THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Hidden Relationship between Conserved Residues and Locally Conserved Phosphate-Binding Structures in NAD(P)-Binding Proteins

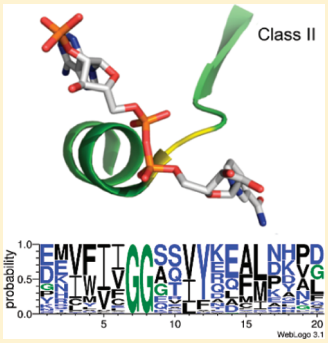
Chih Yuan Wu,[†] Yun Hao Hwa,[†] Yao Chi Chen,[†] and Carmay Lim*[,†,‡]

[†]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan

[‡]Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan

Supporting Information

ABSTRACT: A one-dimensional (1D) motif usually comprises conserved essential residues involved in catalysis, ligand binding, or maintaining a specific structure. However, it cannot be easily detected in proteins with low sequence identity because it is difficult to (1) identify protein sequences suspected to contain the motif, and (2) align sequences with little sequence identity to spot the conserved residues. Here, we present a strategy for discovering phosphate-binding 1D motifs in NAD(P)-binding proteins sharing low sequence identity that overcomes these two hurdles by determining all distinct locally conserved pyrophosphate-binding structures and aligning the same-length sequences comprising each of these structures to identify the conserved residues. We show that the sequence motifs derived from the distinct pyrophosphate-binding structures yield different numbers/spacing of conserved Gly residues. We also show that they depend on the side chain orientations and cofactor type (NAD or NADP). Thus, sequence motifs derived from local similarity of backbone structures without consideration of the cofactor type and/or side chain orientations would reduce their reliability in annotating protein function from sequence alone. The three-dimensional (3D) and 1D motifs comprising conserved residues in nonredundant proteins reveal hidden relationships between the protein structure/function and sequence as well as protein−cofactor interactions.



INTRODUCTION

The number of protein sequences has increased at an unprecedented rate in recent years. Although the number of three-dimensional (3D) protein structures has also increased, the 3D structures of only a small fraction of known protein sequences have been solved. For example, more than 20 million protein sequences have been indexed in the Release 2012_03 UniProt database[1] (~0.54 million in the Swiss-Prot and ~20.6 million in TrEMBL), but only ~81 000 protein structures have been deposited in the Protein Data Bank (PDB).[2] In the absence of structural data, sequence similarity search tools are useful in annotating protein function and in aiding the design of experiments for further studies. If an uncharacterized protein has a sequence homologous to a protein of known function, its function may be predicted by whole sequence comparisons.[3,4] If, however, it exhibits insignificant sequence identity to proteins of known function, its function may be predicted using local sequence or one-dimensional (1D) motifs comprising a few conserved, essential residues.[4−8] The accuracy of a sequence motif, however, relies on overcoming two hurdles: (1) how to identify a set of protein sequences suspected to contain the motif, and (2) how to correctly align sequences sharing low sequence identity to spot the conserved residues.[9]

For proteins with insignificant sequence identity, the residues that are crucial for the protein's fold, stability, and/or function may be conserved not only in residue type yielding a 1D motif, but also in their conformations, forming a structural or 3D motif.[10] A prime example of proteins with low sequence identity whose common function can be associated with a 3D motif is proteins with the Rossmann fold,[11] a well-known 3D motif composed of three β-strands and two helices in the order $\beta\alpha\beta\alpha\beta$. Although Rossmann-fold enzymes catalyze many different reactions,[12] their common function is to bind dinucleotides such as nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP), collectively referred to as NAD(P).[13,14] The fingerprint for this common function is a stretch of 30−35 residues from the first two β strands and the connecting phosphate-binding helix.[14] Alignment of the $\beta\alpha\beta$ sequences forming the functional 3D motif in several NAD(P)-binding proteins revealed a phosphate-binding consensus sequence, G-X_{1-2}-G-X-X-G.[15,16] Thus, although Rossmann-fold proteins show low overall sequence similarity, they possess a conserved 1D motif within a $\beta\alpha\beta$ motif involved in binding the NAD(P) pyrophosphate moiety.

The phosphate-binding G-X_{1-2}-G-X-X-G motif is relatively short, and exceptions to this sequence motif have been found,[14] so it is not considered as a phosphate or pyrophosphate-binding signature in the PROSITE database.[7] Hence, a few studies, based either on 1D sequences or 3D structures, have attempted to extend the definition of the G-X_{1-2}-G-X-X-G motif or discover new sequence motifs for annotating nucleotide-binding function. From analyses of helix−β-strand interactions in Rossmann-fold proteins, Kleiger and Eisenberg[17] found G-X-X-X-G and

Received: February 13, 2012
Revised: April 23, 2012
Published: April 24, 2012

G-X-X-X-A motifs following the phosphate-binding G-X_{1-2}-G-X-X-G motif that stabilize flavin adenine dinucleotide (FAD) and NAD(P)-binding Rossmann folds through C^{α}−H···O hydrogen bonds and van der Waals (vdW) interactions. They proposed an extended the [V/I]-G-X_{1-2}-G-X-X-G-X-X-X-[G/A] sequence motif as an indicator of Rossman folds that bind FAD or NAD(P). Using geometric matching to compare phosphate-binding sites of Rossmann-fold proteins, Brakoulias and Jackson[18] found G-X-X-X-G-I-G, G-X-G-X-V-G, and G-X-G-X-X-G motifs in four clusters of these proteins. Using the motifs described in previous works,[18,19] Gherardini et al.[20] proposed modules to form nucleotide-binding pockets.

Although NAD(P)-binding sequence motifs have been found in Rossmann-fold proteins, several outstanding questions remain. First, are the 3D structures corresponding to a given sequence motif locally conserved? For example, do all structures with the [V/I]-X-G-X_{1-2}-G-X-X-G-X-X-X-[G/A] sequence motif share similar backbone conformations and side chain orientations? Second, are the NAD(P)-binding sequence motifs found in Rossmann-fold proteins also present in other NAD(P)-binding proteins that do not share sequence and overall structural similarity with Rossmann-fold proteins? Third, is the same sequence motif involved in binding the common pyrophosphate moiety of both NAD and NADP?

Since Rossmann-fold proteins with low overall sequence identity, performing different overall functions, possess a 1D motif with conserved residues embedded in a 3D functional motif, we present a strategy to discover 1D phosphate-binding motifs in NAD(P)-binding proteins from the corresponding distinct 3D motifs. Most previous studies employ 1D information to identify 3D motifs,[21−29] but, to the best of our knowledge, they do not comprehensively determine all distinct locally conserved structures corresponding to a given function and use them to identify 1D motifs and conserved residues, as proposed herein. Like the Rossmann-fold proteins, the region for binding pyrophosphates was assumed to contain the first β-strand, the following phosphate-binding α-helix, and the connecting loop. The distinct locally conserved structures corresponding to these pyrophosphate-binding "$\beta\alpha$" segments in *nonredundant* NAD(P)-binding domains were identified with the aid of a 16-letter structural alphabet (see Materials and Methods). These were then used as structural templates to scan *redundant* NAD(P)-binding domains for matching structures. This generated a set of distinct, locally conserved, phosphate-binding structures from all the NAD(P)-binding domains in the current PDB. The sequences comprising each distinct pyrophosphate-binding structure, which are of the same length, were aligned to generate 1D motifs for NAD- and NADP-binding proteins separately; in this way, we overcome the two hurdles mentioned above. The 3D and 1D functional motifs comprising conserved residues in proteins sharing insignificant sequence identity reveal hidden, unexpected relationships between the protein structure and sequence as well as protein−cofactor interactions.

MATERIALS AND METHODS

Data Sets of Redundant and Nonredundant NAD(P)-Binding Domains. Two redundant data sets were created by searching the PDB[2] for ≤3.5 Å X-ray structures of proteins bound to oxidized or reduced NAD(P). For NAD(P) proteins with multiple subunits, only one conformation was included. If two or more structures correspond to proteins with identical sequences, then the highest resolution structure was kept. This generated 503 NAD- and 443 NADP-binding redundant domains.

To generate a set of nonredundant NAD(P)-binding domains, domains sharing identical CATH codes[30] were grouped together, and the domain with the best resolution structure was chosen as the representative of that group. As an example, let d1, d2, and d3 denote three domains with different CATH codes, and let d1+d2, d1+d3, and d2+d3 comprise three proteins X, Y, and Z, whose structures have been solved at 1.0, 1.5, and 2.0 Å, respectively. As the structure of protein X has the best resolution, the d1 and d2 domains in protein X will be chosen as the representative of d1 and d2 domains, respectively. Likewise, the d3 domain in protein Y would be the representative of d3 domains. This procedure yielded 44 NAD- and 33 NADP-binding nonredundant domains, each containing ≥1 domains with a unique CATH code combination, sharing ≤30% sequence identity with other proteins in the nonredundant data set (Supporting Information Table S1).

Converting a 3D Protein Structure to a 1D Structural Letter Sequence. Each protein structure was encoded into its 1D structural sequence according to the structural alphabet,[31] which was derived as follows: The backbone of each protein from a nonredundant protein structure database was represented by consecutive five-residue segments, each described by a vector of eight backbone dihedral angles $V(\psi_{n-2}, \phi_{n-1}, \psi_{n-1}, \phi_n, \psi_n, \phi_{n+1}, \psi_{n+1}, \phi_{n+2})$. The dissimilarity between the two vectors V_1 and V_2 of dihedral angles was measured by the root-mean-square deviation (RMSD) of the dihedral angle values:

$$\mathrm{RMSDa}(V_1, V_2) = \sqrt{\frac{\sum_{i=1}^{4} [\psi_i(V_1) - \psi_i(V_2)]^2 + [\phi_{i+1}(V_1) - \phi_{i+1}(V_2)]^2}{8}} \quad (1)$$

Using an unsupervised cluster analyzer based on the RMSDa of the segments, 16 letters were identified and described in previous work,[31] e.g., the letter *m* describes a central α-helix ($\psi = -47°$, $\phi = -57°$), while the letter *d* describes a central β-strand ($\psi = 135°$, $\phi = -139°$). These 16 letters comprise the structural alphabet.

To convert an *l*-residue protein structure into a string of structural letters, the sequence was scanned using a five-residue sliding window. The structure of each five-residue segment was compared with that of each of the 16 letters, and the letter that had the closest structure (as measured by the RMSDa) to the five-residue segment was assigned to the middle residue of that segment[26] (see Figure 1). The terminal 4 residues of each protein, which cannot be treated as center residues of the five-residue segments, were assigned the letter Z.

Defining Pyrophosphate-Binding $\beta\alpha$ Segments. On the basis of the secondary structures of the NAD(P)-binding domains and pyrophosphate-binding residues assigned by PDBSUM,[32] the $\beta\alpha$ segment was defined as the first residue of the first β-strand to the last residue of the following α-helix. Since favorable interactions between the pyrophosphate moiety and the α-helix dipole is known to drive the formation of NAD(P)-enzyme complexes,[33] those $\beta\alpha$ segments lacking properly oriented phosphate-binding helices were excluded from subsequent analyses. For example, the 1ej2-A structure was not considered because the turn instead of the helix of the $\beta\alpha$ segment binds to the pyrophosphate. $\beta\alpha$ structures with pyrophosphate-binding helices were found in 24 NAD- and 25 NADP-binding nonredundant domains (Table S1) and in 404 NAD- and 327 NADP-binding redundant domains. As the secondary structures assigned by different programs may differ, the pyrophosphate-binding $\beta\alpha$ segment was defined to include a residue flanking each

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644−5652

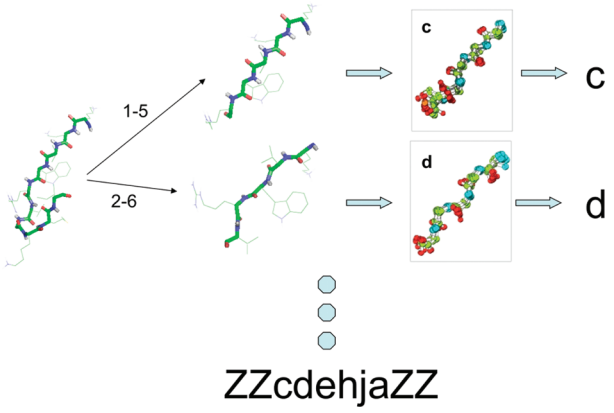


Figure 1. Converting a 3D protein structure to a 1D structural letter sequence. A hypothetical structural sequence *ZZcdehjaZZ* is used as an example. The representative structures of the structural letters *c* and *d* are taken from de Brevern et al., 2000.[31]

side of the respective $\beta\alpha$ sequence. For example, the $\beta\alpha$ sequence in UDP-glucose 6-dehydrogenase (1dlj-A) spans residues 2−19 so the pyrophosphate-binding $\beta\alpha$ segment included residues 1−20.

Defining Pyrophosphate-Binding 1D Structural Sequences. The 3D structures of the pyrophosphate-binding $\beta\alpha$ segments in the nonredundant data set were converted into 1D structural sequences. The pyrophosphate-binding $\beta\alpha$ structural sequences in Table S1 contain various strings of the structural letters *d* and *m*, which represent a β-strand and a helix of varying length, respectively. As they have in common the structural letter *e* or *f* representing a C-terminal β-strand cap, they were aligned according to the position of the first *e* or *f*. Because the $\beta\alpha$ structural sequences of some NAD(P)-binding domains (e.g., 1i24-A, 2dvm, 2ph5, 2pv7, 1gq2, 1pgo) have only two letters before the structural letter *e*, the first few letters were removed until all the structural sequences have only two letters before the letter *e* or *f*. Furthermore, letters were removed from the end of the $\beta\alpha$ structural sequences until the lengths are all 12 to facilitate comparison of their structures (see below). For example, the 21-letter sequence in the PDB entry 1dlj, *Zddedjbfklmmmmmmmm*, was truncated to the 12-letter sequence, *ddedjbfklmmm*. Those NAD(P)-binding domains with identical pyrophosphate-binding structural sequences such as the NAD-binding domains of UDP-glucose 6-dehydrogenase (1dlj-A), hypothetical malate oxidoreductase (2dvm-A), and ferredoxin reductase (2yvf-A) were grouped. This yielded 18 and 17 pyrophosphate-binding structural sequences for the NAD- and NADP-binding domains, respectively (Table 1). However, four of the pyrophosphate-binding structural sequences for the NAD-binding domains (*ddedjbfklmmm*, *ddehjafkmmmm*, *ddehjbfklmmm*, and *ddehjllmmmmm*) are identical to those found for the NADP-binding domains. Hence, 31 pyrophosphate-binding structural sequences were derived from the nonredundant NAD(P)-binding domains (Table 1).

Defining Structural Similarity. To determine whether two structures are similar or distinct, we used two similarity measures; viz., the root-mean-square deviation of C^α atoms (RMSD) and dihedral angles (RMSDa). The structures corresponding to the 12-letter pyrophosphate-binding structural sequences encompassing 16 residues were superimposed using the pair fit function of PyMol.[34] Because the structures of the central residues are better superimposed than those of the end residues, two residues were excluded from each end. For each pair of 12-residue structures without missing or modified residues, the C^α RMSD and the RMSDa values were computed; the latter employed the following equation, viz.,

$$\mathrm{RMSDa}(V_1, V_2) = \sqrt{\frac{\sum_{i=1}^{11}[\psi_i(V_1) - \psi_i(V_2)]^2 + [\phi_{i+1}(V_1) - \phi_{i+1}(V_2)]^2}{22}} \quad (2)$$

In addition to computing the C^α rmsd and RMSDa to assess backbone similarity, the structures were visually inspected to ensure that they also share similar side chain orientations. If two 12-residue structures share similar backbone conformations but were visually found to exhibit different side chain orientations, the differences in the backbone ϕ and ψ angles of each residue were computed, and the largest ϕ and/or ψ angle differences were identified. These differences were used to distinguish structures with similar backbone conformations but different side chain orientations (see next section).

Table 1. Distinct Pyrophosphate-Binding Structures Corresponding to the Structural Sequences Derived from the Nonredundant NAD(P)-Binding Domains

$\beta\alpha$ structural sequence	PDB code of NAD-binder[a]	PDB code of NADP-binder[a]	structural class[b]
cdedjb fklmmm		1get	I
cdedjf kmmmmm		1yve	I
cdehjb fklmmm	2q3e		I
ccedjb fklmmm		1pgo	I
dcehia fkmmmm	1lj8		I
ddedjb fklmmm	1dlj, 2dvm, 2yvf,	1gq2, 1h7x, 2q0l, 3etd	I
ddedjb fkmmmm		1nyt	I
ddedjd fklmmm	1pjs, 1uwk		I
ddedjd fkmmmn	2ph5		I
ddehja fklmmm	1l7e		I
ddehja fkmmmm	1sc6, 2dt5	1yqd	I
ddehjb fklmmm	1c1d, 2jhf	1e5q	I
ddehjb fmmmmm	1t2d		I
ddehjd fklmmm	3d64		I
ddehjd fkmmmm	1lc3		I
Zdehji fklmmm		2i76	I
cdebdf klmmmm		1qzf	II
cdebjf klmmmm		3fs6	II
cdehil mmmlmm	2pv7		III
ddehil mmmlmm	1i24		III
ddehjl lmmlmm		1q0q	III
ddehkl bmklmm		1lua	III
ddehkl mmmlmm		1e6u, 1mb4, 1y7t	III
ddehkl gcjlmm		1zk4	IV
ddeheh jmlmmm	1rlz		V
cdedjf klmmmm	1jq5		VI
ddehja gklmmm		1dqa	VII
ddeehi aklmmm	1qax		VIII
ddehjl lmmmmm	1ib0	1ja1, 1qfz	IX
cdehia klopfk		1d4o	X
ddfbbf kbcfbd	3i9k		XI

[a]PDB entries of the nonredundant NAD(P)-binding domains with the given $\beta\alpha$ structural sequence. [b]Roman numerals denote a distinct pyrophosphate-binding structure, as shown in Figure 2.

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644−5652

RESULTS

Pyrophosphate-Binding 3D Motifs from Nonredundant NAD(P)-Binding Domains. To determine the distinct structures corresponding to the pyrophosphate-binding structural sequences derived from the nonredundant NAD(P)-binding domains in Table 1, we computed the pairwise C^{α} RMSD and RMSDa values. The results in Supporting Information Tables S2A and S2B, along with visual inspection of the side chain orientations, reveal 11 distinct pyrophosphate-binding structures. The first 16 pyrophosphate-binding structural sequences in Table 1 share similar structures with pairwise C^{α} RMSD values $\leq$ 1.6 Å and/or RMSDa values $\leq 44°$; hence their structures were assigned to class I (Figure 2A). The structures corresponding to *cdebdfklmmmm* (1qzf) and *cdebjfklmmmm* (3 fs6) from NADP-binding domains superimpose well with C^{α} RMSD = 0.4 Å, RMSDa = 39°, so they were assigned as class II (Figure 2B). The structures corresponding to *cdehilmmmlmm* (2pv7), *ddehilmmmlmm* (1i24), *ddehjllmmlmm* (1q0q), *ddehklbmklmm* (1lua), *ddehklmmmlmm* (1e6u, 1mb4, 1y7t) superimpose with C^{α} RMSD $\leq$ 1.5 Å and/or RMSDa $\leq 18°$, so they were assigned as class III (Figure 2C). Although these structures also superimpose well with the *ddehklgcjlmm* 1zk4 structure (C^{α} RMSD $\leq$ 1.3 Å), they exhibit different side chain orientations, as shown in Figure 2C,D, and their ϕ_8, ϕ_9, and ψ_9 angles (where the subscript denotes the structural letter position) differ from those of the *ddehklgcjlmm* 1zk4 structure by more than 50, 110, and 70°, respectively (see Supporting Information Table S2C). Hence, the *ddehklgcjlmm* structure was assumed to be distinct and assigned to class IV (Figure 2D). Likewise, the *ddehehjmlmmm* (1rlz) and *cdedjfklmmmm* (1jq5) structures superimpose with C^{α} RMSD = 1.3 Å, but their ϕ_5, ϕ_6, ϕ_7/ψ_7 angles differ by 59, 40, and 113°, respectively (Table S2C), hence they were assumed to be distinct and assigned to classes V (Figure 2E) and VI (Figure 2F), respectively. Along the same vein, the *ddehjagklmmm* (1dqa) and *ddeehiaklmmm* (1qax) structures superimpose well with C^{α} RMSD = 0.7 Å, but their ψ_3, ϕ_4, ψ_4, ϕ_5, ψ_5 and ϕ_6 angles differ by 79, 76, 60, 124, 93 and 155°, so they were assigned to classes VII (Figure 2G) and VIII (Figure 2H), respectively. The remaining three pyrophosphate-binding structural sequences in Table 1 exhibit distinct structures (Figure 2I–K), so their structures were assigned to separate classes (IX, X, and XI).

A Novel Pyrophosphate-Binding 3D Motif from Redundant NAD(P)-Binding Domains. For each distinct pyrophosphate-binding structure in Figure 2A–K, the structure corresponding to the central 12 residues was used to scan the $\beta\alpha$ structures in the 404 NAD- and 327 NADP-binding redundant domains using a 12-residue sliding window, and a match was recorded if the pairwise C^{α} RMSD was $\leq$1.0 Å, the RMSDa was $\leq 30°$, and the side chain orientations of the two structures were similar. Using these criteria, 359 NAD- and 297 NADP-binding redundant domains were found to match the pyrophosphate-binding structures belonging to class I–IV, VI, IX and X (Table 2). The $\Delta\phi_8+\Delta\phi_9+\Delta\psi_9$ differences between each pair of class III and IV structures were >150°, while the $\Delta\phi_5+\Delta\phi_6+\Delta\phi_7+\Delta\psi_7$ differences between each pair of class V and VI structures were >300°, indicating that the class III and IV structures as well as class V and VI structures are distinct.

For the remaining pyrophosphate-binding $\beta\alpha$ structures in the redundant data set that did not match any of the pyrophosphate-binding structures in Figures 2A–K, their structures were superimposed. The 3D structures of 30 $\beta\alpha$ segments were found to superimpose well with C^{α} RMSD < 1.0 Å,

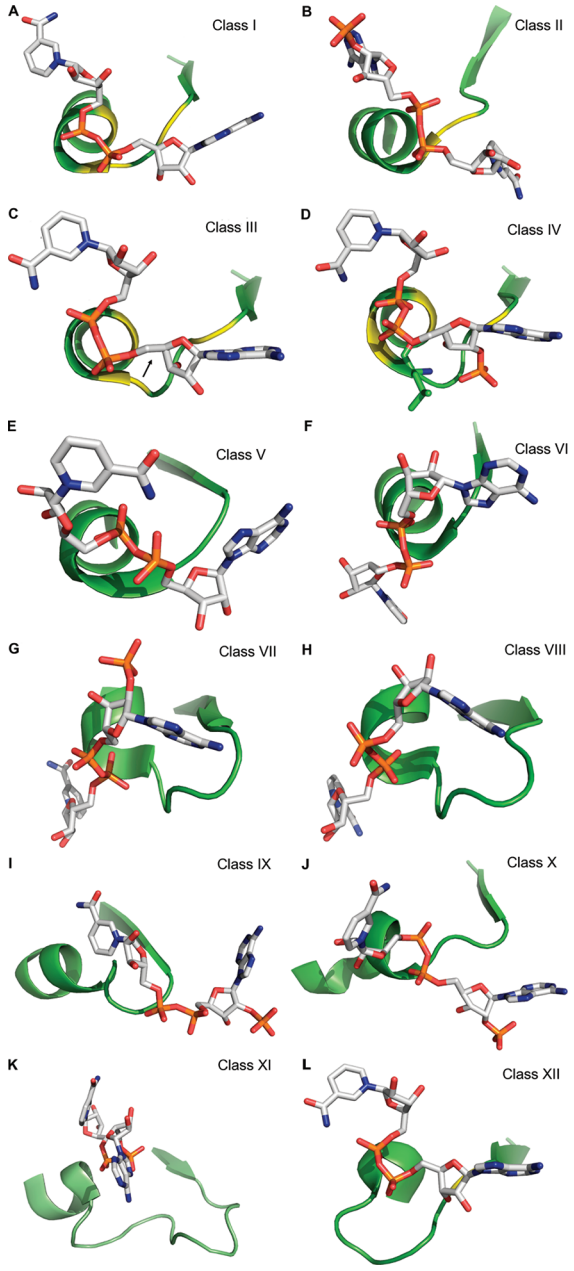


Figure 2. The distinct locally conserved pyrophosphate-binding structures of $\beta\alpha$ segments derived from nonredundant NAD(P)-binding domains. (A) class I: 2jhf-A (193–208), (B) class II: 3 fs6-A (110–125), (C) class III: 1sby-A (6–21), (D) class IV: 1zk4-A (7–22), (E) class V: 1rlz-A (276–291), (F) class VI: 3hl0-A (87–102), (G) class VII: 1dqa-A (646–661), (H) class VIII: 1qax-B (676–691), (I) class IX: 1qfz-A (159–174), (J) class X: 1d4o-A (99–114), (K) class XI: 3i9k-B (158–178), and (L) class XII: 3oig-A (8–23). For each structural class, the structure shown corresponds to the PDB structure with the highest resolution. The secondary structures shown are based on the default PDB definition. The structure of the central 12 residues corresponding to each 12-letter $\beta\alpha$ structural sequence was pair-fitted to the corresponding class I 2jhf-A structure. Although the class III and IV structures share a common backbone conformation (pairwise C^{α} RMSD values <1.0 Å), they exhibit different side chain orientations: the side chain of Leu in 1zk4-A (D), shown in stick, and the corresponding side chain in 1sby-A (C), indicated by the black arrow, point in opposite directions. For each class, the regions containing a conserved Gly in the NAD(P)-binding proteins are highlighted in yellow.

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644–5652

Table 2. Correlation between the Distinct Pyrophosphate-Binding 3D Motifs and the Fold/Function of the NAD(P)-Binding Domain

	NAD-binding domain			NADP-binding domain		
structural class *x*	#[a]	CATH code[b]	EC number[c]	#[a]	CATH code[b]	EC number[c]
I	222	**3.40.50.720**	1.1.1.*n*[d]	81	**3.40.50.720**	1.1.1.*m*[e]
			1.1.99.28			1.1.99.28
			1.2.1.*n*[f]			1.14.13.92
			1.3.1.(24/76)			1.2.1.(12/13/59)
			1.4.1.(1/3/20/21)			1.3.1.(2/24/43)
			1.5.1.(2/11)			1.4.1.(3/16)
			1.6.1.2			1.5.1.10
			2.1.1.107			1.6.4.5
			2.5.1.44			1.8.1.5
			3.2.1.49			4.2.1.10
			3.3.1.1			
			4.99.1.4			
			5.5.1.4			
		3.50.50.60	1.11.1.1		3.50.50.60	1.3.1.2
			1.8.1.(4/7)			1.8.1.(7/9/12)
						1.18.1.2
		3.40.50.10730	4.2.1.49			
II	0			52	3.40.430.10	1.1.1.193
						1.5.1.3
						2.1.1.45
III	78	**3.40.50.720**	1.1.1. *n*[g]	83	**3.40.50.720**	1.1.1.*m*[h]
			1.3.1.(12/26)			1.2.1.(11/38)
			1.5.1.34			1.3.1.*m*[i]
			3.13.1.1			1.5.1.(5/30)
			3.2.1.122			1.6.5.(2/5)
			4.1.1.35			2.3.1.94
			4.2.1.46			4.2.1.47
			5.1.3.*n*[j]			5.1.3.20
			5.4.99.5			
			5.5.1.4			
IV	45	**3.40.50.720**	1.1.1.*n*[k]	60	**3.40.50.720**	1.1.1.*m*[l]
			1.3.1.56			1.3.1.(10/34)
			4.2.1.(107/119)			1.5.1.33
V	1	3.40.910.10	2.5.1.46	0		
VI	10	3.40.50.1970	1.1.1. *n*[m]	1		
			2.5.1.19			
			2.7.1.71			
			4.2.1.10			
			4.2.3.4			
VII	0			1	3.30.70.420	1.1.1.34
VIII	1	3.30.70.420	1.1.1.88	0		
IX	1	3.40.50.80	1.6.2.2	14	3.40.50.80	1.6.2.4
						1.14.13.39
						1.18.1.2
X	0			5	3.40.50.1220	1.6.1.2
XI	1	**3.40.50.720**		0		
XII	27	**3.40.50.720**	1.3.1.9	3	**3.40.50.720**	1.3.1.9
						1.3.1.10

[a]The total number of NAD/NADP-binding domains with $\beta\alpha$ structures of class *x*, whose pairwise C^{α} RMSDs are ≤1.0 Å and RMSDa values are ≤30°. [b]The CATH code of Rossmann-fold domains, 3.40.50.720, is highlighted in bold. [c]The Enzyme Commission numbers of all NAD/NADP-binding domains of the respective CATH code corresponding to $\beta\alpha$ structures of class *x*; 1.3.1.(24/76) denotes 1.3.1.24 and 1.3.1.76. [d]*n* = 1, 3, 8, 14, 18, 22, 24, 27, 28, 31, 35, 38, 45, 67, 95, 103, 282, 284, 290. [e]*m* = 2, 8, 25, 26, 40, 44, 47, 79, 81, 86, 237, 292. [f]*n* = 2, 10, 12, 13, 46, 59. [g]*n* = 1, 37, 49, 203. [h]*m* = 37, 49, 82, 133, 219, 267, 271. [i]*m* = 10, 24, 26, 38, 48,74. [j]*n* = 2, 3, 10, 18. [k]*n* = 30, 35, 47, 50, 53, 62, 63, 107, 141, 159, 178, 268, 275, 304. [l]*m* = 10, 62, 100, 138, 146, 153, 184, 189, 197, 206, 236, 248, 252. [m]*n* = 1, 6, 25, 77.

RMSDa < 30°, and their side chain showed similar orientations. This new pyrophosphate-binding 3D motif was found in enoyl acyl-carrier-protein reductases and was assigned as structural class XII (Table 2 and Figure 2L). The remaining pyrophosphate-binding $\beta\alpha$ structures could be assigned to structural classes I–IV using visual inspection and less stringent

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644–5652

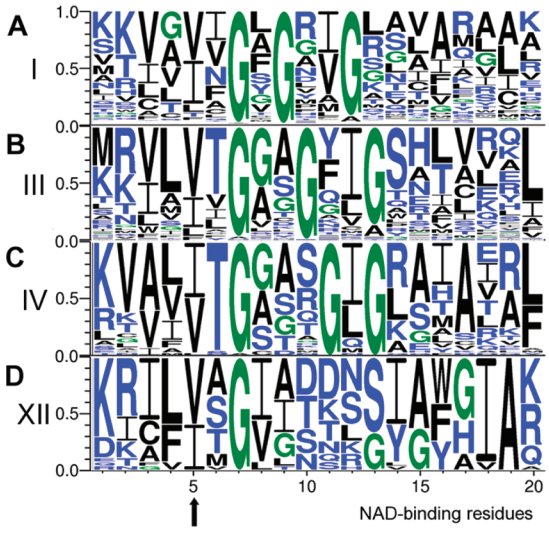


Figure 3. Sequence logos derived from the distinct pyrophosphate-binding $\beta\alpha$ structures of NAD-binding proteins. (A) class I, (B) class III, (C) class IV, and (D) class XII pyrophosphate-binding $\beta\alpha$ structures composed of 20 residues from NAD-binding proteins. Glycine is shown in green, polar residues (S, T, Y, N, Q, H, K, R, D, E) in blue, and nonpolar residues (A, V, L, I, P, W, F, C, M) in black. The arrow indicates the position of the conserved structural letter *e* or *f*, representing a C-terminal β-strand cap.

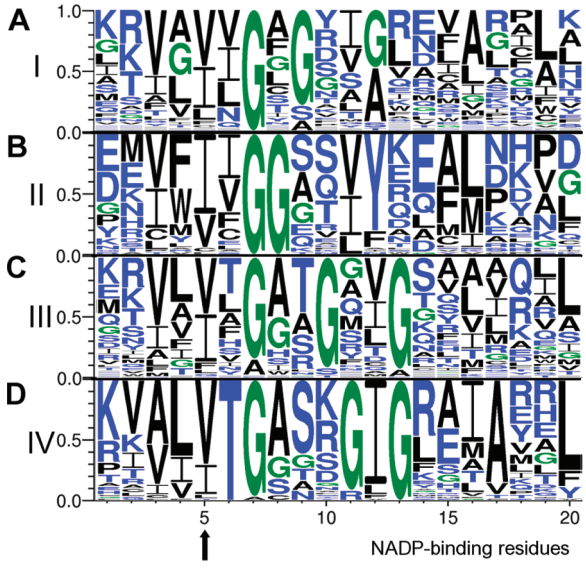


Figure 4. Sequence logos derived from the distinct pyrophosphate-binding $\beta\alpha$ structures of NADP-binding proteins. (A) class I, (B) class II, (C) class III, and (D) class IV pyrophosphate-binding $\beta\alpha$ structures composed of 20 residues from NADP-binding proteins. The coloring scheme is the same as Figure 3.

RMSDa and/or C^{α} RMSD criteria. For example, 14 $\beta\alpha$ structures could be assigned to structural class I–III based on C^{α} RMSD < 1 Å, but the RMSDa values range from 31 to 56°.

Correlation between the Distinct Functional 3D Motifs and the Protein Fold/Function. Why are there so many distinct locally conserved structures for the same function of binding the NAD(P) pyrophosphate? To address this question, we evaluated whether similar pyrophosphate-binding structures correlate with the NAD(P)-binding domain fold and function by listing the CATH codes[30] and Enzyme Commission (EC) numbers[35] of the NAD(P)-binding domains of each pyrophosphate-binding structural class. The results in Table 2 reveal that a given pyrophosphate-binding structure generally correlates with the NAD(P)-binding domain fold and/or function. For example, the class II, VI, and IX pyrophosphate-binding structures are found in NAD(P)-binding domains with CATH codes 3.40.430.10, 3.40.50.1970, and 3.40.50.80, respectively. Although the class I, III, IV, and XII pyrophosphate-binding structures are found in Rossmann-fold domains (CATH code 3.40.50.720), they generally correlate with the EC numbers. For example, Rossmann-fold enzymes with the class XII structure share a common EC number of 1.3.1.9, characteristic of enoyl acyl-carrier-protein reductases, whereas those with class I, III, and IV structures do not exhibit these EC numbers. However, there are a few exceptions; e.g., Rossmann-fold enzymes with EC number 1.1.1.1 exhibit three classes of pyrophosphate-binding structures, viz., I, III, and VI.

NAD and NADP Pyrophosphate-Binding Sequence Motifs. Sequences comprising each pyrophosphate-binding structural class whose structures exhibit pairwise C^{α} RMSD $\leq$ 1.0 Å and RMSDa $\leq 30°$ were aligned provided they exceed 25 (see above and Table 2). The NAD pyrophosphate-binding sequences (Figure 3) were aligned separately from the respective NADP pyrophosphate-binding sequences (Figure 4). The multiple sequence alignment results for NAD- and NADP-binding domains are represented as sequence logos[36,37] in Figures 3 and 4, respectively, where the residues at each position are ordered from most to least frequent from the top to bottom. At a given position, the height of each residue is the probability of observing that residue; hence, the overall height of each stack of letters has unit probability. Glycine is shown in green, polar residues (S, T, Y, N, Q, H, K, R, D, E) in blue, and nonpolar residues (A, V, L, I, P, W, F, C, M) in black. At a given position, if the frequency of a residue is >0.90, the residue (highlighted in bold underlined) is deemed to be conserved; alternatively, if the combined frequencies of residues with common physicochemical properties according to Taylor,[38] e.g., tiny residues (G, A, S, C, T) or aromatic residues (W, H, Y, F), exceeds 0.95, the group of residues (denoted in square brackets) is considered to be conserved. The frequencies of the conserved residue(s) are listed in Supporting Information Table S3. The consensus sequences derived from alignment of the sequences corresponding to the class I, III, IV and XII pyrophosphate-binding structures of NAD-binding domains (Figure 3) and class I, II, III, and IV pyrophosphate-binding structures of NADP-binding domains (Figure 4) are listed in Table 3. These show that the highly conserved residues (in bold underlined) comprising the pyrophosphate-binding structures are not positively charged, but involve at least one glycine.

Differences between NAD and NADP Pyrophosphate-Binding Sequence Motifs. Comparison of the sequence logos for the common structural classes (I, III and IV) in Figures 3 and 4 show that even though NAD- and NADP-binding proteins share the same locally conserved pyrophosphate-binding structure, they do not necessarily share the same sequence motif. For the class I pyrophosphate-binding structure, although the number of NADP-binding proteins (81) is less than half that of NAD-binding proteins (222), certain conserved residues in NAD-binding proteins are found to be less conserved in NADP-binding proteins: The second and third conserved glycines in NAD-binding domains (Figure 3A) can be replaced by other tiny residues such as Ala or Ser in the NADP-binding proteins (Figure 4A). The last residue of the class III structure in NAD-binding proteins is generally nonpolar (Leu, Ile, Ala, Phe, or Val, Figure 3B), but is quite variable in

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644–5652

Table 3. Consensus NAD(P)-Binding Sequences Corresponding to Distinct Pyrophosphate-Binding Structures and Their Statistical Significance

		consensus phosphate-binding sequence	% of random hits	
structure class	#[a]	NAD-binding proteins	reverse[b]	shuffled[c]
I	222	[VILCAF]-X_3-**G**-X-**G**-X-[IVAMLF]-**G**-X_6-[ALICVFMW]	1.2	1.2
III	78	[VILWF]-X-[VIL]-X-**G**-X_2-**G**-X_2-[GA]-X_6-[LIAFV]	0.80	0.76
IV	45	[AVI]-[LVIFA]-[IV]-**T**-**G**-[GAS]-X_2-**G**-X-**G**-X_6-[LFA]	0.00018	0.00018
XII	27	[LFV]-[VI]-X-**G**-[IVL]-X_4-[SG]-X-[AG]-[WFY]-X-[IV]-**A**	0.0021	0.0056
		NADP-binding proteins		
I	81	[VILF]-X-**G**-X-[GSA]-X_2-[GAS]-X_6-[LAIFWCG]	14	14
II	52	[VICL]-X-[IVC]-X-**G**-**G**-X_2-[VIL]-[YFA]-X_2-[AFMCLV]-[LMIVF]	0.038	0.021
III	83	[VIALF]-X-[VIFL]-X-[GA]-X_2-**G**-X_2-**G**	3.7	3.4
IV	60	[AVIC]-[LIV]-[VIL]-**T**-**G**-[AGSC]-X_3-[ILF]-**G**-X_6-[LFY]	0.0015	0.0009
IV	60	[AVIC]-[LIV]-[VIL]-**T**-**G**-[AGSC]-X_2-[GR]-[ILF]-**G**-X_6-[LFY]	0	0.00018

[a]The total number of NAD/NADP-binding domains with $\beta\alpha$ structures of class x. [b]The number of matches in 535 248 randomized sequences generated by taking the reverse sequence in the Swiss-Prot database. [c]The number of matches in 535 248 randomized sequences generated by shuffling residues in windows of 20 residues in Swiss-Prot sequences.

the NADP-binding proteins (Figure 4C). For the class IV structure, the second invariant Gly in NAD-binding domains (Figure 3C) is often replaced by Arg in NADP-binding domains (Figure 4D), which is found near the terminal phosphate attached to the A-ribose. This suggests an alternative consensus sequence for the class IV NADP pyrophosphate-binding structure (Figure 4D); viz., [AVIC]-[LIV]-[VIL]-**T**-**G**-[AGSC]-X_2-[GR]-[ILF]-**G**-X_6-[LFY]. Furthermore, despite the fewer NAD-binding proteins compared to NADP-binding proteins for the class IV structure, the residue after the second conserved Gly in NADP-binding domains (Figure 4D) is generally nonpolar (Ile, Leu, or Phe), whereas both polar and nonpolar residues are found at this position in NAD-binding domains (Figure 3C).

Statistical Significance of the NAD and NADP Pyrophosphate-Binding Sequence Motifs. To determine whether a given NAD(P) sequence motif corresponding to a locally conserved pyrophosphate-binding 3D structure is statistically significant, it was used to search for similar sequences in two sets of randomized sequences using the ScanProsite tool.[39] Each sequence in the UniProtKB/Swiss-Prot database[1] was randomized by (i) taking the reverse sequence or (ii) shuffling the residues in windows of 20 residues. The results in Table 3 show that the 1D motifs appear statistically significant: The average percentage of matches in the two sets of randomized sequences for the consensus NAD(P)-binding sequences is $\lesssim$1.2%, except for the NADP-binding consensus sequences corresponding to structural class III ($\sim$4%) and class I (14%). The latter has the most number of random hits, whereas the class IV NAD/NADP-binding consensus sequences have the least. Among the two consensus sequences for the class IV NADP pyrophosphate-binding structure, the [AVIC]-[LIV]-[VIL]-**T**-**G**-[AGSC]-X_2-[GR]-[ILF]-**G**-X_6-[LFY] sequence has less random matches than that without the second conserved glycine, [AVIC]-[LIV]-[VIL]-**T**-**G**-[AGSC]-X_3-[ILF]-**G**-X_6-[LFY].

DISCUSSION

A 1D motif usually comprises conserved essential residues involved in catalysis, ligand binding, or maintaining a specific structure. In proteins sharing high sequence identity, 1D motifs can be easily identified by multiple sequence alignment of homologous sequences to reveal highly conserved residues. In proteins with low sequence identity, 1D motifs cannot be easily detected because it is difficult to (1) identify protein sequences suspected to contain the motif, and (2) align sequences with little sequence identity to spot the conserved residues.[9] These two problems are overcome herein: Since the conformations of essential aa residues are usually conserved, the first problem is overcome by choosing protein sequences that comprise a locally conserved 3D structure involved in a given function. The second problem is overcome by trimming the protein sequences comprising the 3D functional motif to the same length, so there is no sequence alignment problem even though the proteins may share low overall sequence identity.

Using pairwise C^α RMSD, RMSDa (eqs 1 and 2), and ϕ and ψ angle differences of each residue comprising the 3D motif, $\sim$74% of the NAD(P)-binding domains in the current PDB, belonging to nine superfamilies (nine different CATH codes, Table 2), can be grouped into 12 distinct pyrophosphate-binding structures (Figure 2). Even though these NAD(P)-binding domains share low overall sequence identity, 1D pyrophosphate-binding motifs were found by aligning the same-length NAD/NADP-binding sequences comprising a distinct pyrophosphate-binding 3D structure (Figures 3 and 4). Note that truncation of the $\beta\alpha$ segments to the same length of 12 structural letters or 16 residues would not affect the classification of the pyrophosphate-binding structures/1D motifs as the trimmed $\beta\alpha$ structures all contact the NAD(P) phosphate moiety, as shown in Figure 2. The importance of deriving sequence motifs/logos from local similarity of not only backbone structures, but also side chain orientations is exemplified by the class III and IV pyrophosphate-binding structures, which exhibit similar backbones but different side chain orientations (Figure 2C,D) and different spacing between the conserved glycines, i.e., **G**-X-X-**G**-X-X-**G** in the class III structure, but **G**-X-X-X-**G**-X-**G** in the class IV structure. The large $\Delta\phi$ and $\Delta\psi$ differences of certain residues account for the different side chain orientations, which yield different sequence motifs. Thus, methods employing only a *single* similarity measure such as backbone RMSD values would wrongly group these two classes of pyrophosphate-binding structures together, so aligning their sequences would result in incorrect sequence preference.

Comparison with Previous Sequence Motifs in NAD(P)-Binding Proteins. The phosphate-binding **G**-X_{1-2}-**G**-X-X-**G** motif in the short loop connecting the first β-strand to the first α helix was first found by Rossmann et al.[15,40] from an alignment of the sequences of dogfish lactate dehydrogenase; pig, lobster, and yeast glyceraldehyde-3-phosphate dehydrogenase; horse liver alcohol dehydrogenase; and bovine glutamate dehydrogenase. Here, we find that the **G**-X-**G**-X-X-**G** and **G**-X-X-**G**-X-X-**G** motifs correspond to two distinct structures: class I and class III, respectively. Hence, different positioning of the conserved glycines yields different pyrophosphate-binding structures. Furthermore, we find that the sequence motif corresponding

dx.doi.org/10.1021/jp3014332 | J. Phys. Chem. B 2012, 116, 5644−5652

to a given class of pyrophosphate-binding structure depends on the type of cofactor. For the class I pyrophosphate-binding structure, Ala often replaces the third Gly in NADP-binding domains (compare Figures 4A and 3A). As the third conserved Gly is thought to enable close packing of the first β-strand and the first α-helix, the larger Ala may disrupt this close packing, thus enabling the structure to accommodate the additional NADP terminal phosphate.[14]

By analyzing helix−β-strand interactions in Rossmann-fold proteins, Kleiger and Eisenberg[17] proposed an extended [V/I]-X-**G**-X_{1-2}-**G**-X-X-**G**-X-X-X-[G/A] sequence motif as an indicator of Rossmann folds that bind FAD or NAD(P). For the **G**-X-**G**-X-X-**G** motif derived from class I structures in Figures 3A and 4A and the **G**-X-X-**G**-X-X-**G** motif derived from class III structures in Figures 3B and 4C, Val or Ile is indeed often found two residues before the first conserved Gly; its position corresponds to the conserved structural letter *e* or *f* in the structural sequences in Table 1. However, the fourth residue after the third conserved Gly in the sequences belonging to structural classes I and III is *not* a conserved Gly or Ala, but is variable. Brakoulias and Jackson[18] used geometric matching to cluster phosphate-binding sites of Rossmann-fold NAD(P)-binding proteins with similar 3D structure, and found **G**-X-X-X-**G**-**I**-**G**, **G**-X-**G**-X-**V**-**G**, and **G**-X-**G**-X-X-**G** motifs in four clusters of these proteins. The **G**-X-X-X-**G**-**I**-**G** and **G**-X-**G**-X-**V**-**G** motifs are subsets of the class IV and class I NAD-binding consensus sequences (Figure 3), whereas the **G**-X-**G**-X-X-**G** motif is a subset of the class I NADP-binding consensus sequence (Figure 4A).

Relationship between 3D and 1D Motifs. Previous works have proposed the following roles for the conserved glycines in the G-X_{1-2}-G-X-X-G motif in Rossmann-fold proteins: The first Gly permits a tight turn of main chain, which is important for positioning the second Gly; the missing side chain of the second Gly allows for close contact of the main chain to the NAD(P) pyrophosphate, while the third Gly enables the close packing of the first β-strand and the first α-helix.[14,41] However, the results herein show that the second and third glycines are not necessarily conserved in the different classes of pyrophosphate-binding structures. For example, sequences derived from the class XII pyrophosphate-binding structure of Rossmann-fold proteins exhibit only one conserved glycine (see Figure 3D), while sequences derived from the class I pyrophosphate-binding structure of Rossmann-fold NADP-binding domains often have small Ala or Ser residues in lieu of the second and third glycines (Figure 4A). On the other hand, sequences derived from *non*-Rossmann-fold NADP-binding domains with the class II structure exhibit two successive conserved Gly (Figure 4B).

The findings herein provide physical-chemical insight as to how NAD/NADP-binding proteins that share low sequence identity bind the NAD(P) phosphate moiety. That none of the conserved residues comprising the sequence motifs derived from the distinct 3D functional motifs are positively charged implies that the locally conserved $\beta\alpha$ structures present in NAD(P)-binding proteins bind the phosphate via favorable interactions with the α-helix dipole.[33] A novel finding is that the sequence motifs in Figures 3 and 4 have in common a highly conserved Gly. Interestingly, in the class I and II pyrophosphate-binding structures, the $\varphi-\psi$ conformations of this highly conserved Gly prevent mutation to other residues: in three exceptions where this Gly is replaced by Ala in NAD-binding proteins (2gsd-A, 2ixa-A, and 2nad-A), the backbone φ and ψ angles of Ala are *not* in the generally allowed region in the Ramachandran plot according to the Molprobity program.[42] In the class I, III, IV, and XII pyrophosphate-binding structures, this Gly is in close contact with the NAD(P) A-ribose, suggesting that these distinct locally conserved structures may help to bind the negatively charged NAD(P) pyrophosphate and A-ribose, which in turn enables protein recognition of the adenine. Directions for future works include (i) elucidating the roles of the conserved residues found in each pyrophosphate-binding structural class, and (ii) determining locally unstable regions that may be involved in binding the pyrophosphate in the NAD(P)-binding domains[43] that do not possess pyrophosphate-binding 3D motifs and identifying their conserved residues, if present.

CONCLUSION

In summary, this work has shown that many different 3D structures are associated with the same function (binding the NAD(P) pyrophosphate), and they differ in the number/spacing of conserved glycines. It has underscored the importance of carefully determining all distinct 3D motifs associated with a given function according to both their backbone conformations and side chain orientations. As more structures are solved in the future, our strategy would yield more accurate pyrophosphate-binding sequence motifs, which, in conjunction with other sequence-based methods for functional residue annotation, could enable functional annotation of uncharacterized proteins. Our strategy can also be applied to unravel 1D motifs in other protein superfamilies that contain 3D phosphate-binding motifs such as proteins that bind organic cofactors such as FAD, ATP, ADP, GTP, and GDP.

ASSOCIATED CONTENT

Supporting Information

PDB entries, C^{α} RMSD and RMSDa values, and frequencies. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Address: Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan. FAX: 886-2-2788-7641. E-mail: carmay@gate.sinica.edu.tw.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank C. Satheesan Babu and Yu-Ming Lee for helpful discussions. This work was supported by the National Science Council, Taiwan [NSC 95-2113-M-001-038-MY5] and Academia Sinica.

REFERENCES

(1) Magrane, M.; UniProt Consortium. *Database* **2011**, bar009.
(2) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; et al. *Acta Crystallogr. D* **2002**, *58*, 899−907.
(3) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(4) Watson, J. D.; Laskowski, R. A.; Thornton, J. M. *Curr. Opin. Struct. Biol.* **2005**, *15*, 275−284.
(5) Sigrist, C. J.; Cerutti, L.; Hulo, N.; Gattiker, A.; Falquet, L.; Pagni, M.; Bairoch, A.; Bucher, P. *Briefings Bioinf.* **2002**, *3*, 265−274.
(6) Mathura, V. S.; Schein, C. H.; Braun, W. *Bioinformatics* **2003**, *19*, 1381−1390.

(7) Hulo, N.; Bairoch, A.; Bulliard, V.; Cerutti, L.; De Castro, E.; Langendijk-Genevaux, P. S.; Pagni, M.; Sigrist, C. J. A. *Nucleic Acids Res.* **2006**, *34*, D227−D230.
(8) Wass, M. N.; Sternberg, M. J. E. *Bioinformatics* **2008**, *24*, 798−806.
(9) Bailey, T. *Methods Mol. Biol.* **2008**, *452*, 231−251.
(10) Wu, C. Y.; Chen, Y. C.; Lim, C. *Nucleic Acids Res.* **2010**, *38*, e150.
(11) Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.* **1973**, *76*, 241−256.
(12) Kuppuraj, G.; Sargsyan, K.; Hua, Y.-H.; Merrill, A. R.; Lim, C. *J. Phys. Chem. B* **2011**, *115*, 7932−7939.
(13) Schulz, G. E. *Curr. Opin. Struct. Biol.* **1992**, *2*, 61−67.
(14) Bellamacina, C. R. *FASEB J.* **1996**, *10*, 1257−1269.
(15) Rossmann, M. G.; Liljas, A.; Branden, C. I.; Banaszak, L. T. *Enzymes* **1975**, *11*, 61−102.
(16) Dym, O.; Eisenberg, D. *Protein Sci.* **2001**, *10*, 1712−1728.
(17) Kleiger, G.; Eisenberg, D. *J. Mol. Biol.* **2002**, *323*, 69−76.
(18) Brakoulias, A.; Jackson, R. M. *Proteins: Struct, Funct. Bioinf.* **2004**, *56*, 250−260.
(19) Ausiello, G.; Gherardini, P. F.; Gatti, E.; Incani, O.; Helmer-Citterich, M. *BMC Bioinf.* **2009**, *10*, 182.
(20) Gherardini, P. F.; Ausiello, G.; Russell, R. B.; Helmer-Citterich, M. *Nucleic Acids Res.* **2010**, *38*, 3809−3816.
(21) Artymiuk, P. J.; Poirrette, A. R.; Grindley, H. M.; Rice, D. W.; Willett, P. *J. Mol. Biol.* **1994**, *243*, 327−344.
(22) Wallace, A. C.; Borkakoti, N.; Thornton, J. M. *Protein Sci.* **1997**, *6*, 2308−2323.
(23) Kleywegt, G. J. *J. Mol. Biol.* **1999**, *285*, 1887−1897.
(24) Meng, E. C.; Polacco, B. J.; Babbitt, P. C. *Proteins: Struct., Funct. Bioinf.* **2004**, *55*, 962−976.
(25) Torrance, J. W.; Bartlett, G. J.; Porter, C. T.; Thornton, J. M. *J. Mol. Biol.* **2005**, *347*, 565−581.
(26) Dudev, M.; Lim, C. *BMC Bioinf.* **2007**, *8*, 106−118.
(27) Kristensen, D. M.; Ward, R. M.; Lisewski, A. M.; Erdin, S.; Chen, B. Y.; Fofanov, V. Y.; Kimmel, M.; Kavraki, L. E.; Lichtarge, O. *BMC Bioinf.* **2008**, *9*, 17.
(28) Sadowski, M. I.; Jones, D. T. *Curr. Opin. Struct. Biol.* **2009**, *19*, 357−362.
(29) Erdin, S.; Lisewski, A. M.; Lichtarge, O. *Curr. Opin. Struct. Biol.* **2011**, *21*, 180−188.
(30) Pearl, F. M. G.; Bennett, C. F.; Bray, J. E.; Harrison, A. P.; Martin, N.; Shepherd, A.; Sillitoe, I.; Thornton, J.; Orengo, C. A. *Nucleic Acids Res.* **2003**, *31*, 452−455.
(31) de Brevern, A. G.; Etchebest, C.; Hazout, S. *Proteins: Struct., Funct. Genet.* **2000**, *41*, 271−287.
(32) Laskowski, R. A. *Nucleic Acids Res.* **2009**, *37*, D355−D359.
(33) Hol, W. G. J.; Van Duijnen, P. T.; Berendsen, H. J. C. *Nature* **1978**, *273*, 443−446.
(34) DeLano, W. L. *The PyMOL Molecular Graphics System*; 1.2r3pre ed.; Schrodinger, LLC: New York, 2008.
(35) Bairoch, A. *Nucleic Acids Res.* **2000**, *28*, 304−305.
(36) Schneider, T. D.; Stephens, R. M. *Nucleic Acids Res.* **1990**, *18*, 6097−6100.
(37) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188−1190.
(38) Taylor, W. *J. Theor. Biol.* **1986**, *119*, 205−218.
(39) de Castro, E.; Sigrist, C. J. A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P. S.; Gasteiger, E.; Bairoch, A.; Hulo, N. *Nucleic Acids Res.* **2006**, *34*, W362.
(40) Rossmann, M. G.; Moras, D.; Olsen, K. W. *Nature* **1974**, *250*, 194−199.
(41) Wierenga, R. K.; DeMaeyer, M. C. H.; Hol, W. G. J. *Biochemistry* **1985**, *24*, 1346−1357.
(42) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D: Biol. Crystallogr.* **2010**, *66*, 12−21.
(43) Chen, Y. C.; Lim, C. *Nucleic Acids Res.* **2008**, *36*, 7078−7087.