Published in final edited form as:
Proteins. 2004 November 1; 57(2): 294–301.

Rational Proteomics II. Electrostatic Nature of Cofactor Preference in the Short-Chain Oxidoreductase (SCOR) Enzyme Family

Vladimir Z. Pletnev[1,2], **Charles M. Weeks**[1], and **William L. Duax**[1,*]
1*Hauptman-Woodward Medical Research Institute & Dept. of Structural Biology, SUNY at Buffalo, 73 High St., Buffalo, NY 14203, USA*

2*Institute of Bioorganic Chemistry RAS, Ul. Miklukho-Maklaya, 16/10, 117997 Moscow, Russia*

Abstract

The dominant role of long range electrostatic interatomic interactions in NAD/NADP cofactor recognition has been shown for enzymes of the short-chain oxidoreductase (SCOR) family. An estimation of cofactor preference based only on the contribution of the electrostatic energy term to the total energy of enzyme-cofactor interaction has been tested for ~40 known 3D crystal complexes and ~330 SCOR enzymes with cofactor preference predicted by the presence of Asp or Arg recognition residues at specific 3D positions in the β2α3 loop.[1] The results obtained were found to be consistent with ~90% reliable cofactor assignments for those subsets. The procedure was then applied to ~170 SCOR enzymes with completely uncertain NAD/NADP dependence due to the lack of Asp and Arg marker residues. The proposed 3D electrostatic approach for cofactor assignment ("3D_ ΔE_{el}") has been implemented in an automatic screening procedure) and together with the use of marker residues proposed earlier[1] increases the level of reliable predictions for the putative SCORs from ~70% to ~90%. It is expected to be applicable for any NAD/NADP dependent enzyme subset having at least 25-30% sequence identity with at least one enzyme of known 3D crystal structure.

Keywords

short chain oxidoreductase; NAD; NADP; cofactor preference; bioinformatics; computational biochemistry

INTRODUCTION

The short-chain oxidoreductase (SCOR) family of NAD/NADP dependent enzymes[2] includes over 10,000 members identified in sequenced genomes. Of these enzymes, less than 5% have been characterized biochemically, and the three-dimensional crystal structures of ~40 have been reported. These enzymes catalyze oxidation, reduction, epimerization, and synthase reactions with different substrates in a large variety of species. There is only one fully conserved residue (the catalytic Tyr) in the entire SCOR family, and the overall average sequence homology is below 15%[2]. In the first paper in this series, [1] we described the detection of a group of ~40 sequence positions exhibiting 70% or greater conservation that define the fold fingerprint. Of these residues, 12 to 15 make hydrogen-bond contacts with the cofactor. In current databases, the majority of proteins with SCOR fingerprint sequence patterns are

*Author responsible for correspondence. Telephone: 716-856-9600 Ext. 308, Fax: 716-852-6086, Email: duax@hwi.buffalo.edu
Grant sponsor: NIH; Grant number: DK026546.

identified as putative, hypothetical, or probable oxidoreductases/dehydrogenases without defined function or simply as open reading frame products of unknown function.

We have shown that NAD or NADP cofactor preferences can be reliably predicted for ~70% of the SCORs by the presence of Asp or Arg recognition residues, respectively, in specific adjacent sequence positions in the β2α3 turn near the N-terminus[1]. In this paper, we describe a general procedure ("3D_ ΔE_{el}") which takes into consideration the extended 3D "electrostatic" environment of the bound cofactor and can be used to predict cofactor preference for that ~30% of the SCOR enzymes that lack a characteristic Asp or Arg residue.

METHODS AND SOFTWARE

The automated procedure for estimation of enzyme preference for a NAD or NADP cofactor included the following four steps:

1. A subgroup of target sequences that have at least 25-30% identity with the sequence of an enzyme with a known 3D crystal structure was generated from the SWISS-PROT and TrEMBL[3] databases using the Homology-Derived Secondary Structure of Proteins (HSSP) program[4].
2. The sequence of each target was aligned with the sequence of the parent enzyme using program CLUSTALW[5].
3. According to pairwise alignment the target protein sequence was threaded onto the 3D skeleton of the binary complex of the parent enzyme with **a** cofactor using program MODELLER[6]. A sequence identity threshold of 25-30% provides assurance that all members of this subgroup will have a similar fold and permit correct threading.[1]
4. The total potential energy of the 3D structural target model of the enzyme-NAD/NADP complex was minimized by molecular mechanics/dynamics methods using the program XPLOR_3.1[7]. Total energy optimization using CHARMM[7,8] parameterization and Lennard-Jones and Coulomb potential functions for nonbonded and electrostatic interactions, respectively, was applied to each 3D target structure to increase the discriminative signal favoring one or the other possible cofactor. The energy optimization was executed within a 12Å shell around the cofactor with a cutoff distance of 7Å for nonbonded atomic interactions and a dielectric constant of 10. It included 150 initial cycles of molecular mechanics positional refinement, 500 steps (timestep=0.001ps, temp=100K) of short term molecular dynamics followed by 500 cycles of positional refinement. The coordinates of protein C^{α} and cofactor atoms were restrained to initial positions with a force constant of 1.0 Kcal/mol. The positions of the other atoms, including those of the protein side chains, were not restrained. The difference of the electrostatic nature of the (NAD)-O′H and (NADP)-O′PO_3 groups at the adenine ribose C2′ atom was modeled by adding an extra cumulative charge difference of −0.5 to the O2′ atom that is common to both cofactors. During the last cycle of total energy minimization, the energy contribution (E_{el}) from the electrostatic interactions between both cofactors and the enzyme was estimated, and the electrostatic energy difference $\Delta E_{el} = [E_{el}(NADP) - E_{el}(NAD)]$ was computed as a predictor of cofactor preference. The value of ΔE_{el} reflects the preference of an enzyme for the presence or absence of an extra negative charge at the O2′ position. Positive and negative ΔE_{el} values are indicative of NAD and NADP preference, respectively.

Supplemental utility programs in Fortran as well as Perl and shell scripts were written to automate the procedure for cofactor preference assignment for an unlimited number of target enzymes.

RESULTS AND DISCUSSION

Testing the Procedure

Thirty-eight unique 3D crystal structures of functionally assigned short-chain dehydrogenases in complexes with NAD or NADP cofactors (Table I) were used to test the accuracy of cofactor preference assignment (Table II). All 38 structures have quite homologous cofactor-binding pockets (Fig. 1) occurring in similar Rossmann-fold domains. For this subset, the first three steps in the procedure described above were not required. Only the energy minimization step was used, and it was applied separately to the crystallographic coordinates of each enzyme complexed with NAD as well as NADP. After minimization, the root-mean-square deviations (RMSDs) between the C^{α} coordinates of the energy-minimized structure and the corresponding crystallographically observed positions were less 0.5Å.

The data in Table II show that the electrostatic field of the enzyme shell around the cofactor is sensitive to the difference in the electronic nature of the NAD and NADP cofactors, and it is likely to play a decisive role in cofactor recognition. The enhanced population of positively charged Arg and Lys residues around the adenine ribose in the NADP-dependent enzymes, relative to NAD-dependent ones, favors binding of the NADP because of its negatively charged phosphate substituent at the O2′ adenine atom (Fig. 1). The average ΔE_{el} values of +0.9 and −2.7 Kcal/mol for the NAD and NADP subsets, respectively, reflect the strength of the discriminative signal. The cofactor preference signal for NADP-dependent enzymes, which are mostly reductases, is generally strong. The less consistent result for the NAD-dependent enzymes is probably due to the fact that they involve both oxidases (NAD^+ dependent) and reductases (NADH dependent). Our attempts to increase the NAD preference signal by taking into consideration charge differences between NAD+ and NADH were unsuccessful. The effect may be masked by the cumulative error from sequence alignment, 3D skeleton threading, and computational parameterization. No consistent correlation has been found so far between the resolution of the X-ray data (Table I) and the apparent accuracy of the computational prediction (Table II).

The computational results for some functionally similar enzymes from different species indicate different magnitudes of preference for the same cofactor. Examples include NAD-dependent uridine diphosphogalactose-4-epimerase {PDB_AC:1A9Z (*E. coli*) and 1EK5 (human)} as well as NADP-dependent β-ketoacyl-[acyl carrier protein] reductase {1EDO (*Brassica napus*) and 1I01 (*E. coli*)}. Examination of the pairwise sequence alignments and the 3D superimposed structures for the first pair revealed ~30% amino acid differences in a 7Å s hell (significant for electrostatic interactions) around the adenine of the NAD cofactor. Some of the differences (C34H, K37F, R60L, N61D, D98R and N99V using 1A9Z numbering) dominate the local electrostatic field. The reduction of the cumulative positive charge surrounding the adenine in 1EK5 relative to that in 1A9Z probably is the major source for its relatively enhanced preference for NAD. The second pair, 1EDO and 1I01, displays a 40% amino acid difference in the adenine area (Y48T, R50T, A51E, K52N, K78D and L125T using 1EDO numbering), and this difference alters the local electrostatic field. A more positive electrostatic field accompanied by noticeable conformational differences in fragments 49-55 and 102-105 in 1EDO relative to 1I01 accounts for the relatively larger NADP preference of 1EDO.

The cofactor preference predicted by the value of ΔE_{el} is consistent with the experimental biochemical evidence and crystallographic observations for 32 of the 38 test crystal structures. Two apparently incorrect assignments (1BDB and 1DOH) and four questionable ones (1G1A, 1ZID, 1D8A, 1KEP) provide a measure of reliability of the procedure and identify an uncertainty zone, ($+0.5 > \Delta E_{el} > -0.5$), characterized by weak preference for “nad” and “nadp”

at positive and negative ΔE_{el} values, respectively. Outside this zone, the cofactor preference assignment may be considered highly probable.

The procedure has also been tested on several enzymes for which cofactor specificity has been inverted by experimental site-specific mutations (Table III). In these examples,[12-16] most mutations are in the area around the adenine ribose. The predicted cofactor preferences for the wild type and mutated enzymes show good agreement with experimental data. The results indicate that increasing or decreasing the positive electrostatic field in the vicinity of the adenine ribose shifts enzyme preference to NADP or NAD, respectively.

As mentioned above, only step 4 of the procedure described in "METHODS AND SOFTWARE" was applied to the test structures. The validity of steps 2 and 3 was also checked independently. The sequence of the NAD-dependent crystal structure 1FMC was threaded on the skeleton of the NADP-dependent crystal structure 1CYD with which it shares 32% sequence identity and has a RMSD value of 0.8Å for C^{α} atoms within an ~10Å envelope around the cofactor . After energy optimization (step 4), the calculated value of ΔE_{el}=+2.1kcal/mol appeared to be quite close to that of +1.9kcal/mol for 1FMC based on step 4 alone (Table II).

Applications

The HSSP program[4] was used to identify a subgroup of 534 full-length enzymes from the combined SWISS-PROT and TrEMBL databases that have at least 30% sequence identity with the crystal structure of mouse carbonyl reductase (1CYD), and the cofactor preference prediction procedure was applied to this subgroup. The 1CYD structure was chosen because it is one of the SCOR crystal structures having high resolution, good refinement statistics (Table I), and stereochemically correct geometry essential for correct threading of the target sequences. Of the enzymes in this subgroup (most of which are putative), 333 or ~62% have Asp or Arg cofactor residues specifying NAD and NADP preference[1] at 3D positions corresponding to #38 and #39 in the β2α3 loop of the parent 1CYD crystal structure. The "D38" and "R39" subsets contain 118 and 215 structures, respectively. The 29 cases (~5%) that have both Asp38 and Arg39 present compose the "D38_R39" subset. In the "D38" subset, position #39 is occupied by Ile and Leu residues in ~65% of the structures; in the "R39" subset, position #38 is filled by Gly, Ala, Ser and Tyr in ~90% of the structures.

D & R subsets—The automatic "3D_ΔE_{el}" procedure applied to both the "D38" and the "R39" subsets showed over 95% consistency in the cofactor preference predictions compared to those based on Asp or Arg presence alone.[1] For the "D38" subset, 111 enzymes were identified as NAD dependent and 7 (probably erroneously) as NADP dependent (Fig. 2a). For the "R39" subset, 206 enzymes were predicted to be NADP dependent, and 9 were predicted to be NAD dependent (Fig. 2b). Among those nine cases, four enzymes (seq_AC: Q8ZK23, Q8ZGK4, Q92MP6, Q9A7A9) have a Glu residue at position #38.

DR subset—In 24 out of 29 cases, NAD binding was predicted for the "D38_R39" subset (Table IV). The corresponding average value ($<\Delta E_{el}>$ = +1.5 kcal/mol), compared to the +0.9 kcal/mol value for the NAD-dependent test group (mostly "D" subset, Table II), reflects the specificity of the NAD electrostatic environment of the "DR" subset. The only available crystal structure of a SCOR enzyme with this characteristic "DR" pair in the β2α3 loop (1A4U; Table II) revealed the dominant role of the Asp in determining NAD binding, and the ΔE_{el} value of +1.4 kcal/mol is similar to that of the "DR" subset. Among the five indications of NADP preference, two (Q8YJQ6 and Q8YY83) are identified as 3-ketoacyl-[acyl carrier protein] reductases, and in most cases such reductases are NADP dependent.

No D & no R subset—The "3D_ΔE_{el}" procedure was of greatest potential value when applied to the remaining 172 (32%) SCOR enzymes that lack characteristic Asp and Arg residues in the β2α3 loop. The majority of these proteins are putative enzymes with uncertain NAD/NADP dependence. The corresponding data are presented in Table V. Of these 172 predictions, 28 (16.3%) fall within the uncertainty zone of ±0.5kcal/mol , and the corresponding assignments have to be viewed with caution. Among 16 cases (9.3%) with biochemically known cofactor preference, two enzymes (Seq_AC: Q8YJ18 and Q05528) were incorrectly assigned as NADP dependent. It is important to note that the four residues (Gly, Ala, Ser and Tyr) that typically precede Arg39 in the "R39" subset (~90%) also occur frequently (~82%) at the same position in 136 enzymes predicted to be NADP dependent in the "no D38 & no R39" subset. Among these residues, Tyr38 has the highest population (~55%), and it could be considered as an additional single-residue indicator of NADP preference. The limited number of predictions for NAD dependence among enzymes from the "no D38 & no R39" subset does not allow us to draw reliable conclusions. Based on biochemical assignments for 16 structures from this subset (Table V) as well as for 38 crystal structures from the test subset (Table II), the expected error in cofactor prediction is less than 20%.

Dual specificity—Cases in which wild type enzyme has been found to have both NAD and NADP dependence are particularly interesting. For example, in many organisms glucose 6-phosphate dehydrogenase (G6PD) is strictly NADP dependent. However, *L. mesenteroides* G6PD exhibits dual cofactor specificity[17]. The existence of two crystal structures of this enzyme[18] in complexes with both NAD (1H94) and NADP (1H9A) allowed us to compute separately for each complex the value of the electrostatic energy from the interactions between the enzyme and the corresponding bound cofactor. The calculated difference between the two values, −2.4 kcal/mol, indicates a preference for NADP that is in general agreement with the biochemical kinetic study (k_{cat}/K_m for NADP is 10 fold higher than for NAD).[18]

The explanation for the apparent dual specificity may lie in differences in the crystallization conditions involving ionic strength and pH.[18] NAD selectivity was achieved using a protein solution containing $CaCl_2$ at final pH 7.5. In contrast, NADP selectivity was achieved using a solution containing NH_4SO_4 at a lower pH (5.8) thereby increasing the protonation state of several residues, in particular Asp, Glu and His. Such variations, accompanied by local changes in enzyme conformation, undoubtedly produce corresponding variation in the charge environment of the cofactor. Variations in pH and ionic strength in the media (*in vivo* or *in vitro*) may result in significant alterations in charge distribution at the cofactor-binding site leading to altered cofactor preference. Such external influences on cofactor charge environment are not addressed in the calculation of ΔE_{el} presented here. It is also possible that some of the enzymes for which the calculated value of ΔE_{el} falls within the zone of uncertainty (−0.5 to +0.5 kcal/mol) may have dual cofactor specificity. We plan to test this possibility experimentally.

CONCLUSIONS

The role of long-range electrostatic interatomic interactions for NAD/NADP cofactor recognition has been illustrated. A 3D "electrostatic" approach for cofactor preference estimation has been implemented in an automatic screening procedure and successfully applied to a representative subset of 534 functionally known and putative short-chain oxidoreductase (SCOR) family members. This approach has an expected error level ≤ 20%. When used together with the Asp/NAD and Arg/NADP correlation previously described,[1] it increases the level of reliable cofactor assignments for putative SCOR enzymes from ~70% to ~90%. The proposed approach is expected to be applicable for any NAD/NADP-dependent enzyme subset sharing at least 25-30% sequence identity with one parent enzyme of known 3D crystal structure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This research was supported by NIH grant DK026546.

REFERENCES

1. Duax WL, Pletnev V, Addlagatta A, Bruenn J, Weeks CM. Rational proteomics I. Predicting fold and cofactor preference in the short-chain oxidoreductase (SCOR) enzyme family. Proteins 2003;53:931–943. [PubMed: 14635134]
2. Jörnvall H, Höög JO, Persson B. SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. FEBS Letters 1999;445:261–264. [PubMed: 10094468]
3. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 2000;28:45–48. [PubMed: 10592178]
4. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68. [PubMed: 2017436]
5. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680. [PubMed: 7984417]
6. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–81. [PubMed: 8254673]
7. Brunger, AT. X-PLOR (version 3.1) Manual. Yale University; New Haven, CT: 1992.
8. MacKerell, AD.; Brooks, B., Jr; Brooks, CL.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The energy function and its parameterization with an overview of the program. In: Schleyer, PR., et al., editors. The Encyclopedia of Computational Chemistry. 1. John Wiley & Sons; Chichester: 1998. p. 271-277.
9. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540. [PubMed: 7723011]
10. Bernstein FC, Koetzle TF, Williams GJB, Meyer ER, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542. [PubMed: 875032]
11. http://www.ncbi.nlm.nih.gov/Database
12. Cho H, Oliveira MA, Tai HH. Critical residues for the coenzyme specificity of NAD+ -dependent 15-hydroxyprostaglandin dehydrogenase. Arch Biochem Biophys 2003;419:139–146. [PubMed: 14592457]
13. Nakanishi M, Matsuura K, Kaibe H, Tanaka N, Nonaka T, Mitsui Y, Hara A. Switch of coenzyme specificity of mouse lung carbonyl reductase by substitution of threonine 38 with aspartic acid. J Biol Chem 1997;272:2218–2222. [PubMed: 8999926]
14. Galkin A, Kulakova L, Ohshima T, Esaki N, Soda K. Construction of a new leucine dehydrogenase with preferred specificity for NADP+ by site-directed mutagenesis of the strictly NAD+ -specific enzyme. Protein Eng 1997;10:687–690. [PubMed: 9278282]
15. Serov AE, Popova AS, Fedorchuk VV, Tishkov VI. Engineering of coenzyme specificity of formate dehydrogenase from *Saccharomyces cerevisiae*. Biochem J 2002;367:841–847. [PubMed: 12144528]
16. Bocanegra JA, Scrutton NS, Perham RN. Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. Biochemistry 1993;32:2737–2740. [PubMed: 8457541]
17. Levy HR. Glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides*. Biochem Soc Trans 1989;17:313–315. [PubMed: 2753211]
18. Naylor CE, Gover S, Basak AK, Cosgrove MS, Levy HR, Adams MJ. 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 Cryst D Biol Crystallogr 2001;57:635–648. [PubMed: 11320304]

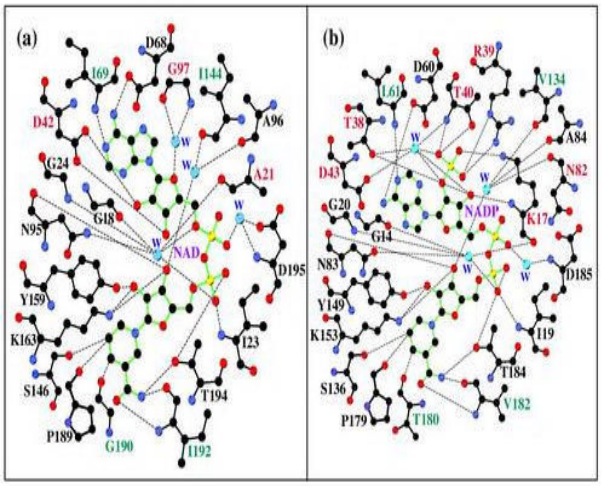

Fig. 1.

LIGPLOT schematics illustrating the hydrogen bond (≤3.35Å) net work around (a) the NAD cofactor in 7α-hydroxysteroid dehydrogenase (1FMC) and (b) the NADP cofactor in mouse carbonyl reductase (1CYD). The names (one letter code) of the residues making identical, similar and distinctly different contacts are shown in black, green and red respectively. The major differences (indicated by red) are mostly located around the adenine part of the cofactor. The conserved water molecules are marked by W symbols.

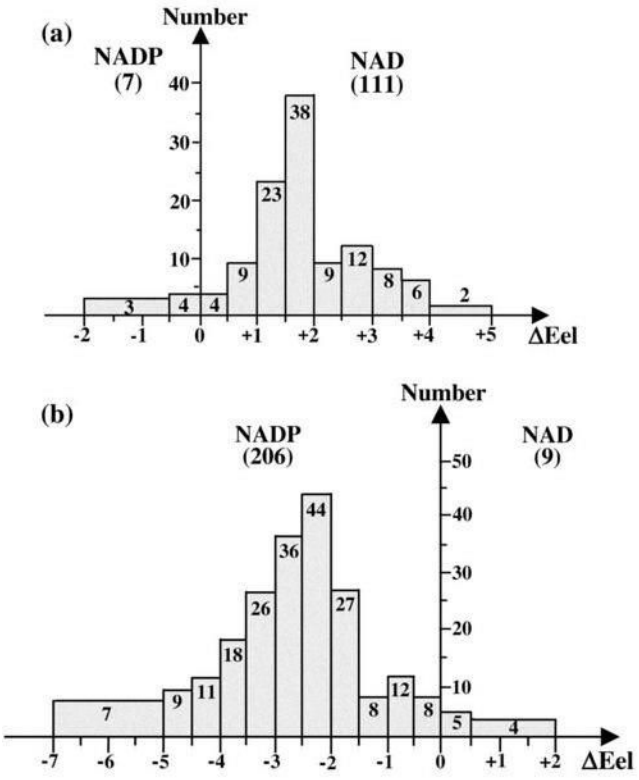

Fig. 2.

The histogram of the cofactor preference assignment for the enzymes from (a) “D38” and (b) “R39” subsets with indicated numbers of predictions for each ΔE_{el} energy shell.

TABLE I

SCOR Enzymes[9] with Known Three-Dimensional X-Ray Structures

No[a]	Protein name	Source	Cofactor	PDB AC[b]	Res (Å)[c]	Rstand (R_{free})[d]
1	Dihydropteridine reductase	Human	NAD	1HDR	2.5	0.17
2		Rat	NAD	1DHR	2.3	0.15
3	3-hydroxyacyl-CoA dehydrogenase	Rat	NAD	1E6W	1.7	0.18 (0.23)
4	7-α-hydroxysteroid dehydrogenase	E. coli	NAD	1FMC	1.8	0.21 (0.25)
5	3/17 β-hydroxysteroid dehydrogenase	Comamonas testosteroni	NAD	1HXH	1.2	0.15 (0.18)
6	dTDP-6-deoxy-L-lyxo -4-hexulose reductase	Salmonella enterica	NAD	1KBZ	2.2	0.19 (0.26)
7	3-α-hydroxysteroid dehydrogenase	Comamonas testosteroni	NAD	1FK8	2.0	0.19 (0.22)
8	Alcohol dehydrogenase	Fruit fly	NAD	1A4U	1.9	0.21 (0.24)
9	UDP-galactose 4-epimerase	Human	NAD	1EK5	1.8	0.19
10		E. coli	NAD	1A9Z	1.9	0.19
11	meso-2,3-butanediol dehydrogenase	Klebsiella pneumoniae	NAD	1GEG	1.7	0.19 (0.21)
12	Sulfolipid biosynthesis protein	Thale cress	NAD	1QRR	1.6	0.17 (0.19)
13	3-α,20-β-hydroxysteroid dehydrogenase	Streptomyces hydrogenans	NAD	2HSD	2.6	0.19
14	Negative transcriptional regulator	Aspergillus nidulans	NAD	1K6X	1.5	0.19 (0.22)
15	dTDP-glucose 4,6-dehydratase	E. coli	NAD	1BXK	1.9	0.19
16		Salmonella enterica	NAD	1G1A	2.5	0.20 (0.25)
17		Streptococcus suis	NAD	1KEP	1.8	0.14 (0.17)
18	Enoyl-[acyl carrier protein]	Oil seed rape	NAD	1D7O	1.9	0.16 (0.23)
19	reductase	Mycobacterium tuberculosis	NAD	1ZID	2.7	0.20 (0.30)
20		E. coli	NAD	1D8A	2.2	0.22 (0.29)
21	Glucose dehydrogenase	Bacillus megaterium	NAD	1GCO	1.7	0.18 0.19
22	Cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase	Pseudomonas sp.	NAD	1BDB	2.0	0.18 (0.23)
23	Mannitol dehydrogenase	Mushroom	NADP	1H5Q	1.5	0.19 (0.21)
24	Biliverdin IX β reductase	Human	NADP	1HDO	1.2	0.12 (0.16)
25	17β-hydroxysteroid dehydrogenase	Human	NADP	1FDT	2.2	0.19 (0.24)
26	Sepiapterin reductase	Mouse	NADP	1OAA	1.3	0.20 (0.22)
27	20β-hydroxysteroid dehydrogenase	Pig	NADP	1N5D	2.3	0.19 (0.25)
28	β-ketoacyl-[acyl carrier protein]	Oil seed rape	NADP	1EDO	2.3	0.19 (0.24)
29	reductase	E. coli	NADP	1I01	2.6	0.23 (0.25)
30	1,3,6,8-tetrahydroxynaphthalene reductase	Rice blast fungus	NADP	1JA9	1.5	0.19 (0.22)
31	Dihydropteridine reductase	Leishmania major	NADP	1E7W	1.8	0.20 (0.24)
32	GDP-mannose 4,6-dehydratase	E. coli	NADP	1DB3	2.3	0.20 (0.23)
33	Carbonyl reductase	Mouse	NADP	1CYD	1.8	0.15 (0.20)
34	GDP-fucose synthetase	E. coli	NADP	1BSV	2.2	0.17
35	ADP-L-glycero-D-mannoheptose 6-epimerase	E. coli	NADP	1EQ2	2.0	0.21 (0.26)
36	Tropinone reductase	Jimsonweed I	NADP	1AE1	2.4	0.16 (0.25)
37		Jimsonweed II	NADP	2AE2	1.9	0.17 (0.21)
38	1,3,8-trihydroxynaphtalene reductase	Rice blast fungus	NADP	1DOH	2.1	0.21 (0.26)

[a] Arabic numbering corresponds to that in Table II.

[b] Protein Data Bank accession codes.[10]

[c] Resolution of the crystallographic data in Angstroms.

[d] Crystallographic R factors (Rstandard and R_{free}).

TABLE II

"Electrostatic" Estimation of the Crystal Structure Subset with Known NAD/NADP Cofactor Preference

No	NAD Dependent PDB_AC[a]	ΔE_{el}[c] (kcal/mol)	Charged Residues[b]	No	NADP Dependent PDB_AC[a]	ΔE_{el}[c] (kcal/mol)	Charged Residues[b]
1.	1HDR	+3.0	D40, R18	23.	1H5Q	−5.1	R21, R43
2.	1DHR	+2.6	D37, R15	24.	1HDO	−5.0	R35, R39, R78
3.	1E6W	+2.4	D41	25.	1FDT	−4.8	R37, K40
4.	1FMC	+1.9	D42	26.	1OAA	−4.8	R18, R43
5.	1HXH	+1.8	D37	27.	1N5D	−4.4	K14, R37, R41
6.	1KBZ	+1.6	D30	28.	1EDO	−4.1	R27, R50
7.	1FK8	+1.5	D32, R34	29.	1I01	−1.3	R15
8.	1A4U	+1.4	D37, R38	30.	1JA9	−3.1	R39
9.	1EK5	+1.3	D33	31.	1E7W	−2.7	K16, R39
10.	1A9Z	+0.5	D31	32.	1DB3	−2.6	K32, R33
11.	1GEG	+1.2	D33	33.	1CYD	−2.3	K17, R39
12.	1QRR	+0.8	D32, R36, R37	34.	1BSV	−1.8	R12, R36
13.	2HSD	+0.7	D37, R16	35.	1EQ2	−1.4	K34, K38, K53
14.	1K6X	+0.6		36.	1AE1	−1.4	K31, R53, E57
15.	1BXK	+0.5	D33, K34	37.	2AE2	−1.0	R19, R41
16.	1G1A	0.0	D32, K33	38.	1DOH	+2.4	R39
17.	1KEP	−0.4	D37, K38				
18.	1D7O	+0.1					
19.	1ZID	−0.2					
20.	1D8A	−0.3					
21.	1GCO	+0.1	R39, K41				
22.	1BDB	−1.7	D36, K37, R41				
$\langle(\Delta E_{el})\rangle$[c]		+0.9		$\langle(\Delta E_{el})\rangle$[c]		−2.7	

[a] Protein Data Bank (PDB) accession code.

[b] Charged residues in the contact shell of the O′H/O′PO3 group of the NAD/NADP cofactor

[c] $\Delta E_{el} = [E_{el}(NADP) - E_{el}(NAD)]$ is the energy difference of the electrostatic interactions between the enzyme and the NADP and NAD cofactors, respectively. $\langle \Delta E_{el} \rangle$ = averaged ΔE_{el}.

TABLE III

NAD/NADP Preference Prediction for the Enzyme Subset with Mutation-Inverted Cofactor Specificity

No.	Enzyme (Seq_AC) [a]	PDB_AC (seq_id %) [bc]	Mutation Sites	Wild Type Specificity	Wild Type ΔE_{el} [d]	Mutant ΔE_{el} [d]
1.	15_PGDH (P15428) [e]	1IY8 (28%)	Q15K,D36A,W37R	NAD	+1.9	−3.6
2.	CR [f]	1CYD (100%)	T38D	NADP	−2.3	+0.7
3.	LeuDH (Q60030) [g]	1C1X (36%)	D203A,I204R,D210R	NAD	+2.2	−2.4
4.	*Sce*FDH [h] (NP_01503311)	2NAD (43%)	D196A,Y197R	NAD	+2.5	−1.4
5.	dhLADH (P00391) [i]	1LVL (37%)	G185A,G189A,E203V, M204R,F205K,D206H,P210R	NAD	+1.6	−2.3

[a] Accession code to SWISS-PROT[3] and NCBI[11] sequence databases

[b] Protein Data Bank accession code

[c] Alignment sequence identity between the target protein and the PDB crystal structure (indicated by PDB_AC) used as a host skeleton for sequence threading

[d] $\Delta E_{el} = [E_{el}(NADP) - E_{el}(NAD)]$

[e] 15-Hydroxyprostaglandin dehydrogenase type I (human)[12]

[f] Carbonyl reductase (mouse)[13]

[g] Leucine dehydrogenase (*Thermoactinomyces intermedius*)[14]

[h] Formate dehydrogenase (*Saccharomyces cerevisiae*)[15]

[i] Dihydrolipoamide dehydrogenase (*Escherichia coli*)[16]

TABLE IV

NAD/NADP Cofactor Preference Prediction for the Putative "D38_R39" Subset*

No.	Seq_AC[a]	ΔE_{el}[b] (kcal/mol)	No.	Seq_AC[a]	ΔE_{el}[b] (kcal/mol)	No.	Seq_AC[a]	ΔE_{el}[b] (kcal/mol)
1.	Q9W4U2	+4.1	11.	Q9ZBX8	+1.3	21.	P73991	+0.8
2.	Q8U673	+2.5	12.	Q56841	+1.3	22.	Q8UA44	+0.8
3.	Q8YDU5	+2.3	13.	Q98EN0	+1.3	23.	Q92PX8	+0.7
4.	Q93FY3	+2.2	14.	Q8Z8K9	+1.2	24.	Q930L3	+0.3
5.	Q986G2	+2.1	15.	Q92NF8	+1.2	25.	Q9L7Y2	−0.3
6.	Q9PCQ2	+1.8	16.	Q8VLS3	+1.2	26.	Q8UB53	−0.4
7.	Q9RJU5	+1.7	17.	Q9EX28	+1.1	27.	Q8YJQ6	−0.9
8.	Q8UE64	+1.6	18.	Q8ZR30	+1.1	28.	Q8YY83	−1.7
9.	Q92MQ9	+1.4	19.	Q8YD94	+1.1	29.	Q8UKE2	−2.1
10.	Q988J7	+1.3	20.	Q9VNF4	+0.8			

*Subset that includes enzymes with both of the specific Asp and Arg residues at the cofactor recognition positions 38 and 39.

[a]Accession code to combined SWISS-PROT/TrEMBL sequence database[3].

[b]$\Delta E_{el} = [E_{el}(NADP) - E_{el}(NAD)]$

TABLE V

NAD/NADP Preference Assignment for 172 Members of Putative “No D & No R” Subset Characterized by Lack of Positionally Specific Asp or Arg Cofactor Recognition Residues: Sequences Identified by SWISS-PROT/TrEMBL Accession Codes (Seq_AC)

NAD assignment ($\Delta E_{el} \geq +0.5$)														
Q9ZNN8*	Q56318	Q9ZW14	P39071*	Q8U3B3	O74159	Q8U759	Q98B18	Q9S2E4	Q9EWT4	Q8TFD5	Q97Y09	Q8RDG3	O49332	Q937L4*
Q98D05	P50171	Q8YF18	Q8Y690	Q8UUM4	Q99RG1	Q8VS64*	Q8ZFH9	Q9I4V1	Q9I1X3	P39345*				
Uncertain cofactor preference with “nad” inclination ($0.0 \leq \Delta E_{el} < +0.5$)														
Q9URL2	Q8W309	Q9URL3	Q9ZGI8	O94749	Q930L5	Q99QK7	O93791	O86034*	Q8RLU4					
NADP assignment ($\Delta E_{el} \leq -0.5$)														
Q9URK8	O00058	Q93LX9	Q9S0N7	O74139	Q8U6D0	O74155	Q944H4	Q9MA93	O74141	O74160	Q8RDH9	Q9ZGC1	O74144	Q93FA7
P50205*	Q9X7T4	Q9HFV6*	O74150	Q98K90	O74158	Q9C496	Q9I6V0	O74140	Q934T1	Q9FK50	Q9KA99	O74149	Q9RKU3	P51831*
Q9HZ96	Q9Y8Y1	Q8TTL5	P50161	Q93S07	O74152	O74153	O74147	Q987G9	O93802	O74157	Q978P7	O74156	Q9K486	O42693
Q00791	Q9UR11	P40397	Q53674	Q98ML3	P73684	P73574*	Q8YVT0	Q9C425	Q8UBE2	Q9RKS8	Q9SQR4	Q9SQR2	Q8UBJ0	Q9K3Z8
O86853	O93874|	Q8VML6	P72332	O67610*	Q92YU8	Q9A322	Q986J1	Q92MQ0	Q98FT0	Q9HWT0	Q8UFJ0	Q9PKF7*	P06234	O53927
Q8YFP3	Q9KGT0	Q92L02	Q9BL81	Q98CV1	Q8UFY9//	Q8VMZ9	Q8XVS6	Q9JW61	Q8Y0J2	Q9JXR1	Q97VV6*	Q8X8I5	Q8ZUD8	P25716*
Q976W9	Q9Z245	Q97CM7	Q8X7R8	Q8YJ18^	Q9RPB0	Q9Z3Y5◆	Q51496	Q9ZB52*	Q54491	Q05528^	Q9ZGD1	Q9KCM1	Q931A8	Q9HK58
Q8X4M2	Q8XJ34	P33368	Q9ZW16	Q9ZGD4	Q8RX32	Q9A514	Q9ZI71	Q9HKZ9	Q8RG25	Q92UX4	Q9ANU6	Q94AT6		
Uncertain cofactor preference with “nadp” inclination ($-0.5 < \Delta E_{el} < 0.0$)														
O74154	O74151	O74145	P52037	Q97IR4	O74138	O74143	Q93M13	Q98HP1	Q92LJ9	O74142	Q9KA03	Q92AK1|	O93790	Q987X7 //◆
Q978I3	Q9RB80	Q9F7E0												

ΔE_{el} = [E_{el}(NADP) − E_{el}(NAD)] is the energy difference of the electrostatic interactions between the enzyme and the NADP and NAD (lower case “nadp” and “nad” abbreviations used for uncertainty zone) cofactors, respectively

In NADP/nadp subset symbol

After ◆ other 6

* Biochemically proved assignment (according to SWISS-PROT/TrEMBL data)

^ Erroneous assignment (according to SWISS-PROT/TrEMBL data)

| terminates the group with Tyr38,

// – Gly38 and

◆ – Ala38/Ser38;