenzyme A-disulfide reductase.[44]

It is interesting that both the naturally NAD-utilizing KARIs, UaKARI and IaKARI, had amino acid insertions relative to other structurally similar enzymes. Insertions and deletions in genes represent a major source of sequence and function variation in natural evolution,[45,46] but are rarely used in protein engineering and directed evolution. In the guide proposed in Chapter 1, classifying the binding mode based on the length of the specificity loop was the first step to identifying a protein engineering strategy for engineering KARI cofactor specificity.[21] These results, however, demonstrate that loop length may be an insufficient indicator of binding geometry.

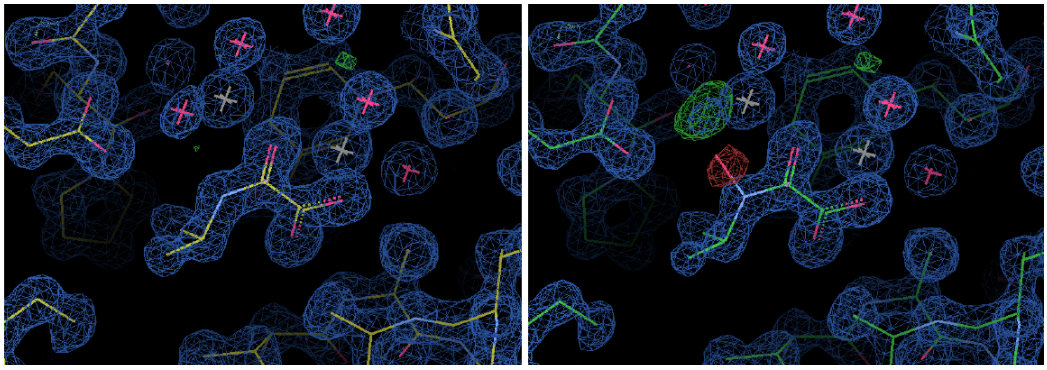


Figure 4-6. Electron density around the substrate of IaKARI^{holo} indicates the dehydroxylation of IpoHA. Left: modeled with *N*-isopropyloxamate (ligand 40E). Right: modeled with *N*-hydroxy-*N*-isopropyloxamate (IpoHA; ligand HIO). The F_o-F_c map (red and green) clearly shows the absence of the *N*-hydroxy moiety. ($2F_o$-F_c map contoured to 1.5σ, F_o-F_c map contoured to 4σ.)

Conformational changes in class I KARIs: Induced fit in IaKARI

The IaKARI structures with and without bound metals or ligands allowed us to observe the changes in the protein structure upon the binding of NADP, metals, and *N*-isopropyloxamate. As in the plant and the *E. coli* class II KARIs, the two domains of IaKARI undergo a rigid body movement, where only the angle between them changes significantly. With metals, cofactor, and inhibitor bound, the Rossmann domains have moved closer together, sealing off the active sites and generating a 'closed' state. Either domain of the ligand-bound protein can be superimposed on its respective apo-protein domain, but the full structures do not superimpose. Figure 4-7 shows a plot of the C^{α}-C^{α} distances along the length of the chain for IaKARI^{holo} and IaKARI^{apo}, aligned by Rossmann domain (blue) or knotted domain (red). With either alignment a distinct jump in the C^{α}-C^{α} distances indicates a 'hinge point' at E189. E189 is on the α1-helix of the knotted domain rather than the β8α1-loop between the domains, which stays rigid with respect to the N-terminal domain. In the ligand-bound structure, the α1-helix is less bent, with an angle of 4.7° compared with 8.7° in the apo structure, as measured using HELANAL-Plus.[38] Figure 4-8 shows this movement structurally (a and b) and quantitatively (c) for IaKARI as well as other structure pairs to be discussed later. However, there is no significant rearrangement in the helix to explain this change in curvature. The changes are subtle and cumulative at the various backbone positions. Only a handful of residues in the protein change rotamer state between the two crystals, most notably H108, K131, D191, and E195, all of which are involved in the binding of the two Mg^{2+} ions. Similarly, R49 and R56, which are involved in binding the NADP cofactor, reorient slightly.

Three regions have notably non-rigid-body behaviour: the N-terminus of the αA-helix (including the latter half of the Rossmann-identifying GXGXXG motif[47]), the β6αF-loop, and the N-terminus of the αF-helix and the α3α4-loop. Indeed, in the ligand-bound structure, the αA-helix and β6αF-loop interact with the α3-helix not of their own chain but of their dimeric partner. This interaction makes up the opposite side of the active site from the D191/E95/α1 face closer to the hinge. A rearrangement of the α3+1-helix was previously identified in the plant KARIs and was thought to be involved in binding IpOHA and 'capping' the active site.[24] In the IaKARI^{apo} structure, a combination of this rearrangement and an increase in the α1-helix curvature means that the interaction between

these regions is absent. The distance between the α-carbons of G134 (on the αF-helix) and A250' (on the dimeric partner's α3-helix) goes from 11.8 Å in the apo structure to 6.6 Å in the ligand-bound structure.

Critically, this closure is centred around the nicotinamide moiety of NADP, in particular the C7N amide. As shown schematically in Figure 4-9a, the side chain of Q28 hydrogen bonds to the backbone amine of G134 in the absence of the cofactor. With cofactor bound, the terminal amide is inserted into this gap, with O7N hydrogen bonding to the backbone amine of G134 and Q28 flipped out of the way. The N7N nitrogen of the amide recruits the α3′-helix, specifically the backbone carbonyl of A250', as well as the side chain of S27. S27 hydrogen bonds to the cofactor, G134 and A250', holding the complex together and ‘capping’ the active site. Figure 4-10 shows this motion in the structure.

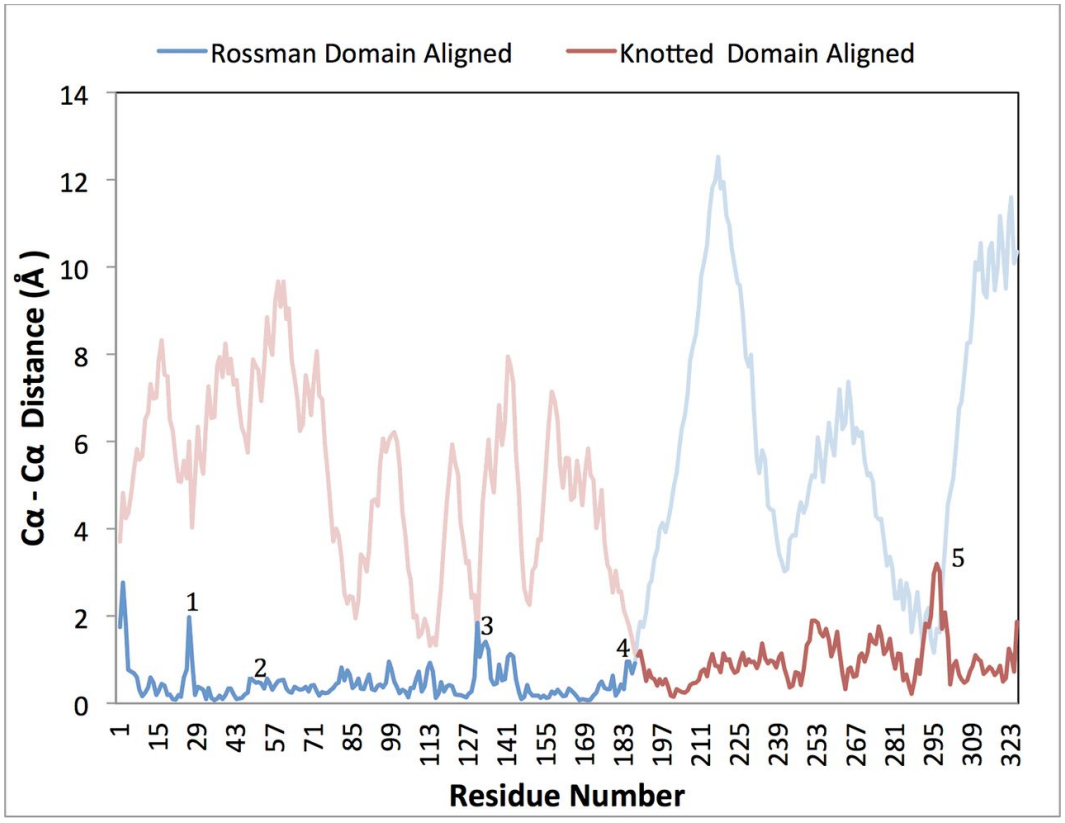


Figure 4-7. C^{α}-C^{α} distances for residues of IaKARI^{holo} aligned against IaKARI^{apo}. Because of the rigid body movement of the interdomain hinge, the Rossmann (blue) and knotted (red) domains have been aligned separately. The non-aligned domains are shown in a lighter color. Points of interest are numbered as follows: (1) αA-loop, (2) specificity loop, (3) β6αF-loop, (4) hinge, (5) α3-helix.

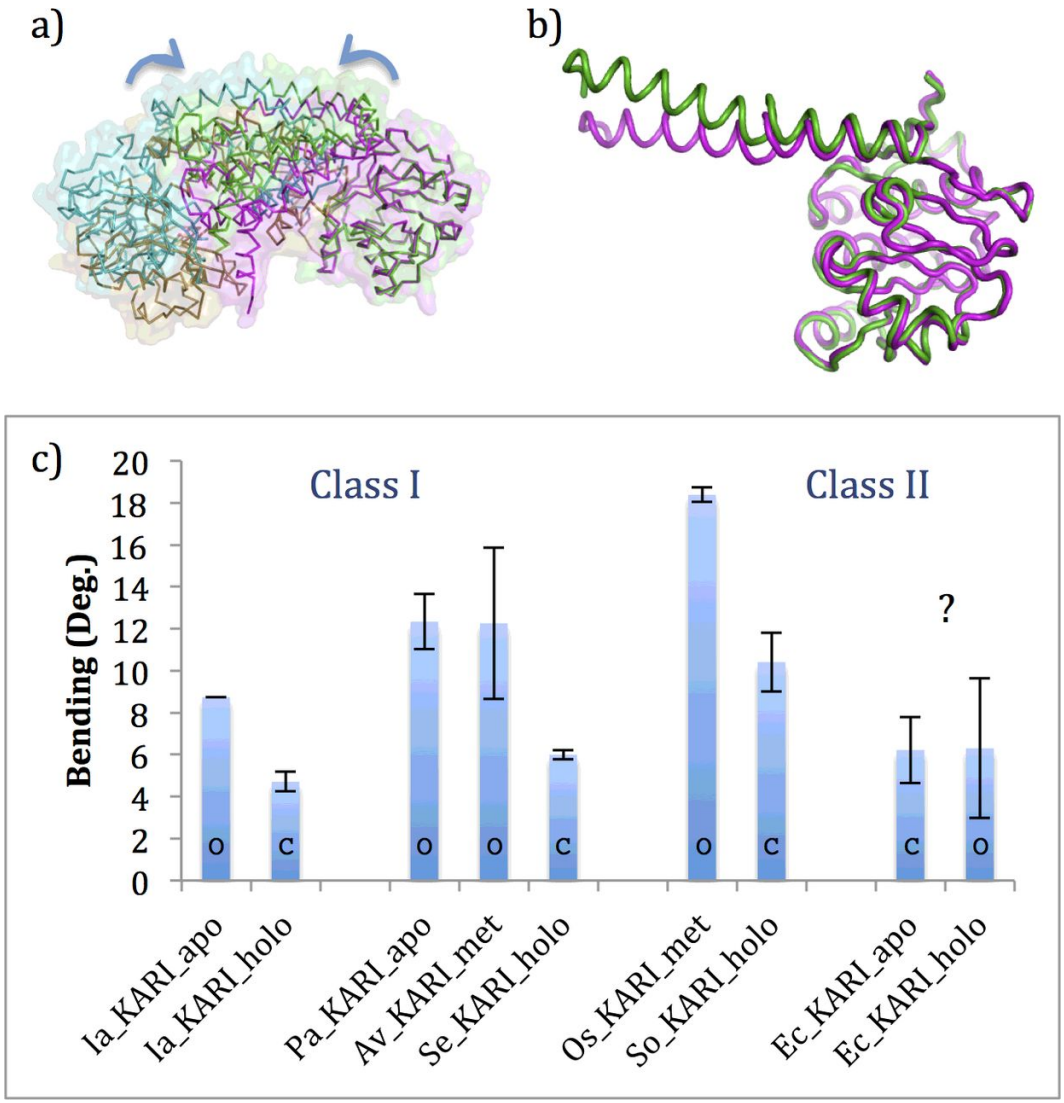


Figure 4-8. Bending of α1-helix is responsible for induced fit. An alignment (a) between the apo and holo dimers of IaKARI. The structures are aligned based on their rightmost Rossmann domains. The apo structure is shown in green and cyan and the holo structure is shown in magenta and tan. The Rossmann domains and α1-helices (b) of the apo (green) and holo (magenta) crystals of IaKARI, showing the change in the bend of the first half of the α1-helix. Quantitative analysis of the bending of the first half of the α1-helix (c). Error bars show the standard deviation of the values for multiple chains in the asymmetric unit of the crystal. Bars are denoted with an 'O' or a 'C' to designate open and closed states of the active site, as assigned visually. EcKARI is a clear outlier, both because it goes from a closed to open state upon binding cofactor and because there is no corresponding change in the α1-helix bending.

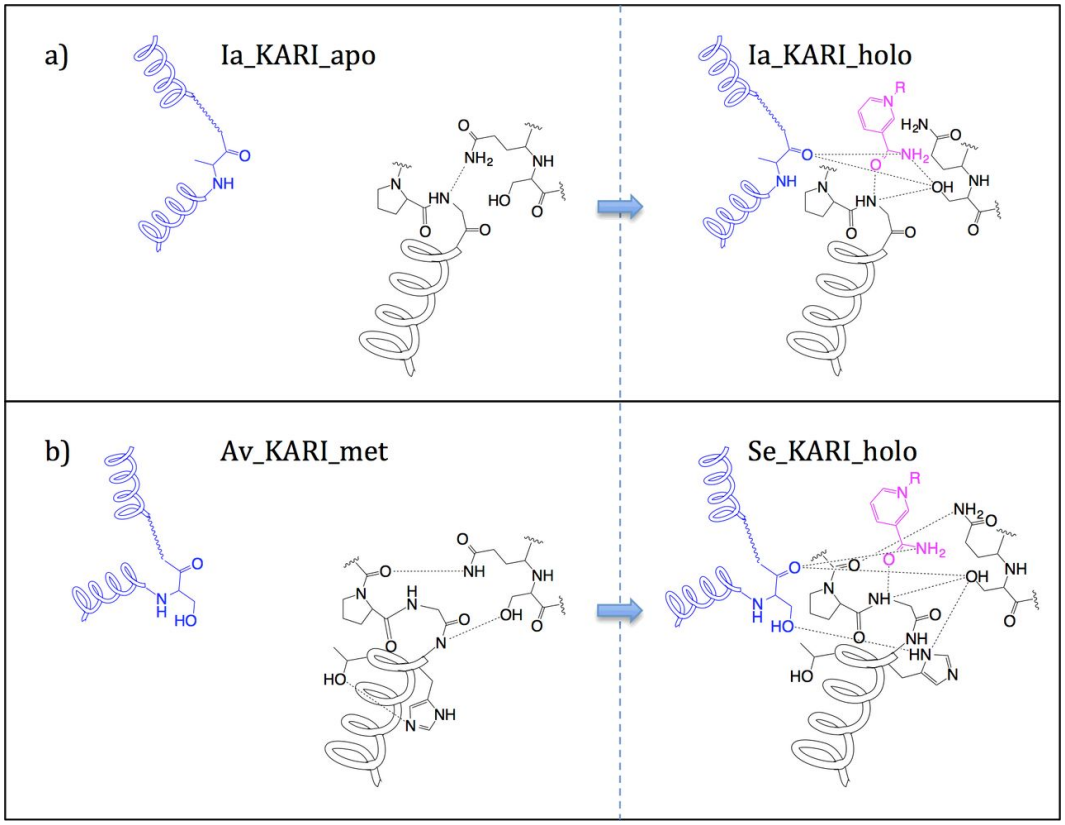


Figure 4-9. Hydrogen bond rearrangements around the nicotinamide amide. Hydrogen bonds involved in the recruitment of the α3′α4′-loop upon cofactor binding in (a) IaKARI^{apo} and IaKARI^{holo}, and (b) AvKARI^{met} and SeKARI^{holo}. R=adenosine diphosphate ribose phosphate.

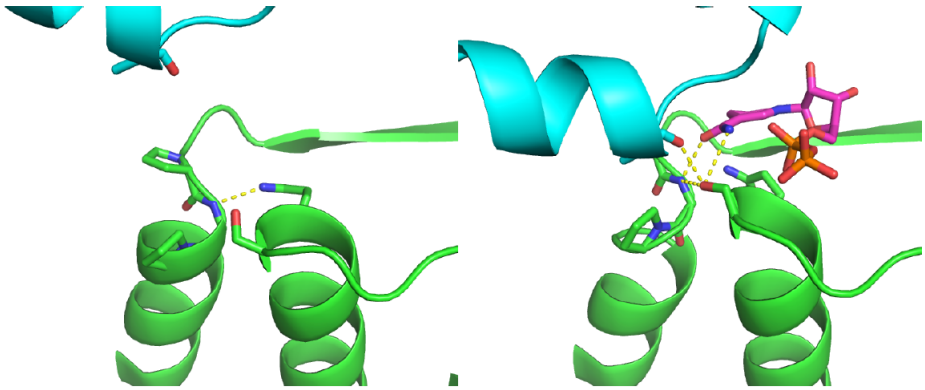


Figure 4-10. Structural rearrangements around the NADP nicotinamide in IaKARI. Binding of the nicotinamide of NADP (magenta) to the αA-helix and the β6αF-loop recruits the α3-helix of the other chain of the dimer (cyan).

Conformational changes in class I KARIs: Induced fit in other class I KARIs

From this pair of structures, it was unclear to what extent the gross conformational change was driven by the smaller-scale rearrangements related to metal binding, cofactor binding, or some combination thereof. However, we were also able to obtain a structure of another KARI, from *A. vinelandii*, with metal ions (Mg^{2+} and Fe^{2+}) bound in the active site, but without a cofactor or a substrate analogue. Because AvKARI^{met} is from a mesophile and possesses a seven-residue NADP-dependent β2αB-loop, we have compared it not to IaKARI, but to the structures of *Pseudomonas aeruginosa* KARI (PaKARI^{apo}), which was crystallized in the apo form[18] and *S. exigua* (SeKARI^{holo}), which was crystallized in Chapter 1 with cofactor, substrate analogue, and metals.[21] Additionally, AvKARI has higher sequence identity (Table S4-1) to these than to IaKARI. Below, residues will be discussed using the numberings in AvKARI; the corresponding numberings in PaKARI and SeKARI are given in Table S4-2.

The structure of AvKARI^{met} aligns very well with that of PaKARI^{apo}, deviating only in the β2αB-loop and at the C-terminus. The former deviation is probably a result of increased flexibility in this loop due to the lack of a cofactor. Both structures are in an 'open' state, with their active sites exposed to solvent. SeKARI^{holo} is in the 'closed' state and to align AvKARI^{met} to SeKARI^{holo}, we were forced to align the N- and C-termini separately, as in Figure 4-7. As with IaKARI, it is clear that each domain is separate and rigid, with a hinge between them and again this hinge is the α1-helix. This provides an unambiguous answer to the question above: the bending of the α1-helix is driven not by the binding of metal ions to the acidic residues on this helix, but by the binding of the cofactor.

It is worth noting that of the six metal-binding residues, only K130, D190, and Q226' undergo rotamer changes upon metal binding, and that these are much more subtle than the changes observed with IaKARI. The residues of PaKARI^{apo} are therefore more pre-organized for metal binding, although it is not clear whether this is an artifact of crystallization or a function of the protein structure. Nevertheless, this is further evidence that metal ion binding is not a driver of a larger rearrangement.

The gross conformational change between AvKARI^{met} and SeKARI^{holo} closely resembles that observed in IaKARI, as does the correlated reorganization of the α3′α4′-loop

and the αF-helix around the nicotinamide amide group. The reorganization in AvKARI^{met}/SeKARI^{holo} however, involves a more intricate hydrogen bond network (Figure 4-9b, Figure 4-11). In place of A250', AvKARI has S249', the side chain of which is capable of hydrogen-bonding. Likewise, P135 of the αF-helix, which did not participate in the rearrangement, is replaced with H134, which is reoriented upon cofactor binding and hydrogen bonds to the side chains of both S249' and S26. Furthermore, Q27, instead of hydrogen-bonding to the backbone amine of G133 in the cofactor-free structure, hydrogen-bonds to the carbonyl of A131. This hydrogen bond is not interrupted by the binding of the cofactor. A different network of interactions is present in PaKARI^{apo}, but inspection of the electron density unambiguously shows this to be a modelling error (Figure S4-1).

It is interesting that this mechanism (open apo state and closed cofactor-bound state) resembles that of the plant class II KARIs rather than the one from the bacterial class II KARIs. The plant KARIs also undergo rearrangement of the α3+1-helix upon binding of NADP, though the plant KARI conformational change is more complex, involving a reorganization of a long C-terminal tail, absent from class I KARIs, upon cofactor binding.[24] In contrast, *E. coli* KARI adopts an open conformation after binding of cofactor and substrate, the reverse of what is observed in the plant KARIs and the bacterial class I KARIs discussed in this chapter. Nonetheless, the conformational change in *E. coli* KARI also involves rearrangements centered around the nicotinamide amide. In *E. coli* KARI, the α2+1- and α3+1-helices interact with the β6αF-loop in the absence of cofactor; the binding of nicotinamide into the active site collapses the β6αF-loop (as briefly discussed by Wong *et al*.[25]) and breaks this interaction (Figure 4-12). Until more structural information is available, it is impossible to say whether *E. coli* KARI is unique in this regard or what evolutionary pathway led it to have this induced fit mechanism opposite to that of the other KARIs.

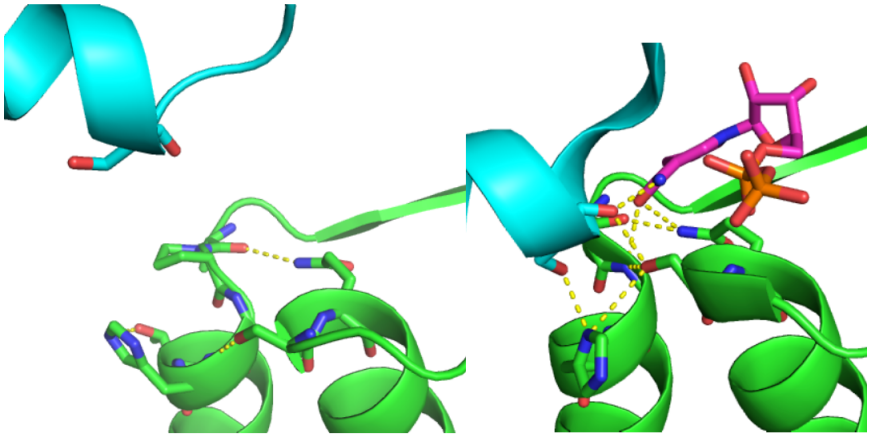


Figure 4-11. Structural rearrangements around the NADP nicotinamide in AvKARI/SeKARI. As with IaKARI, binding of the nicotinamide of NADP (magenta) to the αA-helix and the β6αF-loop recruits the α3-helix of the other chain of the dimer (cyan); in these KARIs more hydrogen bonds are present and undergo further rearrangements.

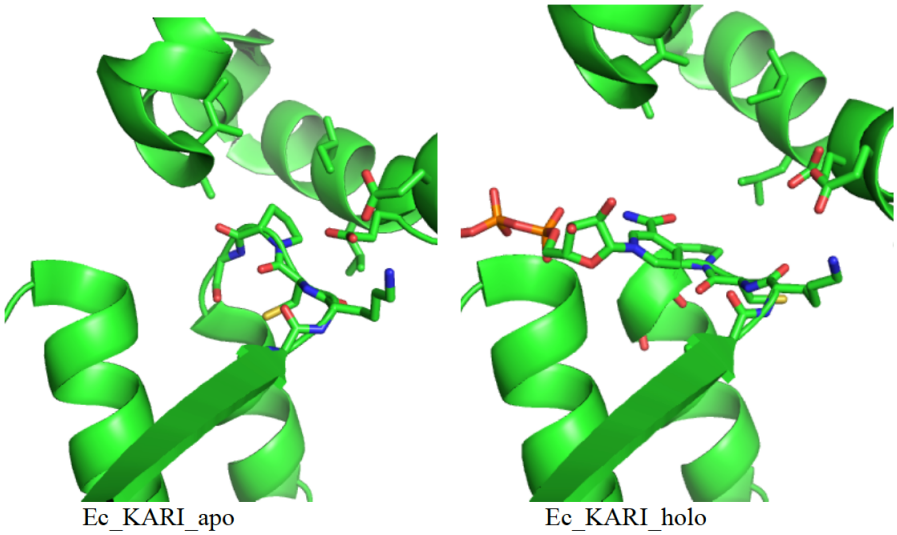


Figure 4-12. Structural rearrangements around the NADP nicotinamide in EcKARI, as described above. Structures from Guddat and coworkers.[25,49]

Conclusions

Understanding and controlling the cofactor specificity of KARI enzymes has been a goal for protein engineers for nearly two decades.[48] These efforts have, by necessity, relied heavily on incomplete structural data to guide mutagenesis strategies. With the addition of the structures described in the present paper to those previously available, we now have representatives of all known canonical KARI cofactor specificity loop lengths. These data show that the six-residue specificity loops use a distinct binding arrangement for the phosphate of NADP, providing a clear structural explanation for results from previous cofactor engineering efforts.[21] Furthermore, structures of two of the naturally NAD-utilizing KARIs show that insertions near the N-terminus of the specificity loop probably determine cofactor specificity, whereas the laboratory engineering of NAD preference involved specific loop substitutions.

These structures also provide insight into the large conformational change associated with KARI cofactor binding and catalysis. The rearrangement observed in these class I KARIs facilitates hydrogen bond interactions with the nicotinamide group and effectively closes the enzyme active site through motion of the α1-helix. This geometry resembles that observed in plant KARIs,[24] but is distinct from that of *E. coli* KARI, which goes from a closed conformation to an open one upon cofactor binding.[25] The principal study on order of ligand binding in KARIs was conducted using *E. coli* KARI,[1] raising the question of whether the kinetic mechanism for ordered binding has been maintained throughout divergent evolution. The structures provided in this chapter answer several outstanding questions about the structural and functional diversity of KARIs and bring us closer to a comprehensive understanding of this enzyme class.

References

1. Chunduru, S. K., Mrachko, G. T., and Calvo, K. C. (1989) Mechanism of ketol acid reductoisomerase - Steady-state analysis and metal-ion requirement. *Biochemistry* **28**, 486-493
2. Brinkmann-Chen, S., Cahn, J. K. B., and Arnold, F. H. (2014) Uncovering rare NADH-preferring ketol-acid reductoisomerases. *Metabolic Engineering* **26**, 17-22
3. Dumas, R., Biou, V., Halgand, F., Douce, R., and Duggleby, R. G. (2001) Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase. *Accounts of Chemical Research* **34**, 399-408
4. Dumas, R., Butikofer, M. C., Job, D., and Douce, R. (1995) Evidence for 2 catalytically different magnesium-binding sites in acetohydroxy acid isomeroreductase by site-directed mutagenesis. *Biochemistry* **34**, 6026-6036
5. Halgand, F., Dumas, R., Biou, V., Andrieu, J. P., Thomazeau, K., Gagnon, J., Douce, R., and Forest, E. (1999) Characterization of the conformational changes of acetohydroxy acid isomeroreductase induced by the binding of Mg^{2+} ions, NADPH, and a competitive inhibitor. *Biochemistry* **38**, 6025-6034
6. Mrachko, G. T., Chunduru, S. K., and Calvo, K. C. (1992) The pH-dependence of the kinetic-parameters of ketol acid reductoisomerase indicates a proton shuttle mechanism for alkyl migration. *Archives of Biochemistry and Biophysics* **294**, 446-453
7. Tyagi, R., Lee, Y. T., Guddat, L. W., and Duggleby, R. G. (2005) Probing the mechanism of the bifunctional enzyme ketol-acid reductoisomerase by site-directed mutagenesis of the active site. *FEBS Journal* **272**, 593-602
8. Hasegawa, S., Uematsu, K., Natsuma, Y., Suda, M., Hiraga, K., Jojima, T., Inui, M., and Yukawa, H. (2012) Improvement of the redox balance increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions. *Applied and Environmental Microbiology* **78**, 865-875
9. Park, J. H., Lee, K. H., Kim, T. Y., and Lee, S. Y. (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proceedings of the National Academy of Sciences* **104**, 7797-7802
10. Park, J. H., and Lee, S. Y. (2010) Fermentative production of branched chain amino acids: A focus on metabolic engineering. *Applied Microbiology and Biotechnology* **85**, 491-506
11. Vogt, M., Haas, S., Klaffl, S., Polen, T., Eggeling, L., van Ooyen, J., and Bott, M. (2014) Pushing product formation to its limit: Metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metabolic Engineering* **22**, 40-52
12. Dumas, R., Cornillon-Bertrand, C., Guigue-Talet, P., Genix, P., Douce, R., and Job, D. (1994) Interactions of plant acetohydroxy acid isomeroreductase with reaction intermediate analogs - Correlation of the slow, competitive, inhibition-

kinetics of enzyme-activity and herbicidal effects. *Biochemical Journal* **301**, 813-820

13. Epelbaum, S., Chipman, D. M., and Barak, Z. (1996) Metabolic effects of inhibitors of two enzymes of the branched-chain amino acid pathway in *Salmonella typhimurium*. *Journal of Bacteriology* **178**, 1187-1196
14. Lee, Y. T., Ta, H. T., and Duggleby, R. G. (2005) Cyclopropane-1,1-dicarboxylate is a slow-, tight-binding inhibitor of rice ketol-acid reductoisomerase. *Plant Science* **168**, 1035-1040
15. Liu, X. H., Chen, P. Q., Wang, B. L., Li, Y. H., Wang, S. H., and Li, Z. M. (2007) Synthesis, bioactivity, theoretical and molecular docking study of 1-cyano-*N*-substituted-cyclopropanecarboxamide as ketol-acid reductoisomerase inhibitor. *Bioorganic and Medicinal Chemistry Letters* **17**, 3784-3788
16. Liu, X. H., Doug, W. L., Zhang, C. Y., Wang, B. L., Ma, Y., and Li, Z. M. (2008) KARI and its inhibitors. *Progress in Chemistry* **20**, 1788-1797
17. Atsumi, S., Hanai, T., and Liao, J. C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89
18. Ahn, H. J., Eom, S. J., Yoon, H. J., Lee, B. I., Cho, H. J., and Suh, S. W. (2003) Crystal structure of class I acetohydroxy acid isomeroreductase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **328**, 505-515
19. Taylor, W. R. (2000) A deeply knotted protein structure and how it might fold. *Nature* **406**, 916-919
20. Virnau, P., Mallam, A., and Jackson, S. (2011) Structures and folding pathways of topologically knotted proteins. *Journal of Physics - Condensed Matter* **23**
21. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951
22. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Research* **42**, D222-D230
23. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metabolic Engineering* **13**, 345-352
24. Leung, E. W. W., and Guddat, L. W. (2009) Conformational changes in a plant ketol-acid reductoisomerase upon Mg^{2+} and NADPH binding as revealed by two crystal structures. *Journal of Molecular Biology* **389**, 167-182
25. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G., and Guddat, L. W. (2012) Bacterial and plant ketol-acid reductoisomerases have different mechanisms

of induced fit during the catalytic cycle. *Journal of Molecular Biology* **424**, 168-179

26. Dumas, R., Job, D., Ortholand, J. Y., Emeric, G., Greiner, A., and Douce, R. (1992) Isolation and kinetic-properties of acetohydroxy acid isomeroreductase from spinach (*Spinacia oleracea*) chloroplasts overexpressed in *Escherichia coli*. *Biochemical Journal* **288**, 865-874
27. Bursey, E. H., and Burgess, B. K. (1998) Characterization of a variant iron protein of nitrogenase that is impaired in its ability to adopt the MgATP-induced conformational change. *Journal of Biological Chemistry* **273**, 16927-16934
28. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D - Biological Crystallography* **66**, 125-132
29. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D - Biological Crystallography* **62**, 72-82
30. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D – Biological Crystallography* **67**, 271-281
31. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallographica Section D - Biological Crystallography* **66**, 22-25
32. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674
33. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallographica Section D – Biological Crystallography* **58**, 1772-1779
34. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D – Biological Crystallography* **67**, 355-367
35. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**, 235-242
36. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D - Biological Crystallography* **60**, 2126-2132
37. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
38. Kumar, P., and Bansal, M. (2012) HELANAL-Plus: A web server for analysis of helix geometry in protein structures. *Journal of Biomolecular Structure and Dynamics* **30**, 773-783

39. Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., and Pebay-Peyroula, E. (1997) The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 angstrom resolution. *EMBO Journal* **16**, 3405-3415
40. Thomazeau, K., Dumas, R., Halgand, F., Forest, E., Douce, R., and Biou, V. (2000) Structure of spinach acetohydroxyacid isomeroreductase complexed with its reaction product dihydroxymethylvalerate, manganese and (phospho)-ADP-ribose. *Acta Crystallographica Section D - Biological Crystallography* **56**, 389-397
41. Naylor, C. E., Gover, S., Basak, A. K., Cosgrove, M. S., Levy, H. R., and Adams, M. J. (2001) $NADP^+$ and NAD^+ binding to the dual coenzyme specific enzyme Leuconostoc mesenteroides glucose 6-phosphate dehydrogenase: Different interdomain hinge angles are seen in different binary and ternary complexes. *Acta Crystallographica Section D - Biological Crystallography* **57**, 635-648
42. Kavanagh, K. L., Klimacek, M., Nidetzky, B., and Wilson, D. K. (2003) Structure of xylose reductase bound to NAD^+ and the basis for single and dual co-substrate specificity in family 2 aldo-keto reductases. *Biochemical Journal* **373**, 319-326
43. Falini, G., Fermani, S., Ripamonti, A., Sabatino, P., Sparla, F., Pupillo, P., and Trost, P. (2003) Dual coenzyme specificity of photosynthetic glyceraldehyde-3-phosphate dehydrogenase interpreted by the crystal structure of A(4) isoform complexed with NAD. *Biochemistry* **42**, 4631-4639
44. Wallen, J. R., Paige, C., Mallett, T. C., Karplus, P. A., and Claiborne, A. (2008) Pyridine nucleotide complexes with *Bacillus anthracis* coenzyme A-disulfide reductase: A structural analysis of dual NAD(P)H specificity. *Biochemistry* **47**, 5182-5193
45. Britten, R. J., Rowen, L., Williams, J., and Cameron, R. A. (2003) Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National Academy of Sciences* **100**, 4661-4665
46. Tóth-Petróczy, Á., and Tawfik, D. S. (2013) Protein insertions and deletions enabled by neutral roaming in sequence space. *Molecular Biology and Evolution* **30**, 761-771
47. Wierenga, R. K., De Maeyer, M. C. H., and Hol, W. G. J. (1985) Interaction of pyrophosphate moieties with α-helixes in dinucleotide-binding proteins. *Biochemistry* **24**, 1346-1357
48. 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**, 83-89
49. Tyagi, R., Duquerroy, S., Navaza, J., Guddat, L. W., and Duggleby, R. G. (2005) The crystal structure of a bacterial class II ketol-acid reductolsomerase: Domain conservation and evolution. *Protein Science* **14**, 3089-3100
50. Bond, C. S. (2003) TopDraw: A sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311-312

C h a p t e r 5

ARTIFICIAL DOMAIN DUPLICATION REPLICATES EVOLUTIONARY HISTORY OF KETOL-ACID REDUCTOISOMERASES

Material from this chapter appears in Cahn J. K. B., Brinkmann-Chen S., Buller A. R., Arnold F. H. (2016). **Artificial domain duplication replicates evolutionary history of ketol-acid reductoisomerases**, *Protein Science* Epub ahead of print, and is reprinted by permission from John Wiley and Sons.

Abstract

The duplication of protein structural domains has been proposed as a common mechanism for the generation of new protein folds. A particularly interesting case is the class II ketol-acid reductoisomerase (KARI), which putatively arose from an ancestral class I KARI by duplication of the C-terminal domain and corresponding loss of obligate dimerization. As a result, the class II enzymes acquired a deeply embedded figure-of-eight knot. To test this evolutionary hypothesis we constructed a novel class II KARI by duplicating the C-terminal domain of a hyperthermostable class I KARI. The new protein is monomeric, as confirmed by gel filtration and X-ray crystallography, and has the deeply-knotted class II KARI fold. Surprisingly, its catalytic activity is nearly unchanged from the parent KARI. This provides strong evidence in support of domain duplication as the mechanism for the evolution of the class II KARI fold and demonstrates the ability of domain duplication to generate topological novelty in a function-neutral manner.

Introduction

The mechanisms by which new protein folds arise and diversify are difficult to test experimentally and remain important open questions in evolutionary biology, molecular biology, and biophysics.[1-4] The presence of internal symmetry in proteins has led to the hypothesis that ancient domains arose by fusion of short peptide ancestors[5,6] and that they further diversified by fusion with other duplicate or non-duplicate domains.[7-10] One possible outcome for tandem domain duplication is that it provides a structural alternative to homo-oligomerization, with multiple monomers combined on a single polypeptide chain forming a pseudo-oligomer.[8,11] This evolutionary model has been used as a basis to deconstruct

proteins in fold classes with high symmetry, such as β-propellers and β-trefoils, into oligomerizing monomers[5,12-15] and for reconstructing novel proteins by monomer fusion.[5,15-20] Modular domains have also been recombined in new stoichiometries, such as with ankyrin repeats.[21,22]

The evidence supporting domain duplication as a mechanism for fold evolution is strongest in cases where monomeric and dimeric forms of a protein are known and are shown to be functionally equivalent,[1,11] such as with chorismate mutases[23] and isocitrate dehydrogenases.[24] A particularly interesting example is that of the class II ketol-acid reductoisomerase (KARI).[25,26] KARI (EC 1.1.1.86, also known as acetohydroxyacid isomeroreductase (AHAIR)) is a key enzyme in the biosynthetic pathway for the branched-chain amino acids where it converts (*S*)-2-acetolactate (S2AL) into (*R*)-2,3-dihydroxyisovalerate (RDHIV) via an alkyl migration and an NAD(P)-dependent reduction.[27] KARIs are divided into two classes on the basis of sequence length and oligomerization state (Figure 5-1). Following nomenclature introduced in 2003 by Ahn and coworkers,[25] class I KARIs are dimers (or oligomers of dimers) with monomers that are ~340 amino acids in length and composed of an N-terminal Rossmann domain that binds NAD(P) and a C-terminal knotted domain. Each C-terminal domain intertwines with the knotted domain of a second monomer and contributes to the active site (two per dimer).[25,28] The monomeric class II KARIs, on the other hand, possess two copies of the knotted domain that fold together to mimic the dimeric state of the class I enzymes and circumvent the requirement for dimerization to form the (single) active site (Figure 5-1).[26,29] The term pseudosesquimer, from the Latin root referring to one-and-a-half, might be more appropriate to describe the class II KARIs, but we use the term pseudodimer to be in line with prior literature.[8] The all-α knotted domain has no known sequence or structural similarity to any stand-alone domain in nature, suggesting that its stability or ability to fold is compromised in the absence of the partner Rossmann domain. The two knotted domains are extensively intertwined and when fused form the eponymous figure-of-eight knot. The knot of class II KARIs was the first protein knot identified[30] and remains the most deeply embedded knot observed in a protein structure.[31]

Knotted proteins offer a particularly challenging test of the domain duplication hypothesis. Whereas the simple or symmetrical proteins discussed above fold in a fashion

similar to how their individual repeats fold,[20,21] introduction of topological complexity such as a protein knot can dramatically alter the folding pathway by introducing the requirement that one portion of the polypeptide thread through another.[32,33] Previously, it was shown that a synthetic trefoil-knotted pseudodimer could be formed by domain duplication when a 9-residue linker was introduced to span the two domains.[33] However, domain duplication must produce a functional protein even before subsequent adaptation to the new topology, and no studies have addressed how readily this can occur. Though a domain-duplicated protein having identical knotted domain sequences was a necessary intermediate in the evolution of the class II KARI fold, the putative duplicated domains within class II KARIs now share only 12–24% sequence identity. The functionality of an evolutionarily naïve KARI pseudodimer is particularly uncertain because the active sites each include residues from both knotted domains. As such, the catalytic activity will be highly sensitive to the relative positioning of these domains. Here we sought to test domain duplication in a contemporary class I KARI and determine whether it could generate a folded, knotted, and, most importantly, functional class II KARI protein.

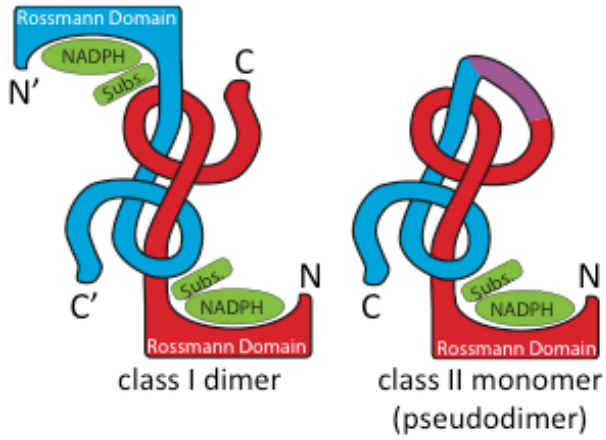


Figure 5-1. Cartoon representation of the class I KARI dimer (left) and class II KARI monomer (right), illustrating the figure-of-eight knot, the duplication of the knotted domain, and the fact that the substrate makes contact with the Rossmann domain as well as with components of both knotted domains.

Materials and Methods

Cloning and expression

Each 2Ia_KARI gene construct was codon-optimized using Integrated DNA Technologies' (San Diego, CA) Codon Optimization Tool and obtained from IDT as gBlocks. The gBlocks were cloned into pET22b(+) vectors using Gibson cloning.[34] Electro-competent *E. coli* BL-21(DE3) 'E. cloni Express' cells from Lucigen (Middleton, WI) cells were transformed with the constructs. For expression, 1 L of Terrific Broth media was inoculated with 5 mL saturated overnight culture and grown at 37 °C for 4 h before induction with IPTG to a final concentration of 500 μM. Upon induction, the temperature was dropped to 20 °C and protein expression proceeded for 24 h prior to centrifugal harvesting.

Protein purification and characterization

Protein purification via the C-terminal His_6-tags was performed as described.[35] The protein concentration was determined with the Bradford assay. Values for k_{cat} were determined at 85 °C with the standard assay in the presence of 200 mM NAD(P)H, 2.5 mM S2AL, and 10 mM $MgCl_2$ in 100 mM phosphate buffer, pH 7.0. Michaelis-Menten constants for the substrate were determined in the presence of varying S2AL concentrations and 200 mM NADPH. Michaelis-Menten constants for the cofactors were measured under saturating substrate conditions and varying cofactor concentrations in a fluorescence plate reader at 25 °C.

To determine the half-denaturation temperatures (T_{50}) of IaKARI and 2Ia_KARIs, 40mL aliquots of purified enzyme were transferred into PCR tubes. Each tube was assigned a specific incubation temperature on the block of an Eppendorf Master Cycler PCR machine. The measurements were conducted in duplicates. The tubes were incubated in their slots for 60 min, and the reactions were then quenched on ice. Residual activity was determined with the standard activity assay at 25 °C in a plate reader. T_{50} is defined as the temperature at which 50% of the initial activity is retained after 60 min incubation.

Gel filtration analysis

The oligomerization states of IaKARI and the 2Ia_KARI constructs were determined by size exclusion chromatrography (SEC) on a Superdex 75 10/300 GL column with 100 mM phosphate buffer, pH 7.0 using an Akta purifier FPL system (GE Healthcare). Myoglobin (16.7 kDa) and BSA (66.5 kDa) served as standards.

Crystallization and X-ray data collection

For crystallization, purified protein was buffer exchanged into 20 mM Tris pH 7 and concentrated to ~110 mg/mL using an Ultracon centrifugal concentrator (Millipore, Billerica, MA). High-throughput screening of crystallization conditions was conducted at the Beckman Molecular Observatory at the California Institute of Technology. The best crystals were obtained from 2 M sodium formate, 100 mM sodium acetate, pH 4.6. For cryoprotection, the crystal was briefly soaked in a mother liquor containing 20% glycerol prior to flash-freezing in liquid nitrogen. Diffraction data were collected using a Dectris Pilatus 6M detector on beamline 12-2 at the Stanford Synchrotron Radiation Laboratory SSRL at 100 K. Diffraction datasets were integrated with XDS[36] and scaled using SCALA.[37]

Structure solution

For molecular replacement, one monomer from the prior holo structure of Ia_KARI (PDB 4XDZ) was split into two PDB files containing the Rossmann and knotted domains, and Phaser molecular replacement[38] was performed, searching for one copy of the Rossmann domain and two copies of the knotted domain on the basis of the Matthews coefficient[39] as calculated in CCP4.[40] After removal of portions of the chain with poor fit to the density, the structure was rebuilt and refined over several iterations of automated refinement with Refmac5[41] (CCP4 suite[40]) and manual refinement in Coot.[42] TLS operators were included in the final round of refinement.[43]

Results

Design of constructs

We selected as a starting point the class I KARI from *Ignisphaera aggregans* (IaKARI), a hyperthermostable, NADH/NADPH-bispecific enzyme with good heterologous expression in *Escherichia coli*.[28,44] In the previous chapter, we obtained high resolution structures of IaKARI in the apo (without co-crystallized cofactor and inhibitor) and holo (cofactor- and inhibitor-bound) forms, which allowed us to determine the location of the interdomain hinge and describe the role of dimerization in the function of the enzyme.[28] The structure of IaKARI is similar to the other class I KARIs that have been crystallized.[25,28,45]

It was unclear how the class I KARI would respond to the addition of approximately 140 additional amino acids. We therefore designed four potential class II KARIs (Figure 5-2). The first construct, 2Ia_KARI-DD (domain duplication), has a simple duplication of the knotted domain using as a start point for the duplication E189, the position previously identified as the interdomain hinge.[28] The design of the other constructs was based on an alignment between the IaKARI holo structure (PDB 4XDZ[28]) and that of the class II KARI from *E. coli* (EcKARI; PDB 3ULK[46]). The EcKARI structure was used because it is only crystallized class II KARI from a bacterial host. For a description of the notation we use to describe the KARI architecture, the previous chapter. Inspection of the EcKARI structure revealed that the α1+1-helix was truncated at its N-terminal end with respect to the corresponding α1'-helix of the IaKARI dimer. Therefore, in the second construct, 2Ia_KARI-HT (helix truncation), we began the duplicate knotted domain at the first position of this helix which could be aligned to a corresponding position of the EcKARI structure, V197. As a result, 2Ia_KARI-HT is eight amino acids shorter than 2Ia_KARI-DD. Next, we noted that in IaKARI, the final 24 amino acid residues of the knotted domain form the α7-helix, which is absent in EcKARI. Instead, EcKARI possesses a 23-residue loop, the α6α1+1-loop, which links the two knotted domains (Figure 5-3). Therefore, in 2Ia_KARI-LT (loop transplant), we replaced the final residue of IaKARI's α6-helix and all subsequent residues with the final residue from EcKARI's α6-helix and the EcKARI α6α1+1-loop. The α7'-helix, at the C-terminus of the duplicate knotted domain, is intact. Comparing the two knotted domains of EcKARI, one final structural difference is

clear. The second knotted domain of EcKARI does not have an α5-helix, which is replaced by the long α4+1α6+1-loop (Figure 5-3), and the α6+1 helix is shifted and rotated with respect to an alignment with IaKARI. Our fourth and final construct, 2Ia_KARI-2L (two [transplanted] loops), is constructed from 2Ia_KARI-LT by identification of the final positions which can be aligned around this dissymmetry, K264 and E315 of IaKARI, and replacing the 50 residues between them, which compose the C-terminal of the α4'-helix, the α4'α5'-loop, the α5'-helix, the α5'α6'-loop, the α6'-helix, and the α6'α7'-loop, with the 44 residues composing the C-terminal of the α4+1-helix, the α4+1α6+1-loop, the α6+1helix, and the α6+1α7+1 loop of EcKARI.

All 2Ia_KARI gene constructs were codon-optimized for *E. coli* and cloned into the pET22(b)+ vector in frame with a C-terminal His_6-tag.[47] Expression under standard conditions for heterologous expression of KARIs[45] produced folded 2Ia_KARI-DD and 2Ia_KARI-LT with yields comparable to wild-type IaKARI after Ni-NTA purification. 2Ia_KARI-HT and 2Ia_KARI-2L had considerably decreased expression, and were not characterized further.

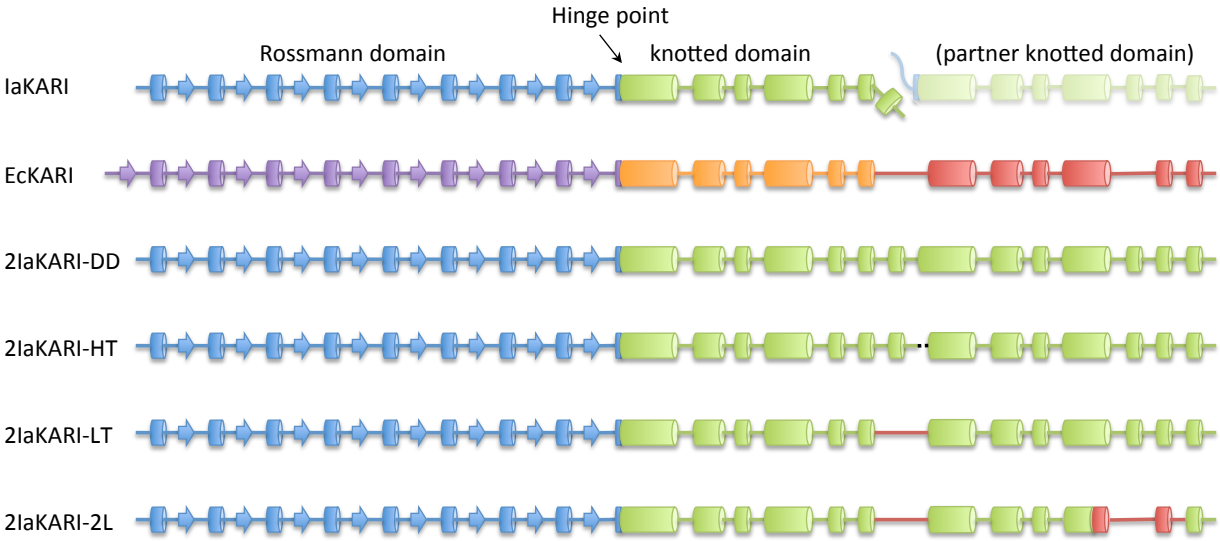


Figure 5-2. Secondary structure schematics of KARIs described in this chapter. Helices are indicated by cylinders and beta-strands by arrows. Sequence elements are colored according to their source and domain: IaKARI sequence elements are blue (Rossmann domain) or green (knotted domain) and EcKARI sequence elements are purple (Rossmann domain), orange (first knotted domain), or red (second knotted domain). The Rossmann domain of the second IaKARI monomer is truncated for clarity. Full sequence alignment is shown in Figure S5-2.

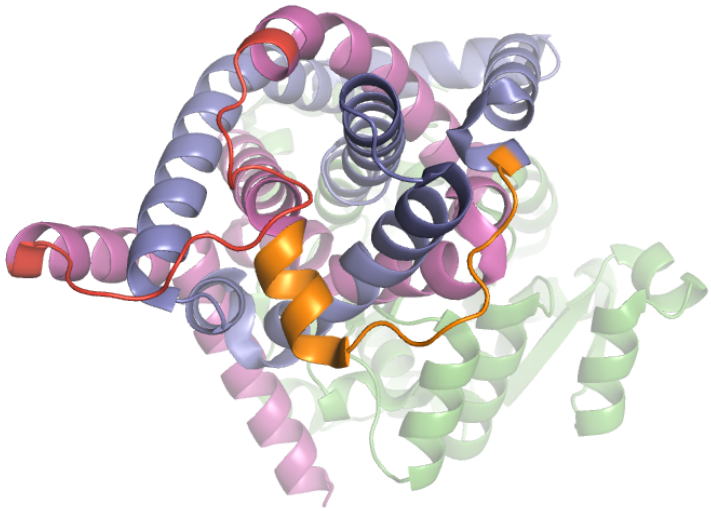


Figure 5-3. A view of EcKARI (PDB 3ULK) showing the two structural deviations from class I KARIs. The Rossmann domain is shown in green, the first knotted domain is shown in purple, and the second is shown in pink. The α6α1+1-loop is shown in orange, connecting the two knotted domains, and the α4+1α6+1-loop is shown in red.

Characterization of the 2Ia_KARIs

Both artificially domain-duplicated 2Ia_KARI-DD and 2Ia_KARI-LT enzymes are highly active for the transformation of (*S*)-2-acetolactate (S2AL) to (*R*)-2,3-dihydroxyisovalerate (RDHIV) at 85 °C. We measured the kinetic parameters of both and compared them to IaKARI wild type (Table 5-1). The NADPH and NADH cofactor K_M values remained below the instrument's 1 mM detection limit. On NADPH, 2Ia_KARI-DD showed a similar k_{cat} to wild type; the k_{cat} of 2Ia_KARI-LT was 3-fold lower (Table 5-1). Like the wild-type enzyme, 2Ia_KARI-LT displays roughly equal catalytic efficiency on NADPH and NADH, while 2Ia_KARI-DD's cofactor preference was slightly shifted toward NADPH (Table 5-2). Surprisingly, the two 2Ia_KARIs had lower K_M values for the substrate S2AL than did the wild type, with a 3-fold reduction observed for 2Ia_KARI-DD and a 6-fold reduction for 2Ia_KARI-LT. Both 2Ia_KARIs therefore had higher catalytic efficiencies for S2AL than the wild-type enzyme (Table 5-1).

2Ia_KARI-DD and 2Ia_KARI-LT were highly thermostable with temperatures of half-activity loss (T_{50}) of 92 and 88 °C, respectively, only slightly less than wild-type IaKARI's 95 °C (Table 5-1). Size exclusion chromatography confirmed the dimeric state of wild-type IaKARI[28]; in contrast, the 2Ia_KARIs elute as monomers in solution (Figure 5-4). The catalytic activity of 2Ia_KARI-DD depended linearly on protein concentration, further indicating that it is monomeric (Figure 5-5).

Table 5-1. Biochemical parameters of IaKARI and two domain-duplicated 2Ia_KARIs.

Enzyme	K_M NADPH [μM]	k_{cat} NADPH [s^{-1}]	k_{cat}/K_M NADPH [$s^{-1}mM^{-1}$]	K_M S2AL [mM]	k_{cat}/K_M S2AL [$s^{-1}mM^{-1}$]	T_{50} [°C]
IaKARI	< 1	3.3	>3,300	3.1	1.1	95
2Ia_KARI-DD	< 1	3.0	>3,000	1.0	3.0	92
2Ia_KARI-LT	< 1	1.0	>1,000	0.5	2.0	88

Each value represents the average of three independent measurements. Enzyme activities were determined in 100 mM potassium phosphate pH 7 with 200 mM NADPH or NADH, 10 mM S2AL, and 10 mM $MgCl_2$. See Materials and Methods for details.

Table 5-2. Comparison between catalytic parameters on NADH and NADPH for selected KARIs.

Enzyme	K_M NADH [μM]	K_M NADPH [μM]	k_{cat} NADH [s^{-1}]	k_{cat} NADPH [s^{-1}]	k_{cat}/K_M NADH [$s^{-1}mM^{-1}$]	k_{cat}/K_M NADPH [$s^{-1}mM^{-1}$]	Catalytic Efficiency Ratio (NADH/NADPH)
IaKARI	< 1	< 1	2.6	3.3	>2,600	>3,300	0.8
2Ia_KARI-DD	< 1	< 1	1.7	3.0	>1,700	>3,000	0.6
2Ia_KARI-LT	< 1	< 1	0.8	1.0	>800	>1,000	0.88

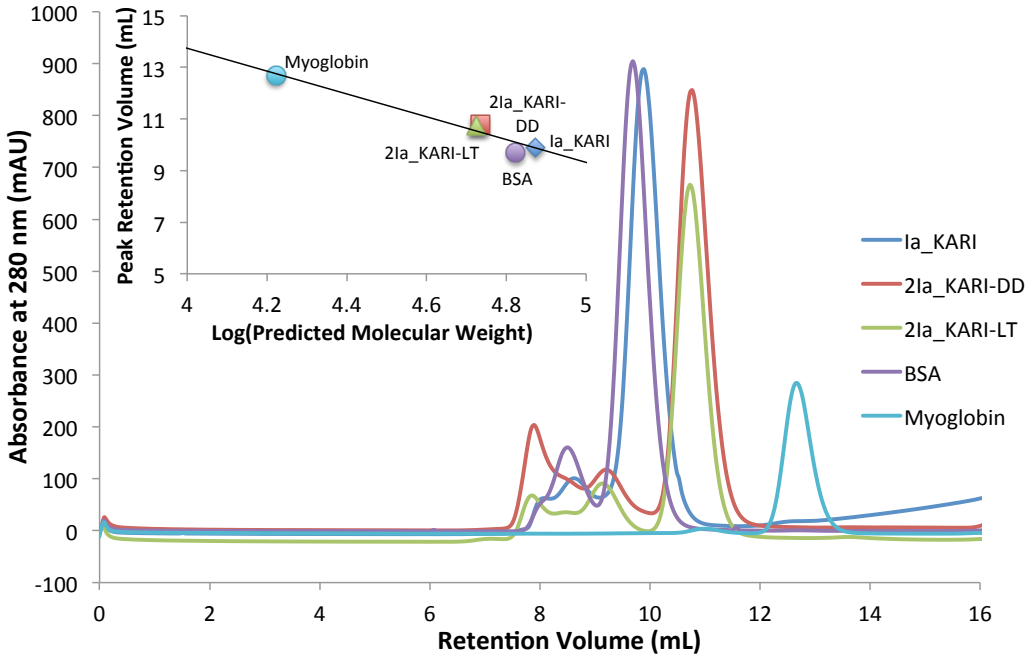


Figure 5-4. Gel filtration chromatography analysis of 2Ia_KARIs. Gel filtration curves are shown for 2Ia_KARI-DD and 2Ia_KARI-LT, along with dimeric IaKARI WT and flanking controls (BSA and myoglobin). Inset shows proper power-law correspondence for predicted monomeric molecular weight.

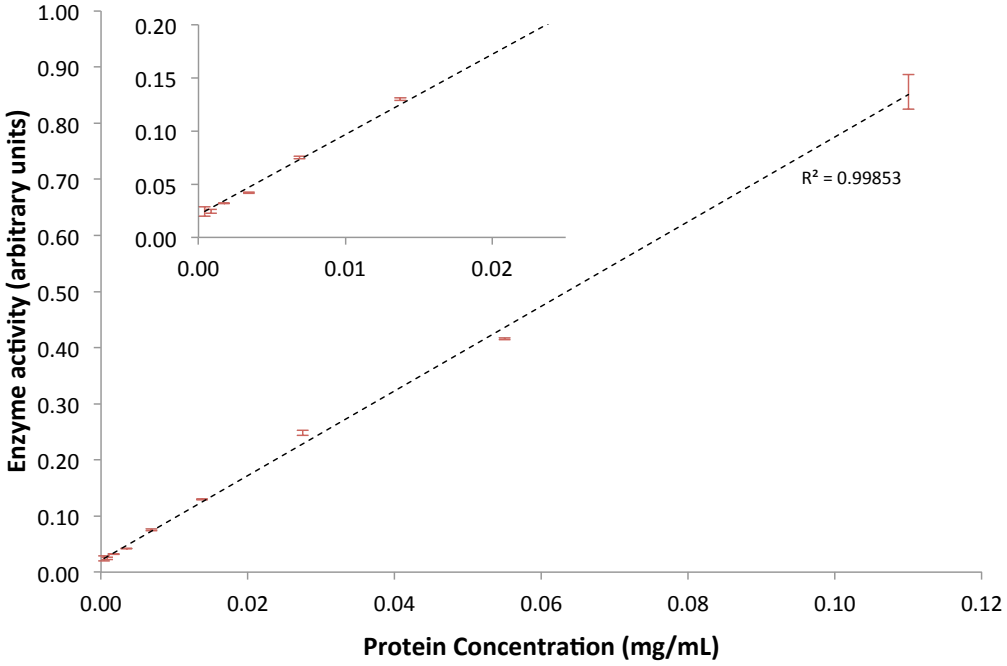


Figure 5-5. Linear relationship between enzyme concentration and activity across a 512-fold range of dilutions for 2Ia_KARI-DD, showing monomeric behavior.

X-ray crystal structure

To confirm that domain duplication had produced a canonical class II KARI fold, we obtained the crystal structure of 2Ia_KARI-DD. This protein was noticeably more soluble than IaKARI: at a protein concentration of 20 mg/mL, the concentration used to crystallize IaKARI,[28] very few wells of our high-throughput crystallization screens gave even amorphous precipitate. However, by increasing the protein concentration to ~110 mg/mL, we were able to obtain two crystals in sodium formate, one of which diffracted to 1.94 Å (Figure 5-6b, Table 5-3). The protein crystallized in the $P\,4_3\,2\,2$ space group with a single molecule of 2Ia_KARI-DD in the asymmetric unit (for crystal packing, see Figure S5-1), and a monomeric biological assembly is predicted by the Protein Interfaces, Surfaces, and Assemblies server (PISA).[48]

The structure of 2Ia_KARI-DD contains NADP and the inhibitor *N*-isopropyloxamate, and its fold closely resembles that of the class II KARIs. Overall, the structure also shows remarkable agreement with the structure of the native IaKARI dimer

in its cofactor-bound state (PDB ID 4XDZ[28]), as shown in Figure 5-6. The Rossmann domain and the active site components from both knotted domains in the new class II enzyme align very well with the IaKARI structure, including the contacts to the cofactor and substrate from the α3+1α4+1-loop. While the overall C^{α} RMSD is 0.3 Å for aligned residues, the deviation is considerably higher in a handful of regions as shown in Figure 5-7. The amino acids immediately before and after the linker as well as those that were in close proximity to the dimeric partner's now-deleted Rossmann domain show the largest perturbations (Figure 5-8). In particular, the α5+1-helix (residues 430–441), exhibits considerable structural heterogeneity, with two distinct parallel conformations observable in the maps, separated by an average of 2.2 Å (Figure 5-9). Interesting, EcKARI lacks an α5+1-helix altogether, possessing instead an extended α4+1α6+1-loop (Figure 5-3). In the wild-type IaKARI structures, the α5'-helix forms significant contacts with the αD'- and αE'-helices of the dimeric partner's Rossmann domain. Elimination of this domain removes these stabilizing contacts from the α5+1-helix, leading to the conformational heterogeneity observed in the domain-duplicated protein. This instability may underlie the eventual evolution away from helicity observed in EcKARI, which lacks an α5+1-helix, as do all other crystallized class II KARIs.[28,29,49]

The residues connecting the first and second knotted domains are disordered in the 2Ia_KARI-DD crystal structure. Interestingly, the final residue of the first knotted domain that could be resolved, M329, is also the final residue that could be resolved in the IaKARI structure, so the linkage between the domains is composed of the same six residues that made up the disordered C-terminus of the parent protein. The first three residues of the second knotted domain are also disordered. This is in contrast to the apparent structural rigidity in the interdomain linker that has evolved in EcKARI and other naturally-occurring class II KARIs.[29,49]

EcKARI undergoes an induced fit conformational change upon binding NADP opposite that of IaKARI and all other KARIs for which the analysis is possible: it goes from an active-site-closed conformation in the apo state to an open one in the cofactor-bound state, whereas IaKARI and others adopt a closed conformation upon cofactor binding by bringing together the Rossmann and knotted domains.[28,46] This raised the question of whether the 2Ia_KARIs would adopt the behavior of their parent class I enzyme

or that of the closest known class II enzyme. The structure of 2Ia_KARI-DD in its cofactor-bound state clearly shows a closed active site, suggesting that the induced fit behavior of the parent class I protein has been preserved upon domain duplication.

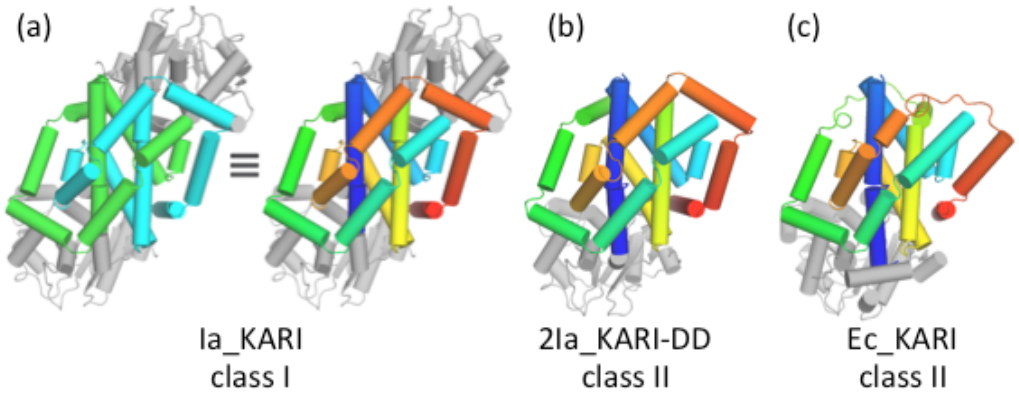


Figure 5-6. Structures of class I and class II KARs, highlighting their knotted domains. Rossmann domains are shown in grey. The class I IaKARI (a; PDB 4DXZ[28]) is shown with knotted domains colored separately to illustrate the dimeric interaction or with spectral coloring corresponding to the class II KARIs. The class II 2Ia_KARI-DD (b) was engineered by duplication and concatenation of the IaKARI knotted domain. The natural class II EcKARI (c; PDB 3ULK[34]) has evolved away from perfect symmetry.

Table 5-3. Data collection and refinement statistics for the structure of 2Ia_KARI-DD (PDB 5E4R). Values in parentheses refer to the highest resolution shell.

Data collection	
Space Group	$P\,4_3\,2\,2$
a, b, c (Å)	103.9, 103.9, 142.1
α, β, γ (°)	90, 90, 90
Resolution (Å)	103.9-1.94 (1.99-1.94)
$R_{p.i.m.}$ (%)	2.6 (22.2)
$\langle i \rangle / \langle \sigma i \rangle$	7.39 (0.74)
Completeness (%)	99.2 (94.4)
Redundancy	10.3 (8.2)
Refinement	
No. reflections	613,895
Rwork/Rfree (%)	15.8/17.9 (21.8/27.1)
No. atoms	
Protein	3753
Ligand	68
Water	237
RMSD	
Bond lengths (Å)	0.020
Bond angles (°)	1.128
Ramachandran plot	
Favored	466
Allowed	14
Outliers	0

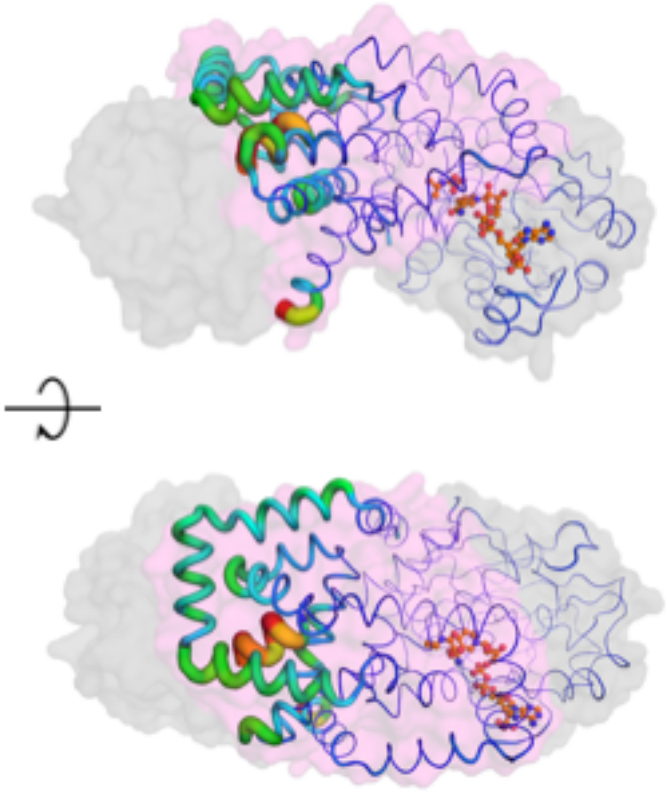


Figure 5-7. The 2Ia_KARI-DD backbone trace highlighting structural differences between it and IaKARI. The outline of IaKARI is shown in grey (Rossmann domains) and magenta (knotted domains). The backbone of 2Ia_KARI-DD has been aligned to it. The coloring and width of the 2Ia_KARI-DD backbone are proportional to the C^{α}- C^{α} distances between the structures (plotted in Figure 5-8). Differences cluster around the interface with the eliminated Rossmann domain. The NADP cofactor and *N*-isopropyloxamate inhibitor are drawn with ball-and-stick, showing that the active site remains entirely unperturbed.

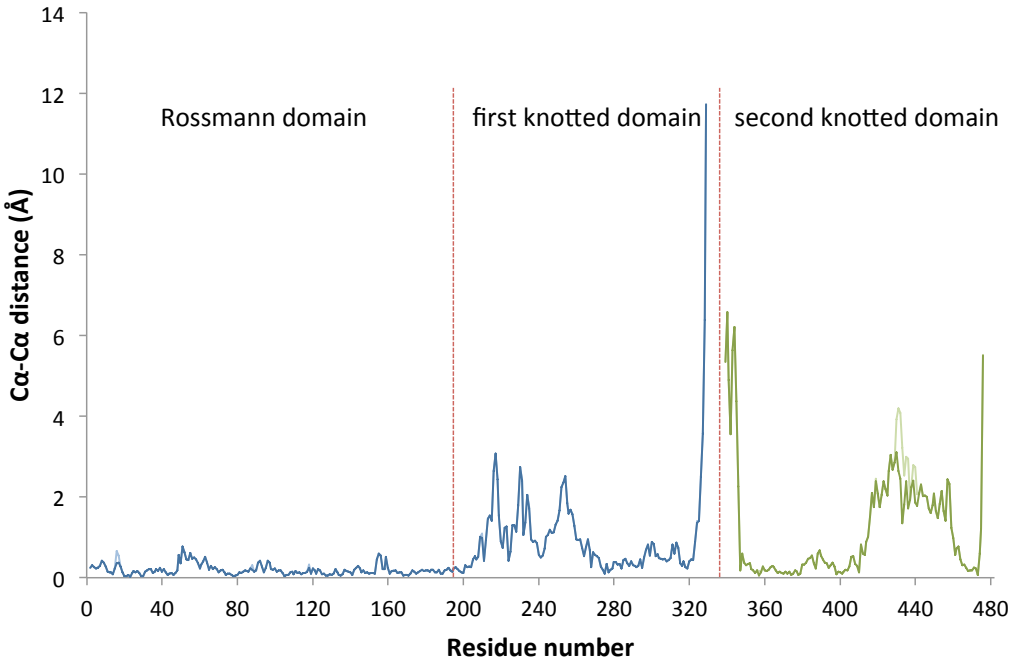


Figure 5-8. The C^{α}- C^{α} distances between 2Ia_KARI-DD and IaKARI. The lines are colored by which monomer of IaKARI is closer, blue for the first monomer and green for the second. When multiple C^{α} are present for a given residue, due to alternate conformations, the greater distance is shown in a lighter color. The residues between 330 and 339 are not structurally resolved.

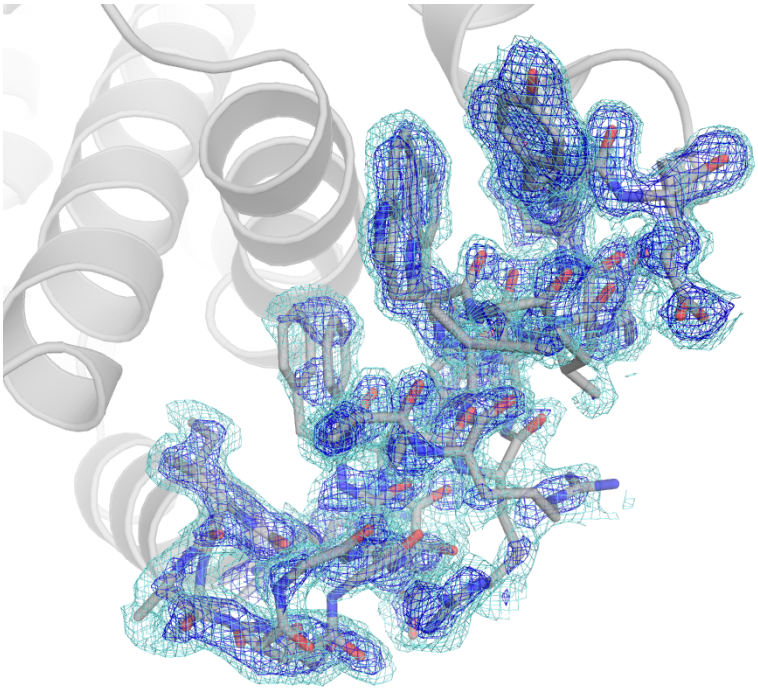


Figure 5-9. The α5+1-helix, showing two distinct conformations visible in the $2F_o$-F_c map, contoured to 1.2σ (blue) and 0.65σ (cyan).

Discussion

Several potential forces have been suggested as favoring the evolution of pseudodimeric folds. Among these are removal of the constraint of symmetry, which allows the two pseudo-monomers to specialize,[8,11,25] or an increase in the range of movement that can be tolerated with a covalent linkage.[8] This demonstration that a crude domain-duplicated enzyme can retain catalytic efficiency comparable to that of the wild-type protein supports the simpler explanation that domain duplication was an essentially neutral change in protein structure. The role of neutral drift in evolution has long been appreciated at a sequence level,[51-53] and our results lend experimental support to the extension of the concept of neutral drift to the generation of structural novelty by domain duplication.[54] Other forms of protein topological variation that do not necessarily come at the cost of function are well known: proteins may tolerate proteolysis or expression from artificially split genes,[55] single-chain proteins have been created from heterodimers,[56] many proteins are known to form functional domain-swapped dimers,[57] and proteins can often be expressed as circular permutants,[4,58] all while preserving function. These studies all

showcase that protein function is contingent upon an arrangement of key residues and associated secondary structures and depends less on the connectivity among them.[55,59] As such, as long as function is preserved, evolution is blind to topology.

While topological changes may have minimal effects on function or even stability of the fold,[55] their effects on the kinetics of folding can be significant. When King *et al*. engineered a comparatively simple and shallow protein knot by dimerization of HP0242, they found that the dimer folded approximately 20 times slower than a control unknotted monomer.[33,60,61] It remains an open question how the folding pathway of 2Ia_KARI-DD resembles or differs from that of the parent IaKARI, or how it compares to that of evolutionarily mature class II KARIs. It would also be interesting to see how 2Ia_KARI-DD might further evolve and whether it takes on the structural or sequence features seen in known class II KARIs, allowing us to replay in real time the evolution of this unique enzyme family.

References

1. Andreeva, A., and Murzin, A. G. (2006) Evolution of protein fold in the presence of functional constraints. *Current Opinion in Structural Biology* **16**, 399-408
2. Deeds, E. J., and Shakhnovich, E. I. (2010) A structure-centric view of protein evolution, design, and adaptation. in *Advances in Enzymology*, John Wiley & Sons, Inc. pp 133-191
3. Abeln, S., and Deane, C. M. (2005) Fold usage on genomes and protein fold evolution. *Proteins: Structure, Function, and Bioinformatics* **60**, 690-700
4. Grishin, N. V. (2001) Fold change in evolution of protein structures. *Journal of Structural Biology* **134**, 167-185
5. Lee, J., and Blaber, M. (2011) Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proceedings of the National Academy of Sciences* **108**, 126-130
6. Lupas, A. N., Ponting, C. P., and Russell, R. B. (2001) On the evolution of protein folds: Are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *Journal of Structural Biology* **134**, 191-203
7. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., and Teichmann, S. A. (2004) Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology* **14**, 208-216
8. Taylor, W. R., and Sadowski, M. I. (2010) Protein products of tandem gene duplication: A structural view. in *Evolution After Gene Duplication* (Dittmar, K., and Liberles, D. eds.), Wiley-Blackwell, Hoboken, NJ. pp 133-162
9. Bashton, M., and Chothia, C. (2002) The geometry of domain combination in proteins. *Journal of Molecular Biology* **315**, 927-939
10. McLachlan, A. D. (1972) Repeating sequences and gene duplication in proteins. *Journal of Molecular Biology* **64**, 417-437
11. Abraham, A.-L., Pothier, J., and Rocha, E. P. C. (2009) Alternative to homo-oligomerisation: The creation of local symmetry in proteins by internal amplification. *Journal of Molecular Biology* **394**, 522-534
12. Richter, M., Bosnali, M., Carstensen, L., Seitz, T., Durchschlag, H., Blanquart, S., Merkl, R., and Sterner, R. (2010) Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. *Journal of Molecular Biology* **398**, 763-773
13. Yadid, I., and Tawfik, D. S. (2007) Reconstruction of functional β-propeller lectins via homo-oligomeric assembly of shorter fragments. *Journal of Molecular Biology* **365**, 10-17
14. Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner, R. (2001) Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nature Structural and Molecular Biology* **8**, 32-36
15. Höcker, B. (2014) Design of proteins from smaller fragments — Learning from evolution. *Current Opinion in Structural Biology* **27**, 56-62

16. Yadid, I., and Tawfik, D. S. (2010) Functional β-propeller lectins by tandem duplications of repetitive units. *Protein Engineering Design and Selection*
17. Bharat, T. A. M., Eisenbeis, S., Zeth, K., and Höcker, B. (2008) A βα-barrel built by the combination of fragments from different folds. *Proceedings of the National Academy of Sciences* **105**, 9942-9947
18. Arnold, T., Poynor, M., Nussberger, S., Lupas, A. N., and Linke, D. (2007) Gene duplication of the eight-stranded β-barrel OmpX produces a functional pore: A scenario for the evolution of transmembrane β-barrels. *Journal of Molecular Biology* **366**, 1174-1184
19. Höcker, B., Claren, J., and Sterner, R. (2004) Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. *Proceedings of the National Academy of Sciences* **101**, 16448-16453
20. Longo, Liam M., Kumru, Ozan S., Middaugh, C. R., and Blaber, M. Evolution and design of protein structure by folding nucleus symmetric expansion. *Structure* **22**, 1377-1384
21. Aksel, T., and Barrick, D. (2014) Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophysical Journal* **107**, 220-232
22. Parmeggiani, F., Huang, P.-S., Vorobiev, S., Xiao, R., Park, K., Caprari, S., Su, M., Seetharaman, J., Mao, L., Janjua, H., Montelione, G. T., Hunt, J., and Baker, D. (2015) A general computational approach for repeat protein design. *Journal of Molecular Biology* **427**, 563-575
23. Ökvist, M., Dey, R., Sasso, S., Grahn, E., Kast, P., and Krengel, U. (2006) 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: Novel fold topology revealed. *Journal of Molecular Biology* **357**, 1483-1499
24. Yasutake, Y., Watanabe, S., Yao, M., Takada, Y., Fukunaga, N., and Tanaka, I. (2002) Structure of the monomeric isocitrate dehydrogenase: Evidence of a protein monomerization by a domain duplication. *Structure* **10**, 1637-1648
25. Ahn, H. J., Eom, S. J., Yoon, H. J., Lee, B. I., Cho, H. J., and Suh, S. W. (2003) Crystal structure of class I acetohydroxy acid isomeroreductase from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* **328**, 505-515
26. Tyagi, R., Duquerroy, S., Navaza, J., Guddat, L. W., and Duggleby, R. G. (2005) The crystal structure of a bacterial class II ketol-acid reductolsomerase: Domain conservation and evolution. *Protein Science* **14**, 3089-3100
27. Chunduru, S. K., Mrachko, G. T., and Calvo, K. C. (1989) Mechanism of ketol acid reductoisomerase - Steady-state analysis and metal-ion requirement. *Biochemistry* **28**, 486-493
28. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W., and Arnold, F. H. (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochemical Journal* **468**, 475-484

29. Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D., and Pebay-Peyroula, E. (1997) The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog determined at 1.65 angstrom resolution. *EMBO Journal* **16**, 3405-3415
30. Taylor, W. R. (2000) A deeply knotted protein structure and how it might fold. *Nature* **406**, 916-919
31. Virnau, P., Mallam, A., and Jackson, S. (2011) Structures and folding pathways of topologically knotted proteins. *Journal of Physics - Condensed Matter* **23**
32. Yeates, T. O., Norcross, T. S., and King, N. P. (2007) Knotted and topologically complex proteins as models for studying folding and stability. *Current Opinion in Chemical Biology* **11**, 595-603
33. King, N. P., Jacobitz, A. W., Sawaya, M. R., Goldschmidt, L., and Yeates, T. O. (2010) Structure and folding of a designed knotted protein. *Proceedings of the National Academy of Sciences* **107**, 20732-20737
34. Gibson, D. G. (2011) Chapter Fifteen - Enzymatic assembly of overlapping DNA fragments. in *Methods in Enzymology* (Christopher, V. ed.), Academic Press. pp 349-361
35. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic Engineering* **13**, 345-352
36. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D - Biological Crystallography* **66**, 125-132
37. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D - Biological Crystallography* **62**, 72-82
38. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674
39. Matthews, B. W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology* **33**, 491-497
40. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**, 235-242
41. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D – Biological Crystallography* **67**, 355-367
42. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D - Biological Crystallography* **60**, 2126-2132

43. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D – Biological Crystallography* **62**, 439-450
44. Brinkmann-Chen, S., Cahn, J. K. B., and Arnold, F. H. (2014) Uncovering rare NADH-preferring ketol-acid reductoisomerases. *Metabolic Engineering* **26**, 17-22
45. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951
46. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G., and Guddat, L. W. (2012) Bacterial and plant ketol-acid reductoisomerases have different mechanisms of induced fit during the catalytic cycle. *Journal of Molecular Biology* **424**, 168-179
47. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nature Biotechnology* **6**, 1321-1325
48. Krissinel, E., and Henrick, K. (2007) Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute.
49. Leung, E. W. W., and Guddat, L. W. (2009) Conformational changes in a plant ketol-acid reductoisomerase upon Mg^{2+} and NADPH binding as revealed by two crystal structures. *Journal of Molecular Biology* **389**, 167-182
50. Lim, N. C. H., and Jackson, S. E. (2015) Mechanistic insights into the folding of knotted proteins *in vitro* and *in vivo*. *Journal of Molecular Biology* **427**, 248-258
51. Kimura, M. (1984) *The neutral theory of molecular evolution*, Cambridge University Press
52. Bloom, J. D., Romero, P. A., Lu, Z., and Arnold, F. H. (2007) Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biology Direct* **2**, 17
53. Gupta, R. D., and Tawfik, D. S. (2008) Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods* **5**, 939-942
54. Liberles, D. A., Teichmann, S. A., Bahar, I., Bastolla, U., Bloom, J., Bornberg-Bauer, E., Colwell, L. J., de Koning, A. P. J., Dokholyan, N. V., Echave, J., Elofsson, A., Gerloff, D. L., Goldstein, R. A., Grahnen, J. A., Holder, M. T., Lakner, C., Lartillot, N., Lovell, S. C., Naylor, G., Perica, T., Pollock, D. D., Pupko, T., Regan, L., Roger, A., Rubinstein, N., Shakhnovich, E., Sjölander, K., Sunyaev, S., Teufel, A. I., Thorne, J. L., Thornton, J. W., Weinreich, D. M., and Whelan, S. (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Protein Science* **21**, 769-785
55. Carey, J., Lindman, S., Bauer, M., and Linse, S. (2007) Protein reconstitution and three-dimensional domain swapping: Benefits and constraints of covalency. *Protein Science* **16**, 2317-2333

56. Somoza, J. R., Jiang, F., Tong, L., Kang, C.-H., Cho, J. M., and Kim, S.-H. (1993) Two crystal structures of a potently sweet protein: Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. *Journal of Molecular Biology* **234**, 390-404

57. Liu, Y., and Eisenberg, D. (2002) 3D domain swapping: As domains continue to swap. *Protein Science* **11**, 1285-1299

58. Lindqvist, Y., and Schneider, G. (1997) Circular permutations of natural protein sequences: Structural evidence. *Current Opinion in Structural Biology* **7**, 422-427

59. Heringa, J., and Taylor, W. R. (1997) Three-dimensional domain duplication, swapping and stealing. *Current Opinion in Structural Biology* **7**, 416-421

60. Sułkowska, J. I., Noel, J. K., and Onuchic, J. N. (2012) Energy landscape of knotted protein folding. *Proceedings of the National Academy of Sciences* **109**, 17783-17788

61. Wang, L.-W., Liu, Y.-N., Lyu, P.-C., Jackson, S. E., and Hsu, S.-T. D. (2015) Comparative analysis of the folding dynamics and kinetics of an engineered knotted protein and its variants derived from HP0242 of *Helicobacter pylori*. *Journal of Physics: Condensed Matter* **27**, 354106

C h a p t e r 6

MUTATIONS IN ADENINE-BINDING POCKETS ENHANCE CATALYTIC PROPERTIES OF NAD(P)-DEPENDENT ENZYMES

Material from this chapter appears in Cahn J. K. B., Baumschlager A., Brinkmann-Chen S., Arnold F. H. (2016). **Artificial domain duplication replicates evolutionary history of ketol-acid reductoisomerases**, *Protein Engineering, Design and Selection* **29**(1), 31-38, and is reprinted by permission from Oxford University Press.

Abstract

NAD(P)-dependent enzymes are ubiquitous in metabolism and cellular processes and are also of great interest for pharmaceutical and industrial applications. Here, we present a structure-guided enzyme engineering strategy for improving catalytic properties of NAD(P)-dependent enzymes toward native or native-like reactions using mutations to the enzyme's adenine-binding pocket, distal to the site of catalysis. Screening single-site saturation mutagenesis libraries identified mutations that increased catalytic efficiency up to 10-fold in seven out of ten enzymes. The enzymes improved in this study represent three different cofactor-binding folds (Rossmann, DHQS-like, and FAD/NAD binding) and utilize both NAD and NADP. Structural and biochemical analyses show that the improved activities are accompanied by minimal changes in other properties (cooperativity, thermostability, pH optimum, uncoupling), and initial tests on two enzymes (ScADH6 and EcFucO) show improved functionality in *Escherichia coli*.

Introduction

Engineering novel or improved metabolic pathways often changes the demands placed on enzymes evolved to carry out their natural functions in specific contexts. For instance, it has been proposed that enzymatic K_M values have evolved to match physiological substrate concentrations,[1] which can change as a result of heterologous expression or pathway engineering that changes steady-state metabolite concentrations. Such changes in metabolic context might require alterations in enzyme kinetics through protein engineering for optimal metabolic flux and cell physiology. Beyond the room for improvement created by a novel physiological context, it has been hypothesized that the

kinetics of many enzymes have not been maximized by evolution, particularly the 'moderately efficient' enzymes of secondary metabolism where kinetic enhancement is of minimal benefit to host fitness.[2,3] Despite the suggestions that there is potential to improve native activities, engineering enzymes for the direct improvement of activities on their natural substrates under biologically relevant reaction conditions has proven challenging and, to our knowledge, broadly unsuccessful.[2,4,5] Instead, enzyme engineering usually improves reactivity toward non-natural substrates, increases promiscuous, reactivities or alters selectivity.[6-9] Unlike these goals, in which functional changes can often be ascribed to specific remodeling of the active site to accommodate or exclude certain substrates or transition states, the precise structural origins of an enzyme's kinetic properties are more enigmatic and to-date have been resistant to prediction. In this chapter, we empirically identify structural positions in NAD(P)-dependent enzymes where mutations can provide significant boosts in enzyme activity and catalytic efficiency.

NAD(P)-dependent enzymes are involved in a wide range of metabolic reactions, which makes them of interest for pharmaceutical and industrial applications. Protein engineering has been used to study and change cofactor binding to NAD(P)-dependent enzymes.[10-13] Previous work from the Arnold laboratory on engineering of cofactor specificity of ketol-acid reductoisomerase (KARI) enzymes revealed that mutation at a single position on a helix that runs parallel to the cofactor adenine moiety improved the catalytic activity of several KARIs, including both wild type and, in Chapter 1, cofactor-switched.[11,12] This finding was replicated in another KARI by Reiße *et al*.[14]

While screening a random mutant library of the *Arabidopsis thaliana* glyoxylate reductase (AtGR1) prepared for an unrelated study (Chapter 7), we observed similar activating effects from a mutation adjacent to the position corresponding to that in the KARIs. AtGR1 with mutation C68R showed a significant improvement in activity in lysate, which we later established was due largely to a ~5-fold decrease in K_M for the substrate (this work). AtGR1 and KARIs possess similar overall folds, with highly similar Rossmann domains, and both are specific for NADP over NAD. Because it is rare to find mutations that boost the activity of an enzyme for its native reaction, we set out to investigate whether modifications at similar positions with respect to the adenine could improve the activities of enzymes with more diverse folds and cofactor utilization profiles.

These amino acids are situated in the internal lining of the adenine-binding pocket. More specifically, they contain atoms located within 5 Å of the N6 atom of the NAD(P) adenine (see Figure 6-1), but are not involved in determining cofactor preference through interaction with the phosphate or hydroxyl in the 2′-position of the ribose. Libraries of enzyme variants made by site-saturation mutagenesis at these positions can be screened rapidly for increased enzyme activity in lysate. We demonstrate that this simple structure-guided engineering strategy works to improve the activities of a surprising range of enzymes, opening the door to improving the catalytic properties of a broad array of industrially relevant enzymes and metabolic pathways.

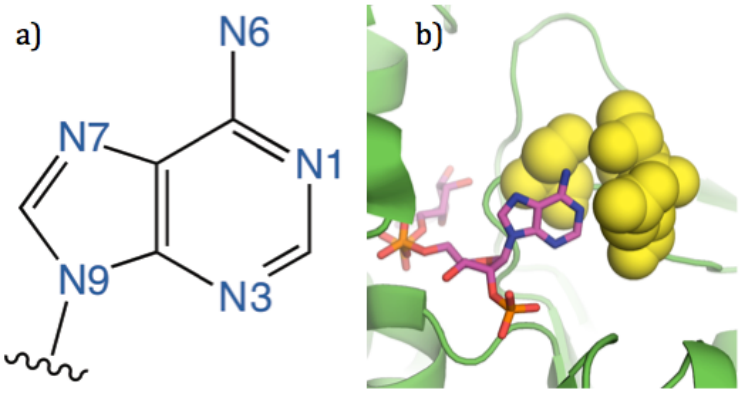


Figure 6-1. The standard numbering of atoms in the NAD(P) adenine moiety (a). An example of the amino acid positions around the adenine N6 (b) mutated in this study (yellow spheres), in this case *S. cerevisiae* ADH6 (ScADH6), whose structure was reported in Valencia *et al.* (PDB 1PIW[15]).

Materials and Methods

Cloning and library construction

All genes were obtained from Integrated DNA Technologies (IDT) as gBlock linear fragments and were cloned into pET22b(+) in frame with the C-terminal His_6-tag for expression in *Escherichia coli* using Gibson cloning[16] with overlap at the T7 promoter and terminator sequences. Mutagenic primers for site-saturation mutagenesis were obtained from IDT and treated as suggested by IDT protocols. Splicing by overlap extension polymerase chain reaction (SOE-PCR) was performed as described previously.[11,17] The quality of the library was assessed using DNA sequencing performed by Laragen (Los Angeles, CA, USA). Standard molecular biology methods were taken from Sambrook *et al*.[18]

Heterologous gene expression for high-throughput screening and protein purification

All expression cultures were grown in Luria-Bertani broth, supplemented with ampicillin for selection (LB+Amp).

For high-throughput screening, pre-cultures of 300 µL LB+Amp were inoculated with single colony forming units (CFUs) in 96-deep-well plates (DWPs) using toothpicks. For each library, 84–88 CFUs were screened, corresponding to 93–95% theoretical library coverage,[19,20] along with the parent protein and the pET22b(+) vector as positive and negative controls. The pre-cultures were grown overnight at 37°C, 200 rpm and 80% humidity. The next day, expression cultures of 600 µL were inoculated with 50 µL of the overnight cultures in 96-DWPs. The pre-cultures were stored at 4°C until the screening was completed to serve as temporary stock from which positive hits were regrown. After incubation of the expression cultures for 4 h at 37°C, 200 rpm and 80% humidity, expression was induced by adding isopropyl thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM. Expression occurred for 21 h at temperatures indicated in Table 6-1 and 200 rpm without humidity control. The expression cultures were harvested through centrifugation, and the DWPs containing the cell pellets were stored at −20°C until screening.

For purified protein, pre-cultures grown overnight at 37°C and 200 rpm were used for inoculation of 200 mL expression cultures to an OD_{600} of 0.05–0.1 and incubated at 37°C and 210 rpm until an OD_{600} of ~0.8 was reached. At this point, the expression cultures were cooled to their expression temperatures (Table 6-1) before induction with IPTG to a final concentration of 0.5 mM and growth for an additional 21 h. The expression cultures were harvested by centrifugation, the supernatant was discarded and the pellets were frozen at −20°C until further use.

Table 6-1. Expression temperatures and screening conditions used for respective enzymes.

Enzyme	**Expression Temperature**	**Substrate**	**Assay Buffer**
AtGR1	25 °C	Sodium glyoxylate	100 mM HEPES, pH 7.8
ScADH6	20 °C	*Trans*-cinnamaldehyde	50 mM sodium phosphate, pH 7.0
LlAdhA	20 °C	Isobutyraldehyde	50 mM sodium phosphate, pH 7.0
DmAdhA	20 °C	Isobutyraldehyde	50 mM sodium phosphate, pH 7.0
KpDhaT	20 °C	Isobutyraldehyde	50 mM sodium phosphate, pH 7.0
EcFucO	20 °C	Furfural	50 mM sodium phosphate, pH 7.0
TeXR	20 °C	D-xylose	50 mM potassium phosphate, pH 7.0
LsNOX	20 °C	Oxygen gas	50 mM sodium phosphate, pH 7.0

Enzyme assays and high-throughput screening

Escherichia coli cells were resuspended in the respective assay buffer (Table 6-1) containing 750 mg/L lysozyme, 10 mg/L DNaseI and 2 mM $MgCl_2$. Lysis was accomplished at 37°C for 1 h. Enzyme activities were then assayed by monitoring NAD(P)H consumption in the presence of the respective substrate (Table 6-1) and 250 μM NAD(P)H at 340 nm in a plate reader.

Thermostability was measured through determination of T_{50}, the temperature where the enzyme activity is reduced to 50% of its initial activity after incubation for 10 min. The pH optimum was determined using a selection of four different buffer systems to cover the relevant pH scale from pH 3 to 10 (pH 3–6: 50 mM sodium citrate buffer, pH 6–8: 50 mM sodium phosphate buffer, pH 8–9: 50 mM Tris–HCl buffer, pH 9.2–10: 50 mM carbonate–bicarbonate buffer). All observations are averages of at least three replicates.

Protein purification and enzyme kinetics

Escherichia coli cell pellets were resuspended in 10 mL buffer A (25 mM Tris, 100 mM NaCl, 20 mM imidazole, pH 7.4) and lysed by sonication. The lysate was centrifuged, and the enzymes were purified via their C-terminal His_6-tag using High Performance (HP) Ni-NTA Sepharose columns (GE Healthcare, Waukesha, WI, USA) on an Äkta Xpress FPLC (GE Healthcare). The concentration of purified protein was determined using the Bradford assay (Bio-Rad, Hercules, CA, USA).

For rate measurements, k_{cat} values were determined using the same assay conditions as above with saturating cofactor and substrate, while Michaelis–Menten constants were determined by varying each individually. At least six cofactor concentrations and at least five substrate concentrations were used for these determinations, and all measurements were performed at least three times. MatLab (Mathworks, Natwick, MA, USA) was used for parameter fitting.

Protein crystallization and structure determination

Screening of crystallization conditions was conducted at the Beckman Molecular Observatory at the California Institute of Technology using commercially available crystal screens. Crystallization occurred with purified $EcFucO^{M185C}$ at a concentration of 15 mg/mL, 10 mM NAD^+ and 10 mM isobutyraldehyde using the sitting drop method at ambient temperature. Crystals were obtained with 12% PEG 3350 and 200 mM NH_4Cl as precipitant. The crystals were soaked with mother liquor containing 17% glycerol and 6.9 mM NAD^+ before flash freezing in liquid nitrogen. Diffraction data were collected using a Dectris Pilatus 6 M detector on beamline 12–2 at the Stanford Synchrotron Radiation

Laboratory at 100 K. Diffraction datasets were integrated with XDS[21] and scaled using SCALA.[22]

The structure of EcFucO (PDB 1RRM[23]) was used for molecular replacement. Refinement was conducted by iterating automatic refinement with Refmac5 (CCP4 suite) and manual refinement using Coot.[24] The structure was deposited in the RCSB Protein Data Bank with accession code 5BR4.

In vivo growth assays

For the *in vivo* growth assays, cells were grown in 24-well round-bottom plates (Invitrogen, Carlsbad, CA, USA). In each well, 3 mL of LB broth with 100 μg/mL ampicillin and 500 μM IPTG were inoculated with 10 μL of saturated overnight culture. After 3 h of growth, an additional 3 mL of LB (with ampicillin and IPTG at the same concentrations) were added, which contained 10 mM of furfural or trans-cinnamaldehyde. After 10 h, the OD_{600} of a 200 μL aliquot was measured on a plate reader (Tecan, Männedorf, Switzerland). All observations are averages of at least five replicates.

Results

Improvement of catalytic properties through mutations around adenine N6

In addition to the KARI enzymes previously studied, we selected eight distinct enzymes representing a range of cofactor preferences and cofactor-binding folds to test whether mutations at positions around the N6 nitrogen of adenine led to higher activity. The enzymes selected for this study include enzymes previously studied in our group and enzymes with potential for industrial applications. *Saccharomyces cerevisiae* cinnamyl alcohol dehydrogenase (ScADH6) is used by yeast for the detoxification of aromatic aldehydes.[25] *Arabidopsis thaliana* glyoxylate reductase (AtGR1) is involved in the glyoxylate-glycolate shuttle for the regulation of photosynthesis.[26] *Lactococcus lactis* and *Drosophila melanogaster* alcohol dehydrogenases (LlAdhA and DmADH) are promiscuous alcohol dehydrogenases previously investigated for use in microbial isobutanol production.[27] *Klebsiella pneumoniae* propanediol dehydrogenase (KpDhaT) is one of the main enzymes for the metabolic pathway in *K. pneumoniae* that enables the species to metabolize glycerol as a sole source of carbon and energy by reduction of 3-

hydroxypropanal to propane-1,3-diol.[28] *Escherichia coli* lactaldehyde reductase (EcFucO) catalyzes the inter-conversion between L-lactaldehyde and L-1,2-propanediol during the anaerobic dissimilation of fucose[29] and aerobic growth on L-1,2-propanediol.[30,31] It also reduces furfural, the dehydration product of xylose and an important fermentation inhibitor in sugar syrups derived from woody biomass, to the less toxic furfuryl alcohol.[31] Xylose reductase is the first enzyme of xylose metabolism in fungi; the enzyme from *Talaromyces emersonii* (TeXR) is a highly active and thermostable member of this family.[32] NAD(P)H oxidase from *Lactobacillus sanfranciscensis* (LsNOX) is an important enzyme in maintaining redox balance[33] and has been used for $NAD(P)^+$ regeneration.[34] Tables 6-2 and 6-3 also include information on two KARIs, EcIlvC from Bastian *et al.*[11] and MrKARI from Reiße *et al.*,[14] which come from the two distinct structural classes of KARIs (Chapter 5). EcIlvC is also notable for undergoing a cofactor-induced conformational change opposite that observed in other KARIs studied (Figure 4-8c).[35]

For each enzyme, we identified residues within 5 Å of the N6 atom of the NAD(P) adenine, based on published crystal structures or homology models produced using the SWISS-MODEL server.[36] A site-saturation mutagenesis library was generated at each position listed in Table 6-2, and the enzyme variants were screened for activity on both NADH and NADPH using the substrates indicated in Table 6-1. Mutants with significant activity enhancements toward either cofactor were purified, and Michaelis–Menten kinetic parameters were determined for the cofactors and the substrate. Improved kinetics were found in variants of five out of the eight new enzymes tested, in addition to the KARIs previously studied by Bastian *et al.*[11] and Reiße *et al*.[14] This number is notable given that these mutations are (i) boosting kinetic properties of native or native-like reactions, (ii) distal to the active sites, and (iii) obtained by screening small saturation libraries made at only a very few positions (between 1 and 4; see Table 6-2).

The combined results are summarized in Table 6-3 where improvements in k_{cat} (1.2–11.0-fold increase) as well as in the K_M for substrate (1.2–11.5-fold decrease) and cofactor (1.1–7.7-fold decrease) are reported for seven out of ten enzymes. (Here, and elsewhere, we use K_M to refer to both the Michaelis constant and the dissociation constant K_H for enzymes displaying cooperativity. Such enzymes are indicated in Table 6-3. No significant changes in the Hill coefficient were observed for these cooperative enzymes.)

Accordingly, the catalytic efficiencies were increased from 1.2-fold to as much as 83-fold with respect to cofactor or substrate (Table 6-3). The improved enzymes contained Rossmann (KARIs, AtGR1, DmADH, ScADH6), FAD/NAD-binding (LsNOX) and DHQS-like (EcFucO) cofactor-binding folds and included enzymes with both cofactor preferences. We were unable to find beneficial mutations at the targeted positions in three enzymes: LlAdhA (Rossmann fold), KpDhaT (DHQS-like fold) and TeXR (TIM barrel fold).

Because we identified beneficial mutations at multiple residues in the LsNOX enzymes, we also tested the double and triple combinatorial mutations. All three double mutants showed activity enhancements with respect to wild type, although not necessarily above the single mutants (Table 6-3). The triple mutant had considerably elevated k_{cat} values but expressed quite poorly compared with the wild-type enzyme and the other single and combinatorial mutants (data not shown). Losses in expression or stability upon the accumulation of mutations are to be expected; stability can often be recovered by further mutagenesis, without compromising activity.[37-39] Although expression levels were not quantified here, no other large changes in purified protein yields were observed for the other mutants of LsNOX or any other protein tested.

The evolutionary fitness of an enzyme, that is how it contributes to the survival and fitness of its source, is determined by factors beyond catalytic efficiency, and thus catalytic efficiency may have been sacrificed by evolution to achieve other properties. In this case, improving activity might be expected to come at the expense of those other properties. To test whether other enzyme properties may have been perturbed by the mutations we identified, we analyzed selected enzymes for changes in pH optimum ($DmADH^{V108I}$) and thermostability ($DmADH^{V108I}$, $EcFucO^{M185A}$, and $EcFucO^{M185C}$). We found no significant changes in the improved variants compared with the wild-type enzymes (Figure 6-2, Table 6-4). Additionally, none of the enzymes characterized were shown to be uncoupled—that is, none consumed cofactor in the absence of substrate—although we did not test LsNOX for uncoupling (its molecular oxygen substrate would require a controlled atmosphere for kinetic measurements). We also examined the *in vivo* activities of two of the enzymes (ScADH6 and EcFucO) whose activities could be directly tied to cell survival.[25,31] Under standard *E. coli* expression conditions, no growth defect was observed as a result of the

mutations (Figure 6-3, top row). Upon the addition of a toxic aldehyde (*trans*-cinnamaldehyde or furfural, respectively), growth was enhanced with the mutant ScADH6 and EcFucO enzymes providing improved reductive detoxification ability compared with the wild-type enzymes (Figure 6-3, bottom row).

Table 6-2. The enzymes tested in this study (below the line) and previously reported (above), including the Structural Classification of Proteins (SCOP) classification of their cofactor-binding fold and a list of positions mutated. Positions indicated in bold are those where one or more beneficial muations were discovered.

Enzyme	PDB code	Cofactor-binding fold	Cofactor preferrence	Positions mutated
EcIlvC	3ULK[40]	Rossmann (c.2.1.6)	NADP	**Q110**[11]
MrKARI	None	Rossmann (c.2.1.6)	NADP	**T84**[14]
AtGR1	3DOJ (Cofactor from 3PEF)[26,41]	Rossmann (c.2.1.0)	NADP	**C68**, A69
ScADH6	1PIW[15]	Rossmann (c.2.1.1)	NADP	S253, **T255**, D256
LlAdhA	4EEX (Cofactor from 4GKV)[27,42]	Rossann (c.2.1.1)	NAD	A242, A245
DmAdhA	1MG5[43]	Rossmann (c.2.1.2)	NAD	D65, V66, R104, **V108**
KpDhaT	3BFJ (Cofactor from 3OX4)[28,44]	DHQS-like (e.22.1.2)	NAD	K187
EcFucO	1RRM[23]	DHQS-like (e.22.1.2)	NAD	T140, **M185**
TeXR	1K8C[45] (Homology)	TIM barrel (c.1.7.1)	NADP	F217, A254, Q280, N281
LsNOX	2CDU[46]	FAD/NAD-binding (c.3.1.5)	Bispecific	**I122**, **I155**, V214, **I243**

Abbreviations: *Arabidopsis thaliana* glyoxylate reductase (AtGR1), *Saccharomyces cerevisae* cinnamyl alcohol dehydrogenase (ScADH6), *Lactococcus lactis* alcohol dehydrogenase (LlAdhA), *Drosophila melanogaster* alcohol dehydrogenase DmADH, *Klebsiella pneumonia* 1,3-propanediol dehydrogenase (KpDhaT), *Escherichia coli* lactaldehyde reductase (EcFucO), *Talaromyces emersonii* xylose reductase (TeXR), *Lactobacillus sanfranciscensis* NAD(P)H-oxidase (LsNOX), dihydroquinoate synthase (DHQS), flavin adenine dinucleotide (FAD).

Table 6-3. Kinetic parameters of characterized enzymes. Numbers in parentheses refer to α, the Hill coefficient. Substrate K_M was assayed in the presence of the cofactor with greater k_{cat}/K_M Due to low enzymatic activity with NADPH as cofactor, k_{cat}^{NADPH} and K_M^{NADPH} could not be determined for wild-type and variant enzymes of EcFucO and DmADH. $K_M^{substrate}$ could not be determined for LsNOX due to experimental constraints.

		k_{cat} (min^{-1})		K_M (μM)			k_{cat}/K_M (mM^{-1} min^{-1})			Specificity (NADH / NADPH)
	Mutation	NADH	NADPH	NADH	NADPH	Substrate	NADH	NADPH	Substrate	
AtGR1	Wild Type	7.3 ± 0.9	27 ± 3.9	76 ± 24	20 ± 6	690 ± 160	96 ± 33	1400 ± 450	39 ± 16	0.07 ± 0.03
AtGR1	C68E	2.7 ± 0.2	13 ± 0.8	66 ± 29	10 ± 3	60 ± 11	41 ± 18	1300 ± 400	220 ± 77	0.03 ± 0.02
AtGR1	C68R	13 ± 3.1	7.3 ± 1.8	53 ± 11	24 ± 9	120 ± 46	250 ± 78	300 ± 140	61 ± 36	0.81 ± 0.44
ScADH6	Wild Type	1300 ± 370	16000 ± 2300	130 ± 51	140 ± 15	170 ± 26	10,000 ± 4800	110,000 ± 20,000	94,000 ± 22,000	0.09 ± 0.05
ScADH6	T255K	3200 ± 87	9000 ± 1100	240 ± 13	37 ± 21	56 ± 4	13,000 ± 810	240,000 ± 140,000	160,000 ± 94,000	0.05 ± 0.03
DmADH	Wild Type	7.8 ± 1.6	-	57 ± 6	-	130 ± 33	140 ± 32	-	60 ± 21	-
DmADH	V108I	9.2 ± 3.4	-	42 ± 6	-	100 ± 24	220 ± 87	-	92 ± 43	-
EcFucO	Wild Type	1.8 ± 0.2	-	70 ± 8 (2.9 ± 0.8)	-	1400 ± 130	26 ± 4.1	-	1.3 ± 0.24	-
EcFucO	M185A	6.2 ± 2.0	-	39 ± 4 (2.7 ± 0.6)	-	390 ± 87	160 ± 54	-	16 ± 6.4	-
EcFucO	M185C	6.5 ± 1.8	-	55 ± 2 (3.0 ± 0.3)	-	910 ± 140	120 ± 33	-	7.1 ± 2.3	-
LsNOX	Wild Type	1200 ± 94	890 ± 170	85 ± 12	73 ± 60	-	14,000 ± 2300	12,000 ± 10,000	-	1.2 ± 0.99
LsNOX	I122V	4000 ± 240	1400 ± 280	20 ± 9.3	110 ± 64	-	200,000 ± 93,000	13,000 ± 7800	-	16 ± 12
LsNOX	I155L	2700 ± 520	3500 ± 360	100 ± 80	32 ± 8.4	-	27,000 ± 22,000	150,000 ± 58,000	-	0.18 ± 0.16
LsNOX	I243M	2800 ± 370	3100 ± 120	150 ± 175	56 ± 48	-	19,000 ± 22,000	55,000 ± 48,000	-	0.34 ± 0.49
LsNOX	I122V-I155L	2700 ± 490	5000 ± 540	70 ± 12	120 ± 30	-	39,000 ± 9600	42,000 ± 11,000	-	0.93 ± 0.34
LsNOX	I22V-I243M	5500 ± 290	4500 ± 49	59 ± 22	31 ± 21	-	93,000 ± 35,000	150,000 ± 98,000	-	0.64 ± 0.50
LsNOX	I155L-I243M	2500 ± 290	2400 ± 93	83 ± 7.6	120 ± 24	-	30,000 ± 4500	20,000 ± 4100	-	1.5 ± 0.38
LsNOX	I122V-I155L-I243M	7900 ± 520	11000 ± 360	93 ± 96	140 ± 300	-	85,000 ± 88,000	79,000 ± 170,000	-	1.1 ± 2.6

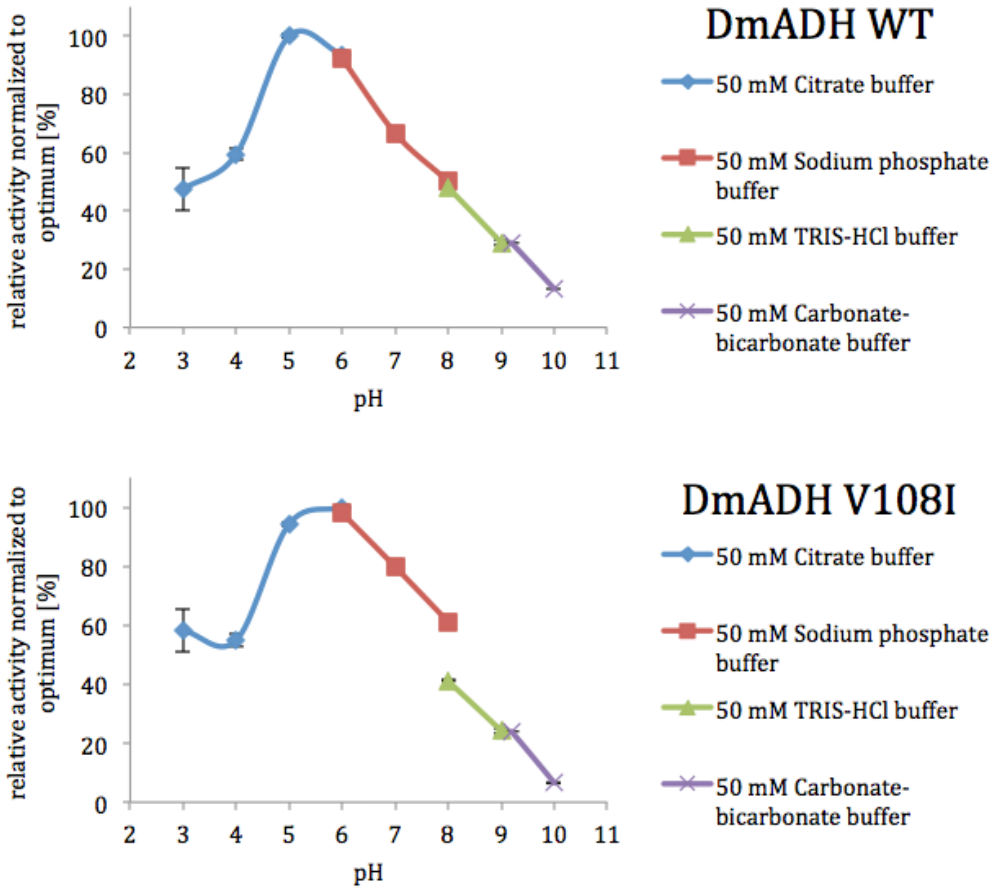


Figure 6-2. The pH activity profiles of DmADH and its V108I mutant.

Table 6-4. Thermostability of indicated enzymes. T_{50} is defined as the temperature at which 50% of the initial activity is retained after 10 min incubation.

Enzyme	T_{50} (°C)
DmADH WT	46.6 ± 0.6
DmADH V108I	48.8 ± 0.8
EcFucO WT	44.7 ± 0.7
EcFucO M185A	45.3 ± 4.0
EcFucO M185C	40.8 ± 2.9

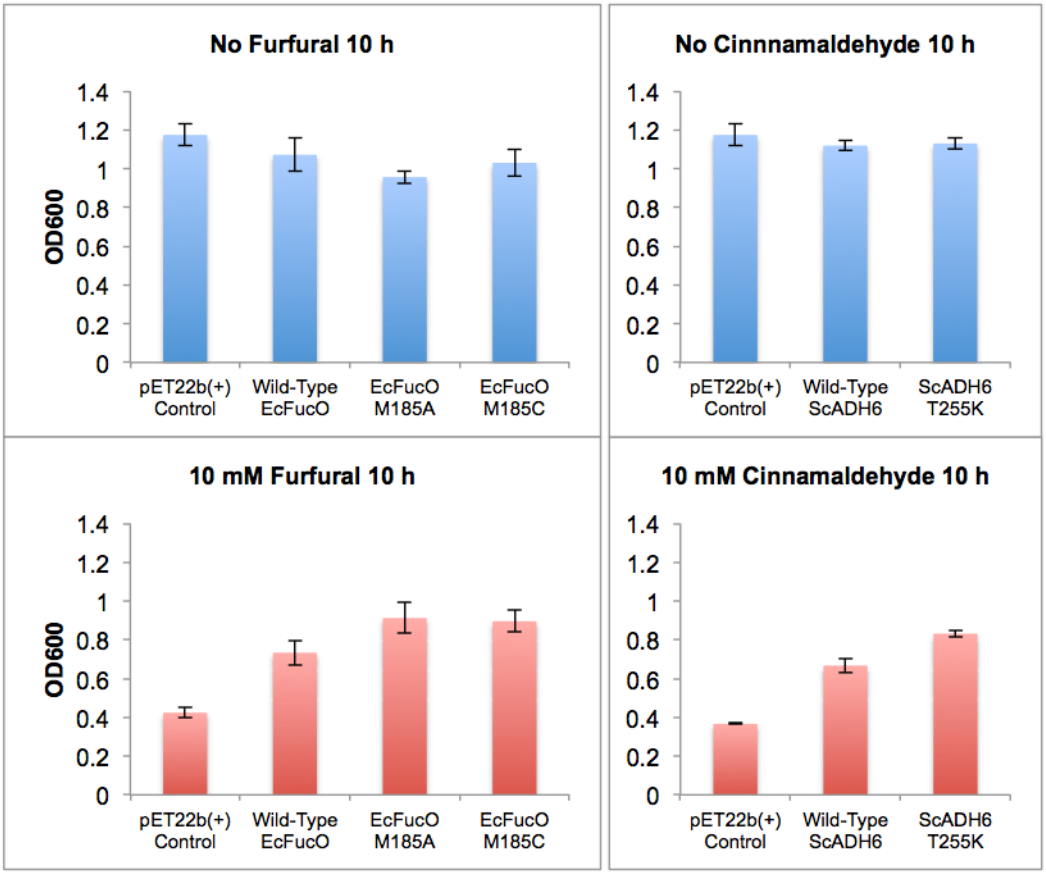


Figure 6-3. *In vivo* growth of cells containing mutant enzymes. Top row shows cells under normal growth conditions. None of the characterized mutations cause a growth defect relative to overexpression of the wild-type enzyme. In the presence of toxic aldehydes (bottom row), the respective wild-type enzymes improved growth rate compared to no enzyme, and the mutant enzymes led to higher growth rates than the wild-type enzymes. Error bars indicate the standard deviation of the measurements.

Structural alterations in EcFucO

The X-ray crystal structure of EcFucO^{M185C} was solved at a resolution of 0.91 Å (Table 6-5), which allowed us to investigate whether the mutation at position M185 caused structural changes in the enzyme or the bound cofactor. In EcFucO^{M185C}, activity was enhanced when M185 was substituted by a smaller and slightly more polar cysteine. The side chain of this cysteine lies almost perfectly along the β- and γ-carbons of the wild-type methionine, and aligning the structure of EcFucO^{M185C} to the previously published wild-

type structure[23] reveals no major changes in the protein structure or the cofactor-binding pocket. However, the axis of the cofactor is tilted slightly, and the adenine is slightly shifted (0.3 Å) in the direction of the active site (Fig. 6-4). This leads to a more significant shift of 1.1 Å at the N1 of the nicotinamide at the other end of the cofactor (Fig. 6-4). As the nicotinamide is the catalytically active part of the cofactor, we assume that this change in position in the active site enhances catalysis in the EcFucO^{M185C} variant, mainly via the 3.6-fold increase in k_{cat}.

In Chapter 1, a similar repositioning of the cofactor was also observed in the crystal structure of SeKARI^{DDV} with mutation I95V close to the adenine of the cofactor.[12] Mutations in SeKARI^{DDV} caused a 1 Å shift of the adenine compared with the parent structure, and we proposed that this readjustment placed the cofactor in a more favorable position for catalysis,[12] although the presence of two additional mutations and a reversal of the cofactor specificity made it impossible to attribute the shift to I95V alone.

Table 6-5. Data collection and refinement statistics for the crystal structure of EcFucO^{M185C} (PDB ID 5BR4). Values in parentheses refer to the highest-resolution shell.

Data Collection	
Space Group	$P\,2_1$
Cell Dimensions	
***a*,*b*,*c* (Å)**	69.7, 68.3, 91.7
***α*,*β*,*γ* (°)**	90, 111.2, 90
Resolution (Å)	85.50-0.91 (0.96-0.91)
$R_{p.i.m.}$ **(%)**	4.4 (166.9)
Mn(*I*)/sd	8.8 (0.4)
Completeness (%)	91.8 (74.6)
Redundancy	2.7 (2.5)
Refinement	
No. of reflections	466,646 (22,931)
R_{work}/R_{free} **(%)**	12.9/14.7 (35.9/37.1)
No. atoms	
Protein	5,818
Ligand/ion	112
Water	1,167
RMSD	
Bond lengths (Å)	0.021
Bond angles (°)	2.035
Ramachandran map analysis	
Favored	774
Allowed	11
Outliers	0

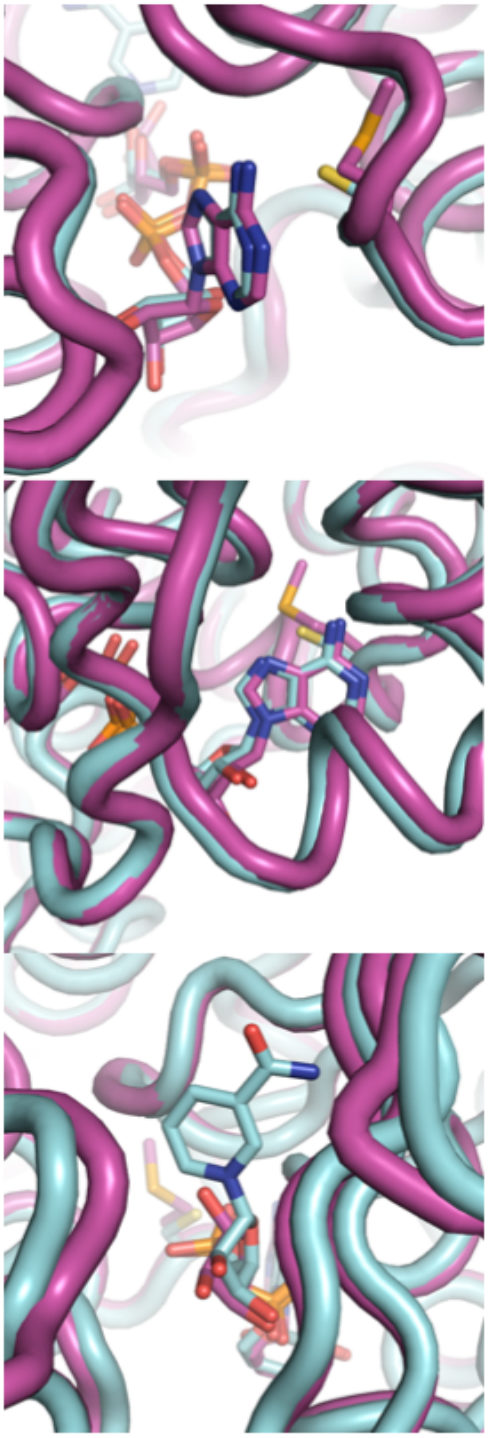


Figure 6-4. Structures of wild-type EcFucO (purple) and the M185C mutant (cyan), showing two angles on the adenine and one of the nicotinamide. The nicotinamide moiety is not resolved in the wild-type structure.

Discussion

It is often assumed in protein and metabolic engineering that enzymes have already been optimized toward their native functions and that their native catalytic efficiencies cannot be improved.[47] Despite evidence that metabolically crucial enzymes display, on average, faster kinetics than those involved in secondary metabolism and the corresponding prediction that the kinetics of secondary metabolic enzymes could be improved,[2] the engineering of more active enzymes has remained elusive.[3] Even when modest successes have been described (e.g., as in Bastian *et al.*[48]), no method for finding activating distal mutations (other than random mutagenesis) has been reported. In this study, we have empirically identified structural positions across a broad category of enzymes where mutations can improve kinetics of native or native-like reactions. Our results with ScADH6 and EcFucO also indicate improved overall functionality in an *in vivo* context where higher activity promotes better growth in the presence of toxic substrates.

We propose a few factors that may contribute to this unusual finding. The mutations identified in this study are remote from the catalytically active centers in the proteins (the average distance between the C^{α} of residues targeted in this study and the hydride-carrying C4N atom of the cofactor is 16.9 Å), lowering the chance of disrupting the active sites. Nevertheless, the extended NAD(P) cofactor can transmit perturbations to the active site and may even magnify them.[49,50] This is particularly relevant given the role of adenosine as a common energetic and recognition 'handle' in the binding of enzyme cofactors.[51,52] Furthermore, despite (or perhaps because of) its ubiquity, structurally diverse binding pockets for adenine have evolved.[51-55] In contrast to moieties with more specialized, conserved binding motifs, adenine binding is governed by a 'fuzzy recognition template' consisting of hydrophobic residues above and below adenine rings and polar residues around its rim.[51,54] This might suggest that the adenine-binding pocket can tolerate mutations that fine-tune the kinetics or energetics of binding as long as these general structural elements are present. Although no pattern in the beneficial mutations described here is readily apparent, such as an increase or decrease in steric bulk, side chain polarity or conformational entropy, a number of structural factors can potentially be linked to the catalytic efficiency of enzymes. We offer two possible mechanisms for the enhancements observed and discuss them below in the specific context of nicotinamide cofactor binding.

In a 1997 study of isocitrate dehydrogenase, Mesecar *et al*. observed that subtle chemical modification of the adenine ring of NADP significantly reduced catalytic activity. High-resolution crystal structures showed how the change to the adenine binding resulted in a slight shift in the position of the nicotinamide with respect to the substrate, and the large decrease in catalytic activity was attributed to this subtle misalignment.[49] In a study of aldo–keto reductases, Campbell *et al.* similarly concluded that the relative positioning of the cofactor and substrate in the active site was the major factor contributing to efficient turnover.[56] Although the structure of wild-type EcFucO does not have the nicotinamide resolved and the structure of EcFucO^{M185C} lacks the substrate, it is possible that the beneficial mutations identified here realign the cofactor in a way that improves catalytic preorganization with respect to the substrate and active site.

Alternately, the mutations may affect the binding and unbinding kinetics of the cofactor. In a comprehensive study of adenine-binding pockets, Nobeli *et al*. found that protein-bound adenine moieties form, on average, only 67% of theoretically possible hydrogen bonds,[54] suggesting that modulating the binding energy to fine-tune k_{on} and k_{off} is more important than achieving the tightest possible binding. Furthermore, for several kinds of NAD(P)-dependent enzymes, it has been shown that long-range conformational changes occur during binding of cofactor and/or substrate, resulting in allosteric cooperativity,[27,45,56,57] which could explain why, in some of the enzymes tested, mutations affect the substrate K_M, even though they are far from the substrate-binding portion of the enzyme.

Bar-Even *et al*. postulate in a recent paper that the sub-optimality of moderately efficient enzymes (enzymes whose second-order rate constants lie well below the diffusion limit) reflects, at least in part, a high proportion of 'futile encounters' between enzyme and substrate before a productive complex forms.[58] In this context, the mutations we observed could increase the likelihood of formation of productive enzyme–substrate complexes either by improving the energetics of proper cofactor alignment or by increasing the rate at which futile complexes dissociate. The observation that enhancements in all three kinetic parameters (k_{cat}, cofactor K_M and substrate K_M) arise from mutations distal to the site of catalysis or substrate binding may seem counter-intuitive. In this context, it is useful to remember that k_{cat} and K_M values are indirect 'black-box' measures of the formation and

dissociation of a catalytically productive enzyme–cofactor–substrate complex.[58] Therefore, cofactor binding that is better suited to the formation of a productive complex can be manifested in the turnover rate (k_{cat}) or may promote substrate binding that is more likely to be catalytically productive. Measurement of the microscopic or single-turnover kinetic parameters governing these reactions may be able to de-convolute these effects and shed more light on how the mutations described here promote activity.

In three out of ten enzymes tested thus far (LlAdhA, KpDHAT, and TeXR), no single mutation at these positions created variants with improved kinetics, indicating that these enzymes already lie at a local fitness optimum with respect to the targeted residues. No clear factor unites these three enzymes as distinct from the others studied, suggesting that the optimality of these positions is determined stochastically by the balance of genetic drift and natural selection.[59,60]

In this study, we have empirically identified structural sites that have a strong effect on activity without themselves being catalytically crucial and demonstrated that we can find mutations that boost the catalytic efficiency of the enzyme through subtle structural and/or energetic changes that cannot be rationally designed or predicted. We propose targeting mutations near the adenine for site-saturation mutagenesis and screening for improved kinetics as a fast and simple way to improve or tune the catalytic properties of NAD(P)-dependent enzymes. With a demonstrated success rate of seven in ten, including the two previously published KARIs, this is the first general strategy that has been proposed for improving a broad category of enzymes for their natural functions. Furthermore, different kinetic properties of the enzymes were changed, including substrate and cofactor affinity as well as turnover rate, allowing for fine-tuning of enzymatic properties for specific applications. Finally, these results demonstrate that many enzymes have room for improvement of catalytic properties *in vitro* and likely also *in vivo*, which holds promise for the engineering of improved biocatalysts and metabolic pathways.

References

1. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology* **5**, 593-599
2. Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402-4410
3. Bar-Even, A., and Tawfik, D. S. (2013) Engineering specialized metabolic pathways — Is there a room for enzyme improvements? *Current Opinion in Biotechnology* **24**, 310-319
4. Tcherkez, G. G., Farquhar, G. D., and Andrews, T. J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* **103**, 7246-7251
5. Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010) Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proceedings of the National Academy of Sciences* **107**, 3475-3480
6. Schmidt-Dannert, C., and Arnold, F. H. (1999) Directed evolution of industrial enzymes. *Trends in Biotechnology* **17**, 135-136
7. Bornscheuer, U. T., and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design. *Current Opinion in Chemical Biology* **5**, 137-143
8. Brustad, E. M., and Arnold, F. H. (2011) Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology* **15**, 201-210
9. Li, Y., and Cirino, P. C. (2014) Recent advances in engineering proteins for biocatalysis. *Biotechnology and Bioengineering* **111**, 1273-1287
10. Hurley, J. H., Chen, R. D., and Dean, A. M. (1996) Determinants of cofactor specificity in isocitrate dehydrogenase: Structure of an engineered $NADP^+ \rightarrow NAD^+$ specificity-reversal mutant. *Biochemistry* **35**, 5670-5678
11. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic Engineering* **13**, 345-352
12. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951
13. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* **18**, 2125-2138

14. Reiße, S., Garbe, D., and Brück, T. (2015) Identification and optimization of a novel thermo- and solvent stable ketol-acid reductoisomerase for cell free isobutanol biosynthesis. *Biochimie* **108**, 76-84
15. Valencia, E., Larroy, C., Ochoa, W. F., Parés, X., Fita, I., and Biosca, J. A. (2004) Apo and holo structures of an NADP(H)-dependent cinnamyl alcohol dehydrogenase from *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **341**, 1049-1062
16. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345
17. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods in Enzymology* **154**, 367-382
18. Sambrook, J., Frisch, E. F., Maniatis, T. (1989) *Molecular Cloning: A laboratory manual.*, Cold Spring Harbor Laboratory Press, New York
19. Patrick, W. M., Firth, A. E., and Blackburn, J. M. (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Engineering* **16**, 451-457
20. Bosley, A. D., and Ostermeier, M. (2005) Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomolecular Engineering* **22**, 57-61
21. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D - Biological Crystallography* **66**, 125-132
22. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D - Biological Crystallography* **62**, 72-82
23. Kumaran, D., and Swaminathan, S. (2009) Crystal structure of lactaldehyde reductase. RCSB Protein Data Bank
24. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D - Biological Crystallography* **60**, 2126-2132
25. Larroy, C., Fernández, M. R., González, E., Parés, X., and Biosca, J. A. (2002) Characterization of the *Saccharomyces cerevisiae* YMR318C (ADH6) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: Relevance in aldehyde reduction. *Biochemical Journal* **361**, 163-172
26. Hoover, G. J., Van Cauwenberghe, O. R., Breitkreuz, K. E., Clark, S. M., Merrill, A. R., and Shelp, B. J. (2007) Characteristics of an *Arabidopsis* glyoxylate reductase: General biochemical properties and substrate specificity for the recombinant protein, and developmental expression and implications for glyoxylate and succinic semialdehyde metabolism in *planta*. *Canadian Journal of Botany* **85**, 883-895
27. Liu, X., Bastian, S., Snow, C. D., Brustad, E. M., Saleski, T. E., Xu, J.-H., Meinhold, P., and Arnold, F. H. (2012) Structure-guided engineering of

Lactococcus lactis alcohol dehydrogenase LlAdhA for improved conversion of isobutyraldehyde to isobutanol. *Journal of Biotechnology* **164**, 188-195

28. Marcal, D., Rego, A. T., Carrondo, M. A., and Enguita, F. J. (2009) 1,3-Propanediol dehydrogenase from *Klebsiella pneumoniae*: decameric quaternary structure and possible subunit cooperativity. *Journal of Bacteriology* **191**, 1143-1151
29. Cocks, G. T., Aguilar, T., and Lin, E. C. (1974) Evolution of L-1,2-propanediol catabolism in *Escherichia coli* by recruitment of enzymes for L-fucose and L-lactate metabolism. *Journal of Bacteriology* **118**, 83-88
30. Chen, Y. M., Lu, Z., and Lin, E. C. (1989) Constitutive activation of the fucAO operon and silencing of the divergently transcribed fucPIK operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. *Journal of Bacteriology* **171**, 6097-6105
31. Wang, X., Miller, E. N., Yomano, L. P., Zhang, X., Shanmugam, K. T., and Ingram, L. O. (2011) Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. *Applied and Environmental Microbiology* **77**, 5132-5140
32. Fernandes, S., Tuohy, M. G., and Murray, P. G. (2009) Xylose reductase from the thermophilic fungus *Talaromyces emersonii*: Cloning and heterologous expression of the native gene (Texr) and a double mutant ($Texr^{K271R+N273D}$) with altered coenzyme specificity. *Journal of Biosciences* **34**, 881-890
33. Jansch, A., Freiding, S., Behr, J., and Vogel, R. F. (2011) Contribution of the NADH-oxidase (Nox) to the aerobic life of *Lactobacillus sanfranciscensis* DSM20451T. *Food Microbiology* **28**, 29-37
34. Petschacher, B., Staunig, N., Müller, M., Schürmann, M., Mink, D., De Wildeman, S., Gruber, K., and Glieder, A. (2014) Cofactor specificity engineering of *Streptococcus mutans* NADH oxidase 2 for $NAD(P)^+$ regeneration in biocatalytic oxidations. *Computational and Structural Biotechnology Journal* **9**
35. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W., and Arnold, F. H. (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochemical Journal* **468**, 475-484
36. Arnold K.; Bordoli L.; Kopp J.; Schwede, T. (2006) The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics*, 195 - 201
37. Bloom, J. D., Arnold, F. H., and Wilke, C. O. (2007) Breaking proteins with mutations: Threads and thresholds in evolution. *Molecular Systems Biology* **3**
38. Gong, L. I., Suchard, M. A., and Bloom, J. D. (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* **2**

39. Fasan, R., Meharenna, Y. T., Snow, C. D., Poulos, T. L., and Arnold, F. H. (2008) Evolutionary history of a specialized P450 propane monooxygenase. *Journal of Molecular Biology* **383**, 1069-1080
40. Wong, S.-H., Lonhienne, T. G. A., Winzor, D. J., Schenk, G., and Guddat, L. W. (2012) Bacterial and plant ketol-acid reductoisomerases have different mechanisms of induced fit during the catalytic cycle. *Journal of Molecular Biology* **424**, 168-179
41. Zhang, Y., and Garavito, R. M. (2011) Crystal structure of gamma-hydroxybutyrate dehydrogenase from *Geobacter metallireducens* in complex with $NADP^{+.}$ RCSB Protein Data Bank
42. Thomas, L. M., Harper, A. R., Miner, W. A., Ajufo, H. O., Branscum, K. M., Kao, L., and Sims, P. A. (2013) Structure of *Escherichia coli* AdhP (ethanol-inducible dehydrogenase) with bound NAD. *Acta Crystallographica Section F – Structural Biology Communications* **69**, 730-732
43. Benach, J., Winberg, J. O., Svendsen, J. S., Atrian, S., Gonzalez-Duarte, R., and Ladenstein, R. (2005) *Drosophila* alcohol dehydrogenase: Acetate-enzyme interactions and novel insights into the effects of electrostatics on catalysis. *Journal of Molecular Biology* **345**, 579-598
44. Moon, J.-H., Lee, H.-J., Park, S.-Y., Song, J.-M., Park, M.-Y., Park, H.-M., Sun, J., Park, J.-H., Kim, B. Y., and Kim, J.-S. (2011) Structures of iron-dependent alcohol dehydrogenase 2 from *Zymomonas mobilis* ZM4 with and without NAD^+ cofactor. *Journal of Molecular Biology* **407**, 413-424
45. Kavanagh, K. L., Klimacek, M., Nidetzky, B., and Wilson, D. K. (2002) The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry* **41**, 8785-8795
46. Lountos, G. T., Jiang, R., Wellborn, W. B., Thaler, T. L., Bommarius, A. S., and Orville, A. M. (2006) The crystal structure of NAD(P)H oxidase from *Lactobacillus sanfranciscensis*: insights into the conversion of O_2 into two water molecules by the flavoenzyme. *Biochemistry* **45**, 9648-9659
47. Tokuriki, N., Jackson, C. J., Afriat-Jurnou, L., Wyganowski, K. T., Tang, R., and Tawfik, D. S. (2012) Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications* **3**, 1257
48. Bastian, S., Rekowski, M. J., Witte, K., Heckmann-Pohl, D. M., and Giffhorn, F. (2005) Engineering of pyranose 2-oxidase from *Peniophora gigantea* towards improved thermostability and catalytic efficiency. *Applied Microbiolial Biotechnology* **67**, 654-663
49. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) Orbital steering in the catalytic power of enzymes: Small structural changes with large catalytic consequences. *Science* **277**, 202-206
50. 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), 251-258

51. Moodie, S. L., Mitchell, J. B., and Thornton, J. M. (1996) Protein recognition of adenylate: An example of a fuzzy recognition template. *Journal of Molecular Biology* **263**, 486-500
52. Denessiouk, K. A., Rantanen, V. V., and Johnson, M. S. (2001) Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* **44**, 282-291
53. Chakrabarti, P., and Samanta, U. (1995) CH/π interaction in the packing of the adenine ring in protein structures. *Journal of Molecular Biology* **251**, 9-14
54. Nobeli, I., Laskowski, R. A., Valdar, W. S. J., and Thornton, J. M. (2001) On the molecular discrimination between adenine and guanine by proteins. *Nucleic Acids Research* **29**, 4294-4309
55. Pyrkov, T. V., Kosinsky, Y. A., Arseniev, A. S., Priestle, J. P., Jacoby, E., and Efremov, R. G. (2007) Complementarity of hydrophobic properties in ATP-protein binding: A new criterion to rank docking solutions. *Proteins* **66**, 388-398
56. Campbell, E., Chuang, S., and Banta, S. (2013) Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily. *Protein Engineering Design and Selection* **26**, 181-186
57. Plapp, B. V. (2010) Conformational changes and catalysis by alcohol dehydrogenase. *Archives of Biochemistry and Biophysics* **493**, 3-12
58. Bar-Even, A., Milo, R., Noor, E., and Tawfik, D. S. (2015) The moderately efficient enzyme: Futile encounters and enzyme floppiness. *Biochemistry* **54**(32), 4959-4977
59. Lynch, M. (2012) Evolutionary layering and the limits to cellular perfection. *Proceedings of the National Academy of Sciences* **109**, 18851-18856
60. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., and Lynch, M. (2012) Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences* **109**, 18488-18492

Section III

Chapter 7: A general tool for engineering the nicotinamide cofactor specificity of enzymes

Chapter 8: Cofactor Specificity Reversal – Structural Analysis and Library Design (CSR-SALAD) users manual

Chapter 7

A GENERAL TOOL FOR ENGINEERING THE NICOTINAMIDE COFACTOR SPECIFICITY OF ENZYMES

Abstract

The need to reverse the cofactor specificity of NAD/NADP dependent enzymes is a recurrent challenge in synthetic biology. Despite more than sixty enzymes having been engineered to switch cofactor specificity, specificity reversal remains difficult to accomplish and no one method easily and reliably identifies a set of mutations sufficient to the task. Drawing inspiration from a comprehensive survey of previous studies and our own engineering successes, we have developed a structure-guided semi-rational strategy for engineering the reversal of cofactor specificity. The efficacy of this strategy has been demonstrated by reversing the specificity of four structurally diverse NADP-dependent enzymes (glyoxylate reductase, cinnamyl alcohol dehydrogenase, xylose reductase, and iron-containing alcohol dehydrogenase). Finally, we have also implemented this approach as an easy-to-use web tool, Cofactor Specificity Reversal – Structural Analysis and LibrAry Design (CSR-SALAD), to enable use by a broad user base.

Introduction

The prospect of a transition from petroleum-based chemistry to bio-based manufacturing is one that is gaining traction. The USDA projects the market share for industrial production of bio-based specialty chemicals and biofuels to grow by more than 10-fold by the end of the next decade.[1] However, engineering highly efficient biosynthetic pathways and processes remains a challenging multi-factorial task, often requiring optimization of the reaction conditions, host strains, pathway expression levels, and the enzymes themselves.[2-4] One common hurdle in the development of efficient pathways is the need to balance the production and consumption of the hydride transport cofactors nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP).[5] Several studies have shown how balancing cofactor availability can increase pathway yields by removing carbon inefficiencies and side products, eliminating oxygen requirements, or improving equilibrium metabolite levels.[6-12] Although metabolic engineering approaches have been used to increase the availability of nicotinamide redox equivalents, the best results have come from the direct engineering of enzymes to alter their

specificities so that a given cofactor produced in one or more pathway steps is consumed in others.[6,10]

Since the first report of engineered specificity reversal in 1990 by Scrutton and co-workers, wherein an NADP-preferring glutathione reductase was engineered to use NAD by the introduction of seven residues from an NAD-preferring homologue,[13] altering nicotinamide cofactor specificity has frequently been a target of protein engineering both for industrial applications as well as to address the fundamental question of how sequence and structure contribute to cofactor specificity. Tables 1-1 and 1-2 contain a list of more than 25 examples of cofactor switching in each direction (NADP-to-NAD and *vice versa*), updated from that of Khoury *et al*.[14] However, as with the field of protein engineering more generally, the engineering of these proteins has occurred on an individual, piecemeal basis, with no single approach having proven consistently effective. Furthermore, such engineering has remained the province of experts, rather than becoming a tool for end-users such as metabolic engineers due to a requirement for specialized knowledge or software, or the equipment for high-throughput screening. Cofactor specificity reversal is a particularly challenging protein engineering task because the phosphate that distinguishes NAD from NADP possesses significant steric and electrostatic bulk. As a result, multiple simultaneous mutations have almost always been required for effective cofactor switching due to strong non-additivities in the effects of mutations.[9,15,16] This renders ineffective most of the common tools in the directed evolution toolkit, such as random mutagenesis by error-prone PCR or iterative single site-saturation mutagenesis, both of which operate on an uphill-walk model for optimization.

In Chapter 1, we developed a simple recipe for the cofactor specificity reversal of any member of the ketol-acid reductoisomerase (KARI) enzyme family.[17] However, the diverse structural motifs that give rise to cofactor specificity[18] prevent the direct extension of this recipe to the multitude of other NAD(P)-dependent enzyme families. In this chapter we provide a general approach that functions independent of protein fold and family. Additionally, we have designed this approach towards sufficient simplicity to make it accessible to non-specialist researchers. This semi-rational approach selects codons containing mixed nucleotides (degenerate codons) at structurally-identified positions so as to focus screening on mutations found to be most beneficial in previous studies. As such, it

balances the desire to keep library sizes small with the practical reality that the complex and dynamic nature of a binding pocket[19-22] and the sensitivity to small changes in binding energy or orientation[23,24] severely limit our ability to rationally predict functional sequences.

To accompany this engineering framework, we have developed an easy-to-use online tool, CSR-SALAD (Cofactor Specificity Reversal – Structural Analysis and LibrAry Design; Figure 7-1ab) which analyzes NAD(P)-bound structures and suggests these small, semi-rational libraries. We demonstrate the efficacy of CSR-SALAD by reversing the cofactor specificity of four NADP-dependent enzymes from three different fold families.

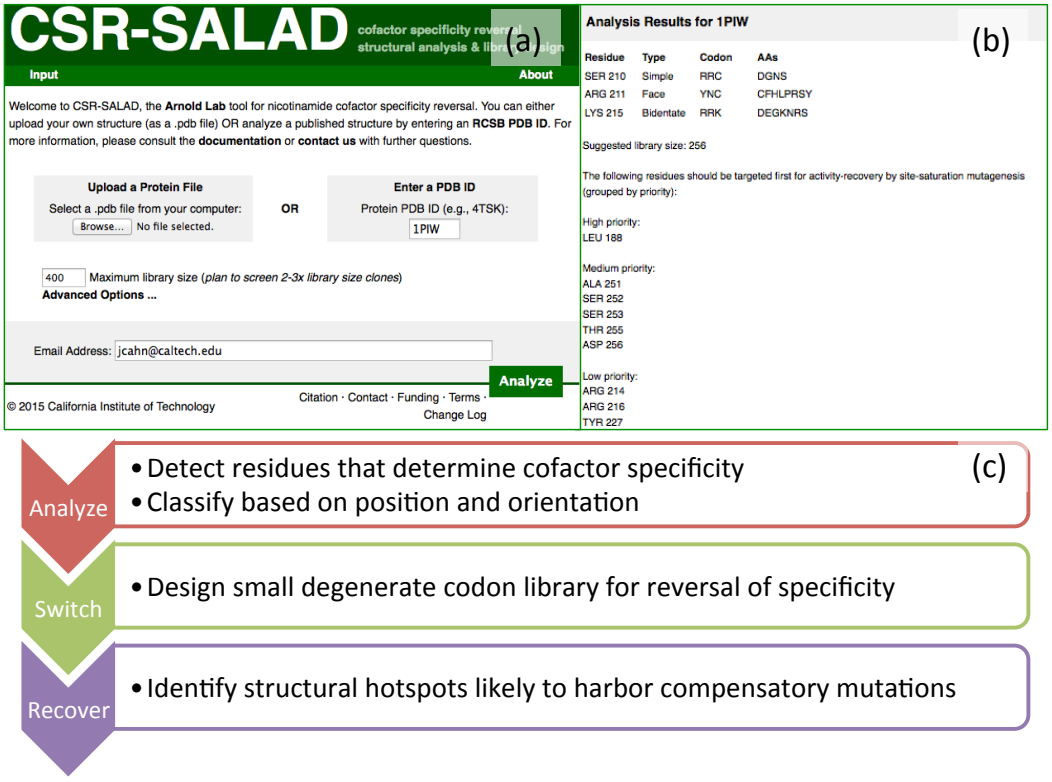


Figure 7-1. The web interface of CSR-SALAD (a), and an example of the output (b). CSR-SALAD performs three tasks (c): structure analysis, design of cofactor-switching libraries, and identification of positions for activity recovery.

Approach

In Chapter 1, we described a three-step process for reversing the cofactor specificity of KARIs: (a) identification of the residues involved in determining cofactor specificity, (b) creation of a small library by combinatorial mutation of these positions and screening for reversed cofactor preference, and (c) recovery of catalytic activity by mutagenesis outside the cofactor specificity-determining region (Figure 1-2).[17] As shown in Figure 7-1c, the approach described herein and implemented in the CSR-SALAD web tool first generalizes these steps to a variety of cofactor-binding folds and both nicotinamide cofactors, and then uses the lessons of the accumulated literature on cofactor specificity to reduce each step to experimentally tractable levels. Computational formalization of these shortcuts serves to provide ease-of-use and reproducibility. The precise details of these formalizations can be found in the CSR-SALAD documentation (Chapter 8); here we focus on the lessons that guide them.

Structural analysis

As might be intuitively expected, nearly all of the experimentally-determined specificity-conferring residues lie in the immediate vicinity of the 2' moiety of the NAD/NADP cofactor (the hydroxyl or phosphate; see Figure 1-1). While residues outside this vicinity have occasionally been mutated in cofactor-switched proteins, they have only rarely been shown to directly contribute to the reversal of specificity. In CSR-SALAD, the specificity-determining residues are therefore defined as those that contact the 2' moiety directly, are in position to contact it through water-mediated interactions, or – specifically for NAD-to-NADP switching – can be mutated to contact the expanded 2' moiety of the NADP cofactor.

To assist with engineering, it is useful to obtain more information about a residue than simply whether or not it contacts the cofactor. Therefore, we have developed a system of residue classifications that describe the role of each residue in forming the cofactor-binding pocket. This classification scheme was informed by that introduced in 1997 by Carugo and Argos.[18] Like those classifications, it was developed using a structural alignment of NAD(P)-bound proteins – in this case, nearly 1,000 proteins instead of 32 – and an analysis of the properties of the contacting residues. These classifications are

assigned on the basis of which atoms of the cofactor a residue contacts and its position with respect to the adenine moiety (examples of the six classifications are shown in Figure 7-2). Structural classifications are important because the useful mutations to a ‘face’ histidine, for instance, might differ considerably from those made to an ‘edge’ or ‘bidentate’ histidine.

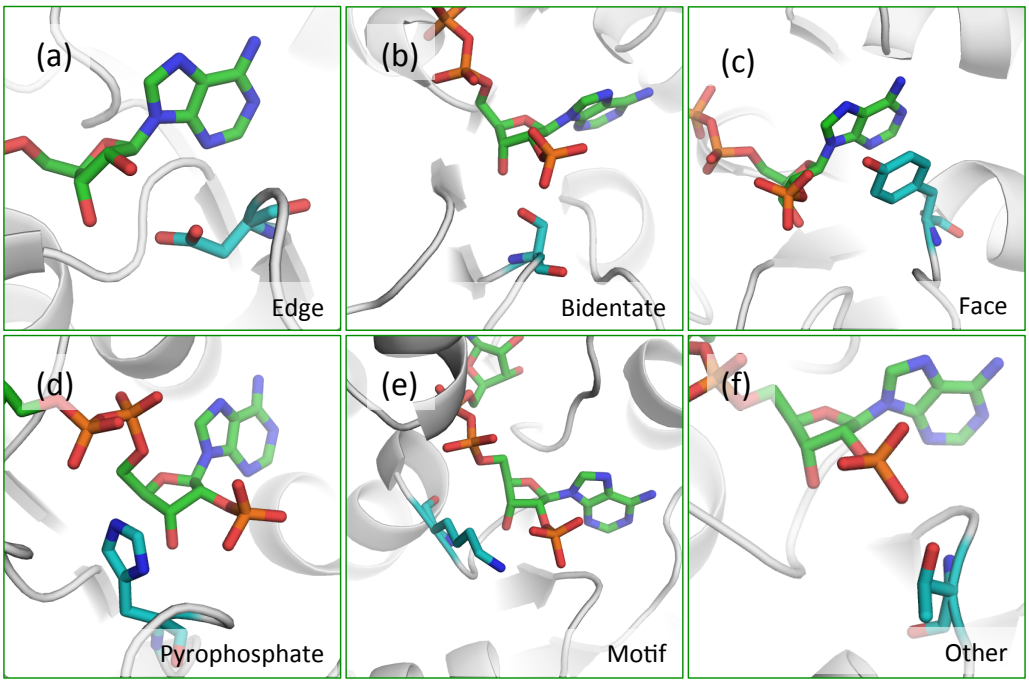


Figure 7-2. Examples of the six structural classifications used by CSR-SALAD. For more details, see the Table 8-2. Figures are from structures 1VC2, 1AMO, 1AMO, 1EZ0, 1CYD, and 1CYD, respectively.

Library design

To keep library sizes small and easy to construct experimentally, CSR-SALAD uses sub-saturation degenerate codon libraries.[25-27] This technique specifies mixtures of nucleotides (see Table 8-1)[28] to generate specific mixtures of amino acids at each targeted position. Previous studies have selected degenerate codons on the basis of natural sequence variability or desired properties (i.e., NYC for hydrophobic amino acids or RRK for charged amino acids);[27] selection of codons for cofactor switching in CSR-SALAD was guided by three considerations:

a) Inclusion of mutations to structurally similar residues that have previously been shown to be useful for cofactor specificity reversal in the desired direction

b) Inclusion of mutations to structurally homologous residues in proteins which naturally bind the desired cofactor

c) Inclusion of mutations based on structural and engineering intuition about binding the desired cofactor or achieving reversed selectivity.

To allow for library sizes to be tailored to the experimental throughput capabilities of the user, CSR-SALAD possesses a range of degenerate codons of different sizes for each residue in each structural class. The precise prioritization of residues within these codons, and the prioritizations of which positions are given the larger codons when library size limitations come into play, are subjective and based on an analysis of the engineering successes of prior studies by ourselves and others. We hope the application of the software will result in feedback that will further improve the algorithm as the available data increases.

Activity recovery

Cofactor-switched enzymes – or any enzymes with multiple mutations away from wild type – often suffer a significant loss of activity,[17,29,30] and compensatory mutations must be found to recover this activity. These mutations re-stabilize or re-activate the protein with the new cofactor, and often lie remote from the cofactor switching mutations. In our

previous work on KARIs in Chapter 1, we recommended using random mutagenesis methods such as error-prone PCR and screening to discover these mutations. While highly effective, this may be beyond the capabilities of non-specialist labs. For CSR-SALAD, we have taken the unprecedented step of using structural information to predict residues with increased probabilities of harboring compensatory mutations. This allows single-site saturation libraries (generally 3–5 of them) to be rapidly screened and the best mutations combined to produce active enzymes. The activity-recovery positions fall into a number categories suggested by common features of previous engineering efforts, as outlined in the CSR-SALAD documentation (Chapter 8). In our hands the most effective have consistently been mutations around the adenine ring, and all of the mutations used in the experimental validation below fall into this category. In Chapter 6 we demonstrated the power of these mutations, in the absence of any other mutations, to boost the activities of native-like reactions for a number of NAD(P)-dependent enzymes.[23] In cofactor-switched enzymes, the effects are even more dramatic, as we discovered by random mutagenesis and screening in several of the KARIs as described in Chapter 1.[17,31]

Materials and Methods

CSR-SALAD development

CSR-SALAD was built in Python, relying heavily on the PDB module of the Biopython package.[32] The analysis component was tested on a representative set of 499 NAD-bound structures and 463 NADP-bound structures selected on the basis of sequence identity and resolution. Degenerate codon selection was assisted by the LibDesign[26] and AA-Calculator[33] tools and optimized using the Ambiguous Nucleotide Tool (ANT) framework.[27] CSR-SALAD has been implemented as a PHP-based web-server that is available at http://www.che.caltech.edu/groups/fha/CSRSALAD/index.html. Further documentation of CSR-SALAD's internal workings and guidelines for use can be found at that website or in the Chapter 8.

Structure analysis

Homology models were created using the SWISS-MODEL server[34]. For proteins crystallized without bound cofactor, cofactor was manually placed by alignment with a homologous protein, followed by manual side-chain rotamer adjustments in PyMol.[35]

Cloning and library creation

Genes were either obtained from Integrated DNA Technologies (IDT) as gBlock linear fragments or cloned from pre-existing vectors, and were cloned into pET22b(+) in frame with the C-terminal His_6-tag for expression in *Escherichia coli* BL21(DE3) 'E. Cloni' (Lucigen) using Gibson cloning[36] with overlap at the T7 promoter and terminator sequences. Mutagenic primers for site-saturation mutagenesis were obtained from IDT and treated as suggested by IDT protocols. Libraries were generated using a modified version of the QuikChange method (Stratagene) as described previously.[37] Some of the libraries screened were based on early versions of the CSR-SALAD algorithm, but all reported mutants are found in the CSR-SALAD libraries generated using default parameters.

Following transformation with library DNA, single colonies were picked with sterile toothpicks and inoculated into 300 µL of Luria broth supplemented with 100 µg/mL ampicillin (LB-Amp) in shallow-well 96-well plates. Following overnight growth at 37 °C with shaking at 225 rpm and 80% humidity, 50 µL of the precultures were added to 600 µL of fresh LB-Amp media in deep-well 96-well plates and grown for 3 h at 37 °C. Then, 50 µL of additional LB-Amp containing 0.25 mM isopropyl thiogalactopyranoside (IPTG) were added and expression was continued at a reduced temperature. For expression temperatures and times, see Table 7-1. The expression cultures were harvested through centrifugation, and the plates containing cell pellets were stored at -20 °C until screening.

Library screening

Assay procedures varied depending on the protein. The following is the general protocol, but deviations for specific proteins are listed in Table 7-1. *E. coli* cells were resuspended in the appropriate lysis buffer (Table 7-1) containing 750 mg/L lysozyme, 10 mg/L DNase I, and 2 mM $MgCl_2$. Lysis was accomplished at 37 °C for 1 h. Enzyme activities were then assayed by monitoring NAD(P)H consumption in the presence of the substrate molecule at 340 nm in a plate reader.

Table 7-1. Details for expression, lysis, and assay conditions for enzymes used for validation of CSR-SALAD.

Enzyme	Expression Conditions	Lysis Conditions	Assay Conditions	Notes
***Arabidopsis thaliana* glyoxylate reductase**	Luria broth 24 °C, 18 h	100 mM hydroxyethyl-piperazineethanesulfonic acid (HEPES) pH 7.8	100 mM HEPES pH 7.8 1 mM sodium glyoxylate	
***Saccharomyces cerevisae* cinnamyl alcohol dehydrogenase**	Luria broth 20 °C, 18 h	33 mM sodium phosphate pH 7 10% 10x BugBuster	33 mM sodium phosphate pH 7 0.5 mM *trans*-cinnamaldehyde	
***Talaromyces emersonii* xylose reductase**	Luria broth 25 °C, 18 h	50 mM potassium phosphate pH 7	50 mM potassium phosphate pH 7 400 mM xylose	
***Thermotoga maritima* iron-containing alcohol dehydrogenase**	Luria broth 20 °C, 18 h	33 mM sodium phosphate pH 7 10% 10x BugBuster	33 mM sodium phosphate pH 7 0.1% (v/v) *n*-butanal (~10 mM)	Enzyme had no prior experimental characterization. Initial tests showed it was more active on *n*-butanal than on propanal or furfural.
***Thermobifida fusca* phenylacetone monooxygenase**	Luria broth 30 °C, 4 h	50 mM potassium phosphate pH 7	50 mM potassium phosphate pH 7 0.5 mM phenylacetaldehyde	Used Breathe Easier membranes (Diversified Biotech) instead of polystyrene lid when growing in 96-well plates. Preferred substrate (phenylacetone) is a Schedule II controlled substance, but PAMO is active with phenylacetaldehyde as well.[38]

Enzyme expression, purification, and kinetic measurements

For larger-scale expression, 5–50 mL pre-cultures were grown overnight and diluted 1:250 into fresh media for expression. After expression cultures reached an OD_{600} of 0.6–0.9, they were induced with IPTG to a final concentration of 0.5 mM. Expression then proceeded for the same times and temperatures used for library expression. Expression cultures were harvested by centrifugation, the supernatant was discarded, and the pellets were frozen at -20 °C.

For purification, cell pellets were resuspended in 10–20 mL buffer A (25 mM Tris, 100 mM NaCl, 20 mM imidazole, pH 7.4) and lysed by sonication or using BugBuster protein extraction reagent (EMD Millipore). The lysate was clarified by centrifugation, and the enzymes were purified via their C-terminal His_6-tag using High Performance (HP) Ni-NTA Sepharose columns (GE Healthcare, Waukesha, WI, USA) on an Äkta Xpress FPLC (GE Healthcare). The concentration of purified protein was determined using the Bradford assay (Bio-Rad, Hercules, CA, USA).

For rate measurements, k_{cat} values were determined using the same assay conditions as above with saturating cofactor and substrate. Michaelis–Menten constants were determined by varying cofactor concentration, and activity was monitored using fluorescence (excitation 340 nm, emission 440 nm) for improved sensitivity. At least six cofactor concentrations were used for these determinations, and all measurements were performed at least three times. MatLab (Mathworks, Natwick, MA, USA) was used for parameter fitting.

Results

Comparison to previous studies

Because this study relies heavily on the results of previous studies, and as a positive control, we first determined whether modestly-sized CSR-SALAD libraries could recapitulate the mutations discovered in previous experiments. Because there may be several sets of mutations that will reverse cofactor specificity for a given enzyme[9,39-41] and because not all results of prior studies have been highly successful, we do not expect a perfect match. Nevertheless, of the 12 previous NADP-to-NAD and 17 NAD-to-NADP studies that reported a reversal of specificity and a final activity greater than 10% of initial activity, CSR-SALAD libraries generated using default parameters contained the best individual mutants (not including distal compensatory mutations) for 7 (58%) and 11 (65%), respectively. (Table 7-2). Full data on mutant recapitulation are found in Tables S7-1 and S7-2.

Table 7-2. Recapitulation of mutants from previous successful nicotinamide cofactor specificity reversals. For a full breakdown of mutant recapitulation, see Tables S7-1 and S7-2.

Percent mutations recapitulated	0%	1-49%	50-99%	100%
NADP-to-NAD	0	1	4	7
NAD-to-NADP	1	4	1	11

Includes only mutants that had k_{cat}/K_M for the switched cofactor greater than the wild type's preferred cofactor and where mutant k_{cat}/K_M was at least 10% of wild type. This count excludes residues distal to the 2' binding pocket.

Experimental validation

To demonstrate the utility of CSR-SALAD for predicting cofactor-switching mutations, we selected a test set of five enzymes drawn from those previously studied in our lab and enzymes with industrial potential: *Arabidopsis thaliana* glyoxylate reductase (GR), *Saccharomyces cerevisae* cinnamyl alcohol dehydrogenase (CinADH), *Talaromyces emersonii* xylose reductase (XR), *Thermotoga maritima* iron-containing alcohol dehydrogenase (FeADH), and *Thermobifida fusca* phenylacetone monooxygenase (PAMO) (Table 7-3). Although CSR-SALAD is capable of processing NAD and NADP-bound structures, we focused only on NADP-preferring enzymes, because NADP-to-NAD switching is generally considered to be more industrially relevant[42-44] and has also proved to be more difficult to achieve in the past (Tables 1-1 and 1-2). The selected enzymes include the most common NAD(P)-binding fold, the Rossmann fold, as well as other folds less commonly used for nicotinamide cofactor binding (TIM barrel, dihydroquinoate synthase-like, and FAD/NAD-binding fold). To demonstrate the robustness of CSR-SALAD analysis, we also included one protein (*T. emersonii* XR) without a crystal structure, but for which the structure of a homologous enzyme with high sequence identity was known, and one enzyme (*A. thaliana* GR) that was crystallized in the absence of the cofactor. For each of these, we generated the library predicted using CSR-SALAD and screened it for cofactor-switched variants. The best of these variants were then subjected to between zero and two rounds of single-site saturation mutagenesis for activity recovery, and the best final variants were purified and characterized for Michaelis-Menten kinetics (Tables 7-3 and 7-4).

For the four oxidoreductases – *A. thaliana* GR, *S. cerevisae* CinADH, *T. emersonii* XR, and *T. maritima* FeADH – cofactor specificity was successfully reversed, and for three of those four the

catalytic activity was recovered back to significant levels (>10% of wild type). Indeed, these represent three of the most successful NADP-to-NAD cofactor specificity reversals ever reported (Table 1-1). For *T. emersonii* XR, which we were also able to switch, the activity recovery was less successful. We suspect that, should greater activity be required, further compensatory mutations could be found by traditional directed evolution methods.[17]

Notably, the sets of mutations found for each enzyme show no significant overlap or patterns. Indeed, the only mutation that shows up twice, serine-to-glycine in *T. emersonii* XR and *T. maritima* FeADH, occurs in residues that are in different structural classes. Further, while each set of mutations contains one or more mutations to a negatively-charged residue, these mutations come from residues with different structural classes in the wild-type binding of the cofactors (Figure 7-3), while the rest of the mutations introduce amino acids whose role in cofactor specificity is less immediately evident. This reinforces the assumptions underlying the CSR-SALAD approach: that the precise determinants of cofactor specificity are often difficult to predict but that active proteins can be selected from libraries containing reasonable sets of potential contacts.

For *T. fusca* phenylacetone monooxygenase (PAMO), a Baeyer-Villiger monooxygenase (BVMO), we were also able to produce a mutant enzyme that consumed NADH faster than NADPH. However, the cofactor consumption of the purified enzyme was found to be nearly completely uncoupled from product formation, with cofactor oxidation rates that were entirely independent of substrate (phenylacetaldehyde) concentration. Because this enzyme displays activity on a considerable range of substrates, including compounds naturally present in *E. coli* lysate, it was impossible to perform a negative screen against uncoupling, and none of the mutations at recovery positions restored coupling. While previous studies on the nicotinamide cofactor specificity of BVMOs have proposed that the 2'-phosphate may be essential for catalytic function,[45] the existence of naturally bispecific BVMOs rebuts this.[46-48] We instead propose that in enzymes such as BVMOs, with multi-step electron-transfer pathways, the precise alignment of the donor-acceptor pairs is much more sensitive to perturbation than with oxidoreductases, resulting in premature loss of the hydride to acceptors such as molecular oxygen. Similar results have been seen in attempts to engineer the specificity of P450 reductases.[49,50] Therefore, successfully switching the cofactor specificities of these enzymes may require the ability to more accurately predict the resultant cofactor geometries, or, more likely, the ability to screen much larger libraries while also monitoring uncoupling or product formation.

Table 7-3: NADP-preferring enzymes used for experimental validation in this study. UniProt IDs, PDB accession codes, and Structural Classification of Proteins (SCOP) classification of their cofactor binding folds are listed for each. Mutations in bold are those used for switching cofactor specificity, others are from recovery libraries. Fold changes from wild-type kinetics are provided, as is the log of the mutant's NADH k_{cat}/K_M divided by the wild-type NADPH k_{cat}/K_M. Specific values for wild-type and mutant enzymes are provided in Table 7-4.

					Fold change from wild type							
					NADH			NADPH			Specificity	log
Enzyme	UniProt	RCSB PDB	Fold (SCOP)	Mutations	k_{cat}	K_M	k_{cat}/K_M	k_{cat}	K_M	k_{cat}/K_M	(NADH/ NADPH)[c]	(NADH^{mut}/ NADPH^{WT})
Arabidopsis thaliana glyoxylate reductase	Q9LSV0	3DOJ[a]	Rossmann (c.2.1.6)	**R31L**, **T32K**, **K35D**, C68R	2.36	0.88	2.67	0.37	4.55	0.08	33	-0.71
Saccharomyces cerevisae cinnamyl alcohol dehydrogenase	Q04894	1PIW	Rossmann (c.2.1.1)	**S210D**, **R211P**, **K215E**, S253P	104.90	0.55	191.36	1.54	0.53	2.90	65	0.77
Talaromyces emersonii xylose reductase	C5J3R6	1K8C[b]	TIM Barrel (c.1.7.1)	**S272G**, **N273D**, **R277Y**, Q280E	13.45	0.44	30.67	0.02	2.48	0.01	4900	-1.48
Thermotoga maritima iron-containing alcohol dehydrogenase	Q9X022	1VHD	DHQS-like (e.22.1.2)	**G36E**, **S38N**, **S39G**	0.88	0.15	5.87	0.20	2.29	0.07	84	0.47
Thermobifida fusca phenylacetone monooxygenase	Q47PU3	2YLR	FAD/NAD pair (c.3.1.5)	**R217T**, **T218E**, **K336Y**	0.29[d]	0.52	0.56	0.00[d]	7.29	0.00	3600	-3.67

[a]Structure 3DOJ has no cofactor, so the cofactor was introduced from structure 3PEF (*G. metallireducens* γ-hydroxybutyrate dehydrogenase) on the basis of a backbone alignment.

[b]*T. emersonii* xylose reductase (XR) has not been crystallized, so a homology model was generated from 1K8C, the structure of an XR from *C. tenuis*.

[c]Specificity is the ratio of NADH k_{cat}/K_M divided by NADPH k_{cat}/K_M for each enzyme. Presented here is fold change.

[d]The mutant enzyme displayed >99% uncoupling for both cofactors.

Table 7-4: Kinetics of proteins used for the validation of CSR-SALAD.

	NADH			NADPH			k_{cat}/K_M Specificity	% Uncoupling	
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	NADH	NADPH
***A. thaliana* GR**	440 ± 50	76 ± 24	5.8 ± 1.9	1600 ± 230	20 ± 6	79.8 ± 26.6	0.07 ± 0.03	N/D[a]	N/D
***At*GR R31L, T32K, K35D, C68R**	1030 ± 190	67 ± 15	15.4 ± 4.5	590 ± 140	91 ± 21	6.5 ± 2.1	2.36 ± 1.04	N/D	N/D
***S. cerevisae* CinADH**	180 ± 70	55 ± 23	3.4 ± 2.0	12350 ± 50	113 ± 46	109.1 ± 43.9	0.03 ± 0.02	N/D	N/D
***Sc*CinADH S210D, R211P K215E, S253P**	19400 ± 7850	30 ± 23	648.7 ± 512.6	18970 ± 1670	60 ± 43	316.9 ± 231.5	2.05 ± 2.20	N/D	N/D
***T. emersonii* XR**	8 ± 1	259 ± 18	0.03 ± 0.003	2540 ± 220	85 ± 73	29.7 ± 25.5	0.001 ± 0.0009	N/D	N/D
***Te*XR S272G, N273D, R277Y, Q280E**	110 ± 40	114 ± 76	0.97 ± 0.74	40 ± 4	212 ± 19	0.2 ± 0.03	5.23 ± 4.04	N/D	N/D
***T. maritima* FeADH**	1310 ± 140	749 ± 403	1.75 ± 0.96	1390 ± 270	402 ± 90	3.46 ± 1.02	0.51 ± 0.32	N/D	N/D
***Tm*FeADH G36E, S38N, S39G**	1160 ± 10	112 ± 112	10.29 ± 10.26	280 ± 80	1175 ± 1538	0.24 ± 0.32	42.70 ± 71.26	N/D	N/D
***T. fusca* PAMO**	30 ± 6	182 ± 68	0.16 ± 0.07	10520 ± 530	26 ± 19	405.14 ± 294.74	0.0004 ± 0.0003	62%	2%
***Tf*PAMO R217T, T218E, K336Y**	8 ± 1	94 ± 33	0.08 ± 0.03	10 ± 3	189 ± 66	0.06 ± 0.03	1.38 ± 0.77	>99%	>99%

[a]N/D = not detected

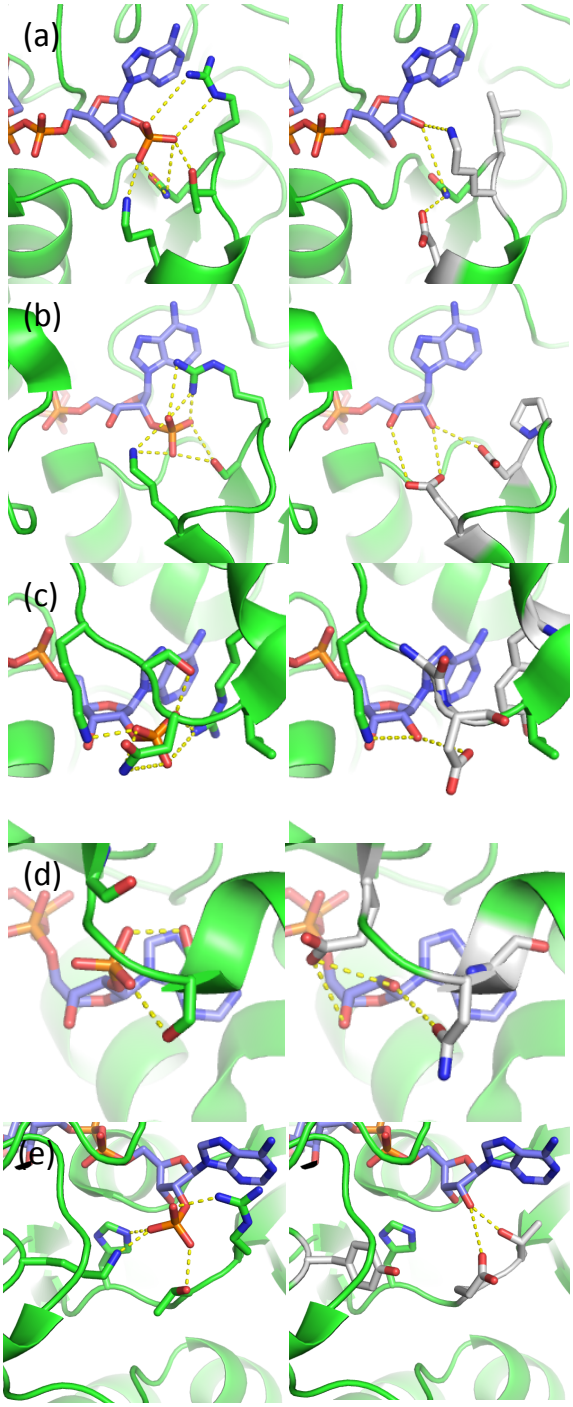


Figure 7-3. Structures of wild type proteins (or homology models) (left) and models of the mutations made (right) for *A. thaliana* GR (a), *S. cerevisae* CinADH (b), *T. emersonii* XR (c), *T. maritima* FeADH (d), and *T. fusca* PAMO (e). Mutations are colored with white carbons, and were introduced in PyMol without any structural minimization or optimization. Selected hydrophilic interactions are shown with yellow dashed lines.

Discussion

Semi-rational protein engineering, whether using degenerate codon libraries or other mutagenesis strategies, has become commonplace as a way to reap the advantages of both rational and evolutionary protein engineering.[2,51-54] In this study, we have created a tool to guide the creation of semi-rational libraries for a common protein engineering task, reversal of nicotinamide cofactor specificity, in an attempt to render it reliable and accessible to a broad base of potential users.

Nicotinamide cofactor specificity engineering provides a number of distinct advantages for an approach of this kind. First, despite the important role of the adenosine end of the cofactor in providing a recognition element and binding anchor for the nicotinamide moiety, there is considerable structural diversity among NAD(P)-binding proteins, both in their overall fold and in the specific motif for binding to the 2'-phosphate of NADP or the 2' hydroxyl of NAD.[18,55] In the past, this structural diversity has been a barrier to transferring cofactor engineering solutions between proteins,[17] but it also suggests that protein engineers do not need to replicate a specific NAD- or NADP-binding mode, and instead that simply providing the right potential contacts in reasonable geometries to allow binding can be sufficient. This hypothesis is central to the semi-rational approach developed in this paper, which makes no predictions about specific binding pocket geometries but instead creates focused diversity that matches the diverse binding modes of known natural and engineered NAD(P)-binding proteins.

Another feature of the cofactor specificity problem that makes it amenable to this approach is that the distinct spectral signature of reduced NAD(P) allows for rapid medium-to-high-throughput screening of libraries in lysate without the need for expensive equipment. Finally, the wealth of previous studies on the structural determinants of cofactor specificity and the numerous previous attempts to reverse specificity have generated a large data set from which we can seek inspiration. Though we validated CSR-SALAD only on NADP-to-NAD switching, the same advantages that apply equally to NAD-to-NADP switching, and CSR-SALAD has been developed to design libraries for this task based on the same criteria.

As protein engineering grows as a discipline, methods will need to be developed that balance reliability with speed and ease-of-use. In current practice, each protein to be engineered is a new problem that must be approached with careful consideration of strategy and methods. With CSR-SALAD, we have developed a framework for reliably performing a common and challenging protein

engineering task with modest experimental effort and with a minimum of required expertise. It remains to be seen whether similar structure-based semi-rational approaches can be developed for other protein engineering targets. Such targets would need to be localizable to a single region or class of residues on the protein, have sufficient prior literature to draw lessons, and have available high-throughput screening methodologies. Nevertheless, we foresee that in the meantime CSR-SALAD will be of great utility for the fields of synthetic biology, metabolic engineering, and biocatalysts, and will make the reversal of cofactor specificity a routine task rather than a formidable engineering endeavor.

References

1. USDA. (2008) Bio-based products: Market potential and projections through 2025. US Department of Agriculture
2. Eriksen, D. T., Lian, J., and Zhao, H. (2014) Protein design for pathway engineering. *Journal of Structural Biology* **185**, 234-242
3. Willrodt, C., Karande, R., Schmid, A., and Julsing, M. K. (2015) Guiding efficient microbial synthesis of non-natural chemicals by physicochemical properties of reactants. *Current Opinion in Biotechnology* **35**, 52-62
4. Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and Lee, S. Y. (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology* **8**, 536-546
5. Wang, Y., San, K.-Y., and Bennett, G. N. (2013) Cofactor engineering for advancing chemical biotechnology. *Current Opinion in Biotechnology* **24**, 994-999
6. Matsushika, A., Watanabe, S., Kodaki, T., Makino, K., Inoue, H., Murakami, K., Takimura, O., and Sawayama, S. (2008) Expression of protein engineered $NADP^+$-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* **81**, 243-255
7. Bengtsson, O., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. F. (2009) Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Biotechnology for Biofuels* **2**
8. Hasegawa, S., Uematsu, K., Natsuma, Y., Suda, M., Hiraga, K., Jojima, T., Inui, M., and Yukawa, H. (2012) Improvement of the redox balance Increases L-valine production by *Corynebacterium glutamicum* under oxygen deprivation conditions. *Applied and Environmental Microbiology* **78**, 865-875
9. Bastian, S., Liu, X., Meyerowitz, J. T., Snow, C. D., Chen, M. M. Y., and Arnold, F. H. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic Engineering* **13**, 345-352
10. Tamakawa, H., Ikushima, S., and Yoshida, S. (2011) Ethanol production from xylose by a recombinant *Candida utilis* strain expressing protein-engineered xylose reductase and xylitol dehydrogenase. *Bioscience, Biotechnology, and Biochemistry* **75**, 1994-2000
11. Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Applied and Environmental Microbiology* **77**, 2905-2915
12. Xiao, Z., Lv, C., Gao, C., Qin, J., Ma, C., Liu, Z., Liu, P., Li, L., and Xu, P. (2010) A novel whole-cell biocatalyst with NAD^+ regeneration for production of chiral chemicals. *PLoS ONE* **5**
13. Scrutton, N. S., Berry, A., and Perham, R. N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**, 38-43
14. Khoury, G. A., Fazelinia, H., Chin, J. W., Pantazes, R. J., Cirino, P. C., and Maranas, C. D. (2009) Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Science* **18**, 2125-2138

15. Rodriguez-Arnedo, A., Camacho, M., Llorca, F., and Bonete, M. J. (2005) Complete reversal of coenzyme specificity of isocitrate dehydrogenase from *Haloferax volcanii*. *Protein Journal* **24**, 259-266
16. Rosell, A., Valencia, E., Ochoa, W. F., Fita, I., Pares, X., and Farres, J. (2003) Complete reversal of coenzyme specificity by concerted mutation of three consecutive residues in alcohol dehydrogenase. *Journal of Biological Chemistry* **278**, 40573-40580
17. Brinkmann-Chen, S., Flock, T., Cahn, J. K. B., Snow, C. D., Brustad, E. M., McIntosh, J. A., Meinhold, P., Zhang, L., and Arnold, F. H. (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110**, 10946-10951
18. Carugo, O., and Argos, P. (1997) NADP-Dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins: Structure, Function, and Bioinformatics* **28**, 10-28
19. Naylor, C. E., Gover, S., Basak, A. K., Cosgrove, M. S., Levy, H. R., and Adams, M. J. (2001) $NADP^+$ and NAD^+ binding to the dual coenzyme specific enzyme Leuconostoc mesenteroides glucose 6-phosphate dehydrogenase: Different interdomain hinge angles are seen in different binary and ternary complexes. *Acta Crystallographica Section D - Biological Crystallography* **57**, 635-648
20. Kavanagh, K. L., Klimacek, M., Nidetzky, B., and Wilson, D. K. (2003) Structure of xylose reductase bound to NAD^+ and the basis for single and dual co-substrate specificity in family 2 aldo-keto reductases. *Biochemical Journal* **373**, 319-326
21. Campbell, E., Chuang, S., and Banta, S. (2013) Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily. *Protein Engineering Design and Selection* **26**, 181-186
22. Plapp, B. V. (2010) Conformational changes and catalysis by alcohol dehydrogenase. *Archives of Biochemistry and Biophysics* **493**, 3-12
23. Cahn, J. K. B., Baumschlager, A., Brinkmann-Chen, S., and Arnold, F. H. (2015) Mutations in adenine-binding pockets enhance catalytic properties of NAD(P)H-dependent enzymes. *Protein Engineering Design and Selection* **29**(1), 31-38
24. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr. (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **277**, 202-206
25. Reetz, M. T., and Wu, S. (2008) Greatly reduced amino acid alphabets in directed evolution: Making the right choice for saturation mutagenesis at homologous enzyme positions. *Chemical Communications* **43**, 5499-5501
26. Mena, M. A., and Daugherty, P. S. (2005) Automated design of degenerate codon libraries. *Protein Engineering Design and Selection* **18**, 559-561
27. Engqvist, M. K. M., and Nielsen, J. (2015) ANT: Software for generating and evaluating degenerate codons for natural and expanded genetic codes. *ACS Synthetic Biology* **4**, 935-938
28. Cornish-Bowden, A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendations 1984. *Nucleic Acids Research* **13**, 3021-3030

29. Fasan, R., Meharenna, Y. T., Snow, C. D., Poulos, T. L., and Arnold, F. H. (2008) Evolutionary history of a specialized P450 propane monooxygenase. *Journal of Molecular Biology* **383**, 1069-1080
30. Bloom, J. D., Arnold, F. H., and Wilke, C. O. (2007) Breaking proteins with mutations: threads and thresholds in evolution. *Molecular Systems Biology* **3**
31. Reiße, S., Garbe, D., and Brück, T. (2015) Identification and optimization of a novel thermo- and solvent stable ketol-acid reductoisomerase for cell free isobutanol biosynthesis. *Biochimie* **108**, 76-84
32. Hamelryck, T., and Manderick, B. (2003) PDB file parser and structure class implemented in Python. *Bioinformatics* **19**, 2308-2310
33. Firth, A. E., and Patrick, W. M. (2008) GLUE-IT and PEDEL-AA: New programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Research* **36**, W281-285
34. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201
35. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
36. Gibson, D. G. (2011) Chapter Fifteen - Enzymatic assembly of overlapping DNA fragments. in *Methods in Enzymology* (Christopher, V. ed.), Academic Press. pp 349-361
37. Engqvist, M. K. M., McIsaac, R. S., Dollinger, P., Flytzanis, N. C., Abrams, M., Schor, S., and Arnold, F. H. (2015) Directed evolution of *Gloeobacter violaceus* rhodopsin spectral properties. *Journal of Molecular Biology* **427**, 205-220
38. Pazmiño, D. E. T., Snajdrova, R., Rial, D. V., Mihovilovic, M. D., and Fraaije, M. W. (2007) Altering the substrate specificity and enantioselectivity of phenylacetone monooxygenase by structure-inspired enzyme redesign. *Advanced Synthesis & Catalysis* **349**, 1361-1368
39. 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**, 83-89
40. 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**
41. Zeng, Q.-K., Du, H.-L., Wang, J.-F., Wei, D.-Q., Wang, X.-N., Li, Y.-X., and Lin, Y. (2009) Reversal of coenzyme specificity and improvement of catalytic efficiency of *Pichia stipitis* xylose reductase by rational site-directed mutagenesis. *Biotechnology Letters* **31**, 1025-1029
42. 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**, 11604-11614
43. Verho, R., Londesborough, J., Penttila, M., and Richard, P. (2003) Engineering redox cofactor regeneration for improved pentose fermentation in Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **69**, 5892-5897
44. Huffer, S., Roche, C. M., Blanch, H. W., and Clark, D. S. (2012) *Escherichia coli* for biofuel production: Bridging the gap from promise to practice. *Trends in Biotechnology* **30**, 538-545

45. Dudek, H. M., Torres Pazmiño, D. E., Rodríguez, C., de Gonzalo, G., Gotor, V., and Fraaije, M. W. (2010) Investigating the coenzyme specificity of phenylacetone monooxygenase from *Thermobifida fusca*. *Applied Microbiology and Biotechnology* **88**, 1135-1143

46. Jensen, C. N., Cartwright, J., Ward, J., Hart, S., Turkenburg, J. P., Ali, S. T., Allen, M. J., and Grogan, G. (2012) A flavoprotein monooxygenase that catalyses a Baeyer–Villiger reaction and thioether oxidation using NADH as the nicotinamide cofactor. *ChemBioChem* **13**, 872-878

47. Jensen, C. N., Ali, S. T., Allen, M. J., and Grogan, G. (2014) Exploring nicotinamide cofactor promiscuity in NAD(P)H-dependent flavin containing monooxygenases (FMOs) using natural variation within the phosphate binding loop. Structure and activity of FMOs from *Cellvibrio* sp. BR and *Pseudomonas stutzeri* NF13. *Journal of Molecular Catalysis B: Enzymatic* **109**, 191-198

48. Riebel, A., Fink, M. J., Mihovilovic, M. D., and Fraaije, M. W. (2014) Type II flavin-containing monooxygenases: A new class of biocatalysts that harbors Baeyer–Villiger monooxygenases with a relaxed coenzyme specificity. *ChemCatChem* **6**, 1112-1117

49. Maurer, S. C., Kühnel, K., Kaysser, L. A., Eiben, S., Schmid, R. D., and Urlacher, V. B. (2005) Catalytic hydroxylation in piphasic systems using CYP102A1 mutants. *Advanced Synthesis & Catalysis* **347**, 1090-1098

50. Fasan, R., Crook, N. C., Peters, M. W., Meinhold, P., Buelter, T., Landwehr, M., Cirino, P. C., and Arnold, F. H. (2011) Improved product-per-glucose yields in P450-dependent propane biotransformations using engineered *Escherichia coli*. *Biotechnology and Bioengineering* **108**, 500-510

51. Quin, M. B., and Schmidt-Dannert, C. (2011) Engineering of biocatalysts: From evolution to creation. *ACS Catalysis* **1**, 1017-1021

52. Chen, T. S., and Keating, A. E. (2012) Designing specific protein–protein interactions using computation, experimental library screening, or integrated methods. *Protein Science* **21**, 949-963

53. Davids, T., Schmidt, M., Böttcher, D., and Bornscheuer, U. T. (2013) Strategies for the discovery and engineering of enzymes for biocatalysis. *Current Opinion in Chemical Biology* **17**, 215-220

54. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. (2012) Engineering the third wave of biocatalysis. *Nature* **485**, 185-194

55. Cahn, J. K. B., Brinkmann-Chen, S., Spatzal, T., Wiig, J. A., Buller, A. R., Einsle, O., Hu, Y., Ribbe, M. W., and Arnold, F. H. (2015) Cofactor specificity motifs and the induced fit mechanism in class I ketol-acid reductoisomerases. *Biochem. J.* **468**, 475-484

Chapter 8

COFACTOR SPECIFICITY REVERSAL – STRUCTURAL ANALYSIS AND LIBRARY DESIGN (CSR-SALAD) USERS MANUAL

Material from this chapter appears online at http://www.che.caltech.edu/groups/fha/CSRSALAD/About.html, and will updated as changes to the CSR-SALAD tool are made.

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, and to render this 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. 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.

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. 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.

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. 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. All user-uploaded files are stored as temporary variables, such that they are erased immediately after the output is generated.

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 8-1. 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.

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.

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.

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 values under 40 will be rejected. Recommended library sizes might be considerably lower than the maximum entered.

Finally, CSR-SALAD allows for a handful of additional options to be selected manually under the 'Advanced Options' panel. 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. Residues of the glycine-rich "fingerprint motif" which interact with the adenine. 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 for the cofactor as a whole and therefore mutations can be very detrimental to overall cofactor binding ability.

iii. Peripheral residues 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".

CSR-SALAD results come in three sections.

i. Messages: In some cases, minor problems in structure parsing may prevent proper 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 below. The second two columns give the library recommendation made by CSR-SALAD, the former being a degenerate codon (see Table 8-1) 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.[1]

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.

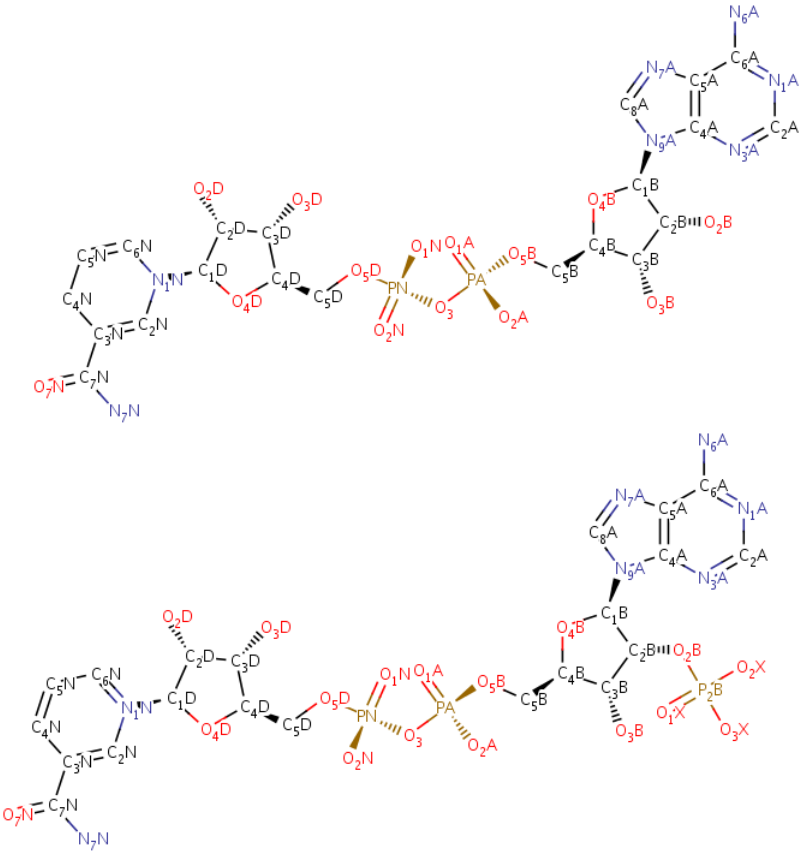


Figure 8-1. Atom labels for NAD(H) and NADP(H) from the PDB.

Table 8-1. The mixed base pair notations[2] 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.

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

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 screening of mutant libraries should be compatible with CSR-SALAD.

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.

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 to to include any mutations from the switching library when designing primers for recovery libraries to avoid accidentally reverting them to wild type.

After cloning libraries, transform and plate on agar at a density that yields distinct single colonies. 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 aliquoted to 96-well assay plates, 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 dilution 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.

While for many applications only in-cell activity is relevant, it is important to characterize kinetics of purified protein 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.

Protocols for protein purification and 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 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. When possible, check for uncoupling by running the reaction in the absence of substrate. 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.

How CSR-SALAD works

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. 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.

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.

Structural analysis

Once the structure has been uploaded and parsed with Bio.PDB, the first step is to select and classify the residues. Table 8-2 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.

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. 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 of the motif 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.[3] 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 8-3. 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. 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.

Table 8-2. Summary of the residue classifications used by CSR-SALAD.

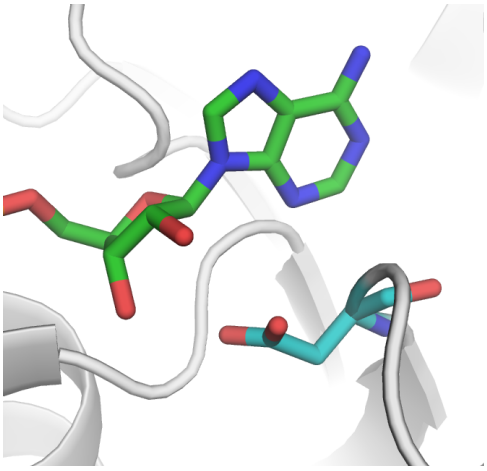
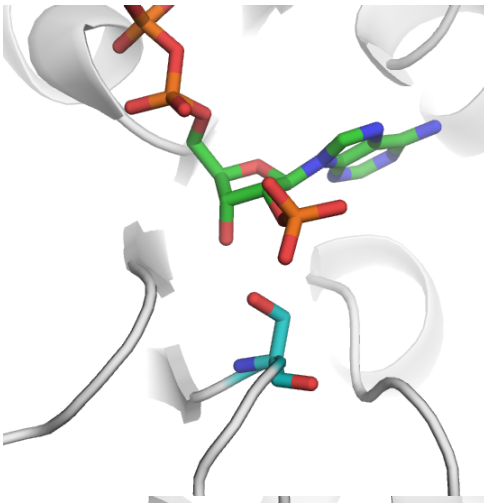
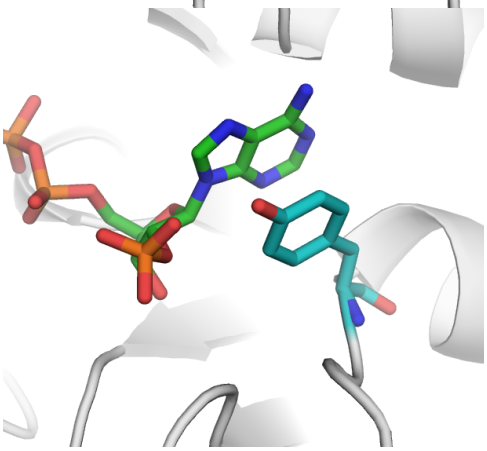
Edge (aka Floor)	These residues lie along the edge of the adenine residue and extend parallel to it toward the ribose moiety, frequently making contact with C2A, N3A, and C4A of the cofactor. In NADP-binding structures, these often do not bind the phosphate, but can be mutated to make interactions with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

Table 8-2. continued

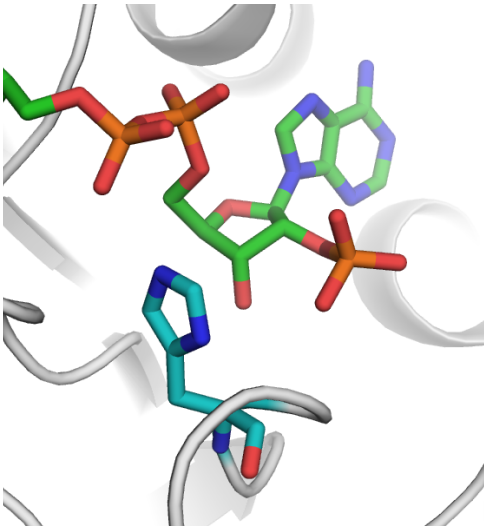
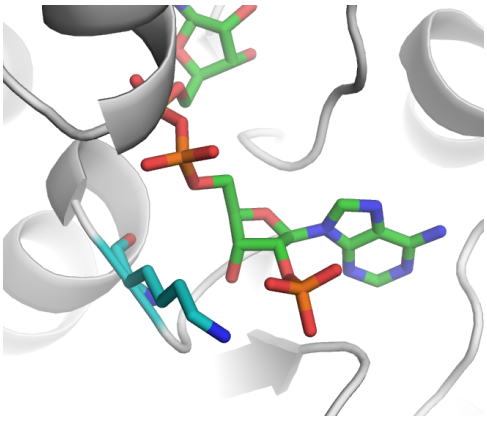
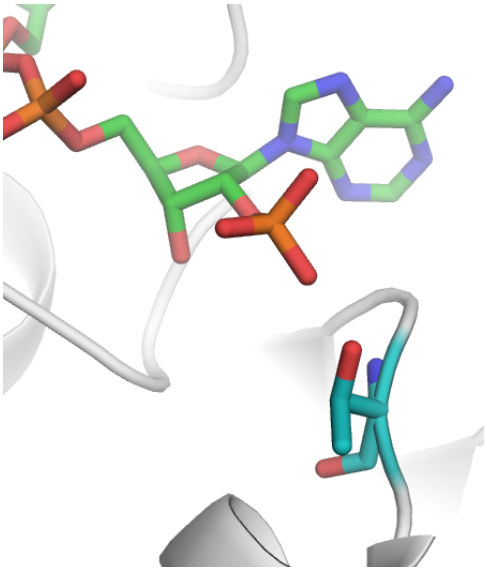
Pyrophosphate	These residues, in addition to making contact with the O2' moiety, also interact with the pyrophosphate moiety that bridges the adenosine and nicotinamide ends of the molecule. As such, these residues are not generally included for mutagenesis because they play a significant role in the 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 8-3. Key atoms of residue side chains used for pseudocenter calculations.

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}

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 8-3. 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 3D 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 $\langle x',y',z' \rangle$ composed of three vectors based on adenine atoms (Figure 8-2). 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 $x' \times y'$.

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, $\langle i,j,k \rangle$, is the vector going from the cofactor C4A atom to the residue pseudocenter. The second, $\langle l,m,n \rangle$ 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 $\boldsymbol{v}$ normal to a residues' side-chain must be calculated. Unlike $\langle i,j,k \rangle$ and $\langle l,m,n \rangle$, $\boldsymbol{v}$ 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 $\boldsymbol{v}$ is calculated as the cross product of two other vectors, $\boldsymbol{t}$ and $\boldsymbol{u}$, as defined by Table 8-4.

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 $\boldsymbol{r}$, 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 $\boldsymbol{r}$ 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 above.

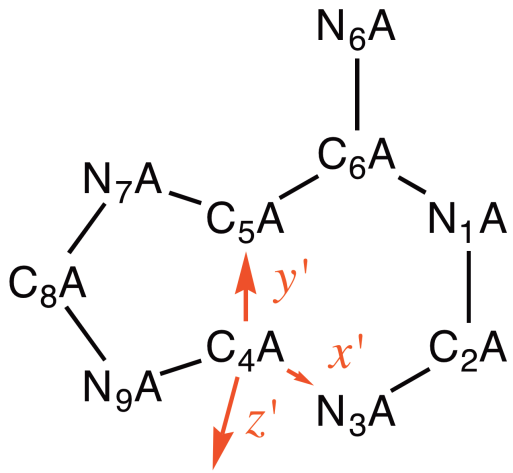


Figure 8-2. The components of *x'y'z'*-space.

Table 8-4. Atom pairs used to calculate vectors ***t*** and ***u*** and thereby $\boldsymbol{v} \equiv \boldsymbol{t} \times \boldsymbol{u}$.

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

Library design

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: (i) which mutations to residues like these have previously been shown to be useful in cofactor switching, (ii) which residues does structural intuition suggest could provide useful contacts in stabilizing the desired cofactor, and (iii) which amino acids exist at corresponding positions in the proteins that naturally bind the desired amino acid.

The degenerate codons used by CSR-SALAD are provided in Tables 8-5 and 8-6. Codons have been 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.

Table 8-5. Codons used for cofactor switching NADP-dependent proteins to NAD.

	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

Table 8-5. continued

	Edge	B/M/P	Face	S/P
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 8-6. Codons used for cofactor switching NAD-dependent proteins to NADP.

	Edge	Bidentate/ Pyrophosphate Motif/	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

Table 8-6. continued

Edge	B/M/P	Face	S/P
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

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 then that library is selected, and the whole process is iterated until that threshold is reached.

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.

The high priority level is composed 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 Chapter 6,[1] defining these residues as those containing atoms within 4.1 Å of N7A, C5A, N6A, or N1A (the top edge of the adenine moiety) but not included in the residues involved in specificity determination.

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 above. 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.

References

1. 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.
2. Cornish-Bowden, A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Research* **13**, 3021-3030
3. Bahar, I., and Jernigan, R.L. (1996) Coordination geometry of nonbonded residues in globular proteins. *Folding and Design* **1**, 357-370.

SUPPLEMENTARY INFORMATION

Supplementary Material S1-1. Citation information for Tables 1-1 and 1-2.

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.

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.

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.

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.

Table S1-1. Analysis of multiple sequence alignments of ketol-acid reductoisomerases (KARIs).

	Conserved residues responsible for cofactor contacts					
	N-terminal	C-terminal				
Domain and loop length		Antepenultimate	Penultimate	Ultimate	Compared sequences, *n*	KARI sequences, %
Bacteria						
6	R(100%)	--	K(65%), R(20%), S(6%)	S(100%)	83	13
7	R(61%), Y(16%), K(12%)	S(78%), G(22%)	--	S(76%), T(18%)	388	61
12	R(93%)	R(87%)	--	S(99%)	117	18
Archaea						
6 + 7	R(64%)	G(77%), S(17%)	--	S(89%)	47	7
Eukaryote						
7	R(100%)	S(57%), G(43%)	--	S(100%)	7	1

Conservational degree at key residues Conservational degree at key residues for specificity identified in β2αB-loop: KARI sequences were obtained from the UniProt database using Swiss-Prot manually reviewed sequence data. To date, 643 reviewed KARI sequences are available: 588 from bacterial, 47 from archaeal, and 7 from eukaryotic origin. Archaeal KARIs seem to have loop lengths of either six or seven residues, eukaryotic KARIs a loop length of seven, and bacterial ones have 6-, 7-, or 12-residue loops.

Table S1-2. Sequence identities (%) of KARIs used in this study

	EcIlvC	**ShKARI**	**SeKARI**	**MaKARI**	**LlKARI**	**AaKARI**
EcIlvC	100	72	20	20	21	22
ShKARI		100	19	22	19	21
SeKARI			100	48	53	50
MaKARI				100	52	56
LlKARI					100	57
AaKARI						100

Supplementary Material S2-1. A simplified version of the python script used to identify putative NAD-utilizing sequences

```
infile = open('alignment.fasta')
lines = infile.readlines()
infile.close()

posns = [0,1,-3,-1] #first, second, preantepenultimate, ultimate

for line in lines:
    if line[0] == '>':
        name = line[:-1]
    else:
        acidcount = 0
        s = line[309:329].replace('-','') #ends of loop determined by visual inspection. Gaps removed.
        if s != '': #catch fragmentary sequences
            for p in posns:
                if s[p] in ['D','E']: acidcount+= 1
            if acidcount!= 0:
                print '%i-point match %s: %s\n' % (acidcount, name, line[304:333])
            if len(s) not in [7,12,6]: #catch sequences of irregular length for visual inspection
                print "irregular length %s: %s\n" %(name, line[304:333])
```

Table S2-1. Table 2-1 with standard errors and Hill coefficients.

Enzyme	K_M or K_H [μM]		Hill coefficient		k_{cat} [s^{-1}]		$k_{cat}/K_{M\ or\ H}$ [$mM^{-1}s^{-1}$]		NADH/NADPH ratio of $k_{cat}/K_{M\ or\ H}$
	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH	
EcIlvC	1,075 ± 370	41 ± 3	-	-	0.3 ± 0.0	3.6 ± 0.4	0.3 ± 0.1	88 ± 11	0.003 ± 0.001
$EcIlvC^{6E6}$	30 ± 6	650 ± 80	-	-	2.3 ± 0.2	0.20 ± 0.02	74 ± 15	0.40 ± 0.05	185 ± 50
$EcIlvC^{P2D1\text{-}A1}$	26 ± 1	> 1,400	-	-	4.3 ± 0.3	0.54 ± 0.20	165 ± 22	< 0.4	> 412
HsKARI	39 ± 1	46 ± 6	-	-	0.12 ± 0.00	0.12 ± 0.01	3.2 ± 0.0	2.7 ± 0.1	1.2 ± 0.2
IaKARI	< 1	<1	-	-	0.02 ± 0.00	0.03 ± 0.00	> 20	>25	~ 0.8
MsKARI	24 ± 1	31 ± 2	3.0 ± 0.0	3.1 ± 0.6	0.06 ± 0.00	0.07 ± 0.00	2.5 ± 0.03	2.1 ± 0.03	1.2 ± 0.1
SwKARI	57 ± 2	44 ± 3	3.6 ± 0.4	3.0 ± 0.0	0.28 ± 0.01	0.22 ± 0.01	5.0 ± 0.2	5.0 ± 0.4	1.0 ± 0.1
AfKARI	5.0 ± 0.5	26 ± 1	3.5 ± 0.4	2.3 ± 0.3	0.1 ± 0.0	0.04 ± 0.00	20 ± 2	1.5 ± 0.1	13 ± 1
DoKARI	32 ± 3	n.a.	3.5 ± 0.3	n.d.	0.25 ± 0.03	n. a.	8.0 ± 1.0	n. a.	-
TpKARI	<1	40 ± 4	n.d.	3.0 ± 1.0	0.46 ± 0.04	0.25 ± 0.04	460 ± 61	6.0 ± 1.0	74 ± 17
UaKARI	1.1 ± 0.1	38 ± 2	3.5 ± 0.2	3.0 ± 0.0	0.22 ± 0.01	0.05 ± 0.00	200 ± 20	1.3 ± 0.1	152 ± 17

All enzymes were His_6-tagged and purified prior to characterization. Enzyme activities were determined in 100 mM potassium phosphate pH 7 with 1 mM DTT, 200 mM NADPH or NADH, 2.5 mM S2AL, and 10 mM $MgCl_2$. Concentrations of the purified enzymes were determined using the Bradford assay. The K_M or K_H values for the cofactors were measured with appropriate dilutions of NADPH and NADH in the presence of saturating concentrations of substrate S2AL. The Hill coefficients are given when applicable. Mutations located within the β2αB-loop of $EcIlvC^{6E6}$: A71S, R76D, S78D, and Q110V. Additional mutations in $EcIlvC^{P2D1\text{-}A1}$: D146G, G185R, and K433E.

n.d. = not determined

n.a. = not active

Table S4-1. Percent identity between class I KARIs as determined by Clustal Omega.

	%ID between AvKARI and
IaKARI	46.9%
PaKARI	93.2%
SeKARI	55.9%

Table S4-2. Amino acid identities and numberings in mesophilic class I KARIs.

Position in PaKARI	Position in AvKARI	Position in SeKARI
H107	H107	H118
K130	K130	K141
D190	D190	S201
E194	E194	E205
E226	E226	E237
E230	E230	E241
S249	S249	S260
H134	H134	H145
S26	S26	S37
P132	P132	P143
G133	G133	G144
A131	A131	G142

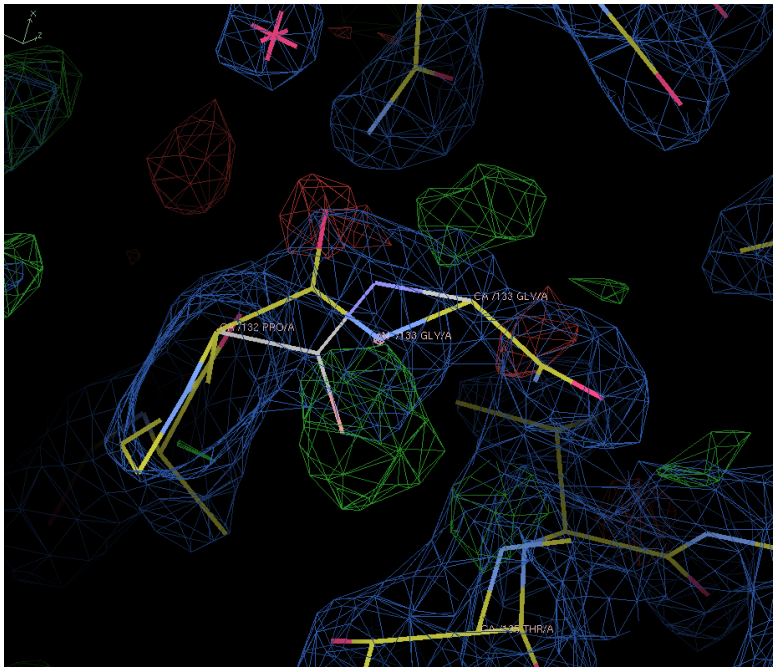


Figure S4-1. Modelling error at P132-G133 in PaKARI^{apo} (PDB 1NP3). The F_o-F_c omit map (red and green) shows the differential density around the peptide bond in the published structure (yellow) that is resolved by implementing a 180° peptide flip (white). This alteration is corroborated by improved backbone hydrogen bond interactions. ($2F_o$-F_c map (blue) contoured to 1.5σ, F_o-F_c map contoured to 2.9σ.)

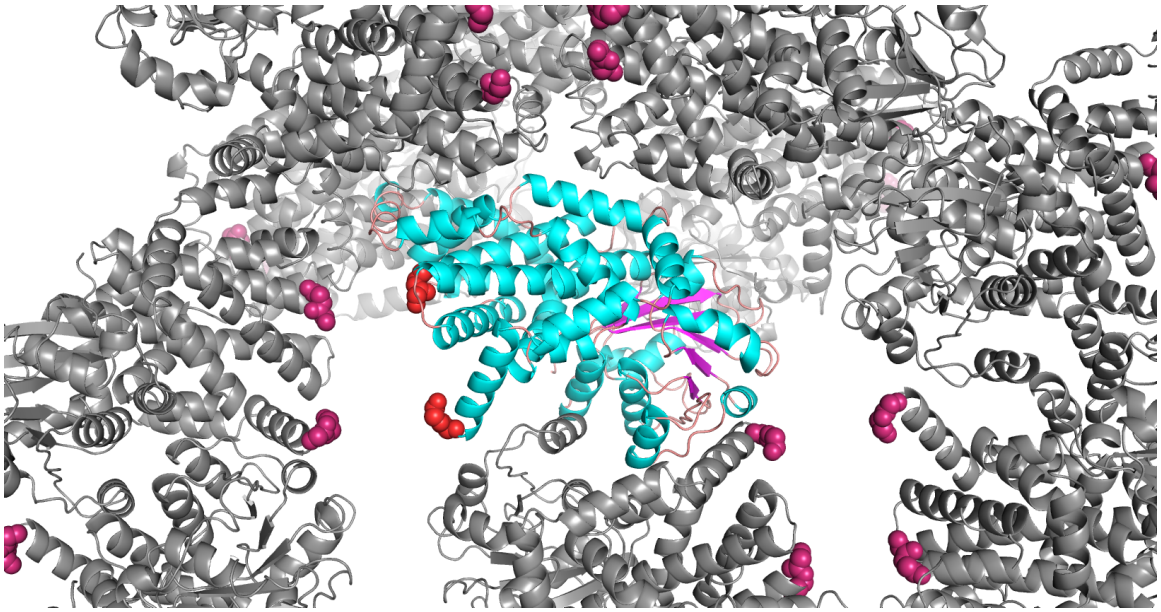


Figure S5-1. The packing of the obtained crystal of 2Ia_KARI-DD. A single protein chain (cyan and magenta) composes the asymmetric unit, and symmetry mates are shown in grey. Spheres represent the residues flanking the structurally unresolved interdomain connector.

```
IaKARI          ------------------MAKIYKDEDI-SLEPIKNKTIAILGYGSQGRAWALNLRDSGLNVVVGLERQG-----DSWRRAIDD
EcKARI          MANYFNTLNLRQQLAQLGKCRFMGRDEFADGASYLQGKKVVIVGCGAQGLNQGLNMRDSGLDISYALRKEAIAEKRASWRKATEN
2Ia_KARI-DD     ------------------MAKIYKDEDI-SLEPIKNKTIAILGYGSQGRAWALNLRDSGLNVVVGLERQG-----DSWRRAIDD
2Ia_KARI-HT     ------------------MAKIYKDEDI-SLEPIKNKTIAILGYGSQGRAWALNLRDSGLNVVVGLERQG-----DSWRRAIDD
2Ia_KARI-LT     ------------------MAKIYKDEDI-SLEPIKNKTIAILGYGSQGRAWALNLRDSGLNVVVGLERQG-----DSWRRAIDD
2Ia_KARI-2L     ------------------MAKIYKDEDI-SLEPIKNKTIAILGYGSQGRAWALNLRDSGLNVVVGLERQG-----DSWRRAIDD

IaKARI          GFKPMYTKDAVAIADIIVFLVPDMVQKSLWLNSVKDFMKKGADLVFAHGFNIHFKIIEPPKDSDVYMIAPKSPGPIVRRSYEMGG
EcKARI          GFKVGTYEELIPQADLVINLTPDK-QHSDVVRTVQPLMKDGAALGYSHGFNIVEVGEQIRKDITVVMVAPKCPGTEVREEYKRGF
2Ia_KARI-DD     GFKPMYTKDAVAIADIIVFLVPDMVQKSLWLNSVKDFMKKGADLVFAHGFNIHFKIIEPPKDSDVYMIAPKSPGPIVRRSYEMGG
2Ia_KARI-HT     GFKPMYTKDAVAIADIIVFLVPDMVQKSLWLNSVKDFMKKGADLVFAHGFNIHFKIIEPPKDSDVYMIAPKSPGPIVRRSYEMGG
2Ia_KARI-LT     GFKPMYTKDAVAIADIIVFLVPDMVQKSLWLNSVKDFMKKGADLVFAHGFNIHFKIIEPPKDSDVYMIAPKSPGPIVRRSYEMGG
2Ia_KARI-2L     GFKPMYTKDAVAIADIIVFLVPDMVQKSLWLNSVKDFMKKGADLVFAHGFNIHFKIIEPPKDSDVYMIAPKSPGPIVRRSYEMGG

IaKARI          GVPALVAVY--QNVSGEALQKALAIAKGIGCARAGVIESTFKEETETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFET
EcKARI          GVPTLIAVHPENDPKGEGMAIAKAWAAATGGHRAGVLESSFVAEVKSDLMGEQTILCGMLQAGSLLCFDKLVEEGTDPAYAEKLI
2Ia_KARI-DD     GVPALVAVY--QNVSGEALQKALAIAKGIGCARAGVIESTFKEETETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFET
2Ia_KARI-HT     GVPALVAVY--QNVSGEALQKALAIAKGIGCARAGVIESTFKEETETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFET
2Ia_KARI-LT     GVPALVAVY--QNVSGEALQKALAIAKGIGCARAGVIESTFKEETETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFET
2Ia_KARI-2L     GVPALVAVY--QNVSGEALQKALAIAKGIGCARAGVIESTFKEETETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFET

IaKARI          VNELKLIVDLIYEKGLTGMLRAVSDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGST
EcKARI          QFGWETITEALKQGGITLMMDRLSNPAKLRAYALSEQ-LKEIMAPLFQKHMDDIISGEFSSGMMADWANDDKKLLTWREETGKTA
2Ia_KARI-DD     VNELKLIVDLIYEKGLTGMLRAVSDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGST
2Ia_KARI-HT     VNELKLIVDLIYEKGLTGMLRAVSDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGST
2Ia_KARI-LT     VNELKLIVDLIYEKGLTGMLRAVSDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEKTA
2Ia_KARI-2L     VNELKLIVDLIYEKGLTGMLRAVSDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEKTA

IaKARI          IETVGRKLREMMFRGMKQISSH----------------------------------------------------------------
EcKARI          FETAPQYEGKI---GEQEYFDKG-------VLMIAMVKAGVELAFETMVDSGIIEESAYYESLHELPLIANTIARKRLYEMNVVI
2Ia_KARI-DD     IETVGRKLREMMFRGMKQISSHETDLFGEQVILVGGIMELIKASFETLVEEGYQPEVAYFETVNELKLIVDLIYEKGLTGMLRAV
2Ia_KARI-HT     IETVGRKLREMMFRGMKQISSH--------VILVGGIMELIKASFETLVEEGYQPEVAYFETVNELKLIVDLIYEKGLTGMLRAV
2Ia_KARI-LT     FETAPQYEGKI---GEQEYFDKGV------VILVGGIMELIKASFETLVEEGYQPEVAYFETVNELKLIVDLIYEKGLTGMLRAV
2Ia_KARI-2L     FETAPQYEGKI---GEQEYFDKGV------VILVGGIMELIKASFETLVEEGYQPEVAYFETVNELKLIVDLIYEKGLTGMLRAV

IaKARI          ----------------------------------------------------------------------------------
EcKARI          SDTAEYGNYLFSYACVPL-----LKPFMAELQPGDLGKA-IPEGAVDNGQLRDVNEAIRSHAIEQVGKKLRGYM-TDMKRIAVAG
2Ia_KARI-DD     SDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGSTIETVGRKLREMMFRGMKQISSH-
2Ia_KARI-HT     SDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGSTIETVGRKLREMMFRGMKQISSH-
2Ia_KARI-LT     SDTAKYGGITVGKFIIDKSVRDKMKIVLERIRSGEFAREWIKEYERGMPTVFKELSELEGSTIETVGRKLREMMFRGMKQISSH-
2Ia_KARI-2L     SDTAKYGGITVGKACVPL-----LKPFMAELQPGDLGKA-IPEGAVDNGQLRDVNEAIRSHAIETVGRKLREMMFRGMKQISSH-
```

Figure S5-2. An alignment of the sequences of IaKARI, EcKARI, and the for designed 2IaKARI constructs. Red text in the 2Ia_KARI constructs shows sequence insertions or deletions coming from EcKARI.

Table S7-1. Recapitulation of successful mutants NADP-to-NAD 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 screened do not include the previously published mutant, blue if the residue is recommended for saturation as a recovery mutation, and black if the mutant would not be found by CSR-SALAD. Mutations indicated in italics are distal to the 2' position.

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

Table S7-1. continued

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

[a]A switched enzyme has $CE_{mut}^{NAD} / CE_{mut}^{NADP} > \mathbf{1}$

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

Table S7-2. Recapitulation of successful mutants NAD-to-NADP 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 screened do not include the previously published mutant, blue if the residue is recommended for saturation as a recovery mutation, and black if the mutant would not be found by CSR-SALAD. Mutations indicated in italics are distal to the 2' position.

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

Table 7-2. continued

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

[a]A switched enzyme has $CE_{mut}^{NADP}/CE_{mut}^{NAD} > \mathbf{1}$

[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.