PROTEINS: Structure, Function, and Genetics 38:95–114 (2000)

New Sequence Motifs in Flavoproteins: Evidence for Common Ancestry and Tools to Predict Structure

Olivier Vallon*
Institut de Biologie Physico-Chimique, CNRS, Paris, France

ABSTRACT **We describe two new sequence motifs, present in several families of flavoproteins. The "GG motif" (RxGGRxxS/T) is found shortly after the βαβdinucleotide-binding motif (DBM) in L-amino acid oxidases, achacin and aplysianin-A, monoamine oxidases, corticosteroid-binding proteins, and tryptophan 2-monooxygenases. Other disperse sequence similarities between these families suggest a common origin. A GG motif is also found in protoporphyrinogen oxidase and carotenoid desaturases and, reduced to the central GG doublet, in the THI4 protein, dTDP-4-dehydrorhamnose reductase, soluble fumarate reductase, steroid dehydrogenases, Rab GDP-dissociation inhibitor, and in most flavoproteins with two dinucleotide-binding domains (glutathione reductase, glutamate synthase, flavin-containing monooxygenase, trimethylamine dehydrogenase . . .). In the latter families, an "ATG motif" (oxhhhATG) is found in both the FAD- and NAD(P)H-binding domains, forming the fourth β-strand of the Rossman fold and the connecting loop. On the basis of these and previously described motifs, we present a classification of dinucleotide-binding proteins that could also serve as an evolutionary scheme. Like the DBM, the ATG motif appears to predate the divergence of NAD(P)H- and FAD-binding proteins. We propose that flavoproteins have evolved from a well-differentiated NAD(P)H-binding protein. The bulk of the substrate-binding domain was formed by an insertion after the fourth β-strand, either of a closely related NAD(P)H-binding domain or of a domain of completely different origin. Proteins 2000;38:95–114.**

Key words: FAD; NAD; NADP; dinucleotide; oxidase; reductase; cofactor binding; Rossman fold

INTRODUCTION

Sequence motifs are those regions of a protein where variability is limited, usually because of structural constraints linked either to folding or to interactions with cofactors, substrates, or other proteins. Much effort is devoted to identifying motifs conserved between distantly related protein families: they can serve as "anchoring sites" for predicting the fold of a polypeptide chain, provided the structure of a distant relative is available. One of the earliest and most fruitful examples of such correlation efforts was the discovery by Rossman and collaborators[1] of a conserved sequence motif, reflecting a common chain fold for several dehydrogenases that use NADH as a cofactor. The basic structure consists of a parallel β-sheet, made up of six strands connected in a right-handed fashion by α-helices that flank the sheet on both sides. The structure has overall twofold rotational symmetry, with the first half of the sheet more specifically involved in binding the adenine moiety, the second the nicotinamide moiety. The nucleotide binds near the C-terminal end of the strands. When the central sheet is viewed with the C-terminal ends of the strands pointing upward and the first helix in foreground (as in Fig. 5), the strands, reading left to right, are in the order 6, 5, 4, 1, 2, and 3. The nicotinamide points slightly to the front, the adenine to the back.

A slightly modified version of this fold was subsequently identified in the FMN-binding protein flavodoxin[2] and in NADPH- and FAD-binding proteins.[3] Structural information on the latter type has come mostly from the analysis of the glutathione reductase (GR)-related family of flavoprotein pyridine nucleotide reductases, which possess one FAD- and one NAD(P)H-binding domain.[4] In this category of proteins, the basic fold of each domain differs slightly from that of dehydrogenases, because the third and fourth strand of the parallel β-sheet are connected not by a helix but by an antiparallel β-sheet (see Schulz[5]) that covers one side of the central sheet. In the NAD(P)H-binding domain of these proteins, the fifth and sixth strands of the central β-sheet are missing. In the FAD-binding domain, the fifth strand is retained but close to the end of the sequence, i.e., after the NAD(P)H-binding domain.

Most of the dinucleotide-binding proteins that adopt the Rossman fold show a series of conserved amino acid residues at a few crucial positions, whereas other regions of the sequence can considerably vary. Alignments of

Abbreviations: CBP, corticosteroid-binding protein; COX, cholesterol oxidase; CMO, cyclic monooxygenases; DAO, D-amino acid oxidase; DBM, dinucleotide-binding motif; FCSD, flavocytochrome c-sulfide dehydrogenase; FMO, flavin-containing monooxygenase = dimethylaniline monooxygenase; GOX, glucose oxidase; GR, glutathione reductase; LAO, L-amino acid oxidase; MAO, monoamine oxidase; PHH, *p*-hydroxybenzoate hydroxylase; Rab-GDI, Rab GDP dissociation inhibitor; SFR, soluble fumarate reductase; TMO, tryptophan 2-mono-oxygenase; TMD, trimethylamine dehydrogenase.

Grant sponsor: Centre National de la Recherche Scientifique; Grant number: UPR1261.

*Correspondence to: Olivier Vallon, Institut de Biologie Physico-Chimique, UPR 9072 CNRS, 13 rue Pierre et Marie Curie, F-75005 Paris, France. E-mail: vallon@ibpc.fr

Received 1 December 1998; Accepted 19 August 1999

protein sequences have been proposed on the basis of pairwise and multiple structural alignments[1,3,6] and have been used to generate a sequence consensus for the central part of the fold. This motif, presented in its most elaborate form by Wierenga et al.,[7] applies to NADH- as well as FAD-binding domains and in a modified form to NADPH-binding domains. In the following, this sequence signature will be referred to as the dinucleotide-binding motif (DBM). The basic DBM consensus can be written:

```
ohxhxGxGxxGxxxhxxhxxxx....hxhxD/E
eeeeee hhhhhhhhhhhhhh    eeeee
```

where o stands for a polar or charged residue and h for a hydrophobic residue. The sequence constitutes the "βαβ " motif: the first two β-strands and the intervening amphipathic α-helix (as indicated by *e* and *h*, respectively, underneath). The conserved central GxGxxG sequence allows close approach of the polypeptide backbone to the ADP ribose and pyrophosphate and close packing of the helix with the β-sheet. The second and third Gs can be replaced by A in NADH or NADPH binding sites.[8] The hydrophobic residues provide hydrophobic interactions between the α-helix and the β-sheet. The terminal residue (generally an E in flavoproteins, a D in NADH-binding proteins) hydrogen bonds to the ribose of the adenine moiety. At that position, a noncharged residue is usually found in NADPH-binding sites, to accommodate the phosphate group of the cofactor (see, however, Baker et al.[9]), followed by a positively charged residue. The importance of these and another residue downstream for NADH/NADPH discrimination has been established by site-directed mutagenesis.[10]

In addition to this motif, a second fingerprint has been identified by Eggink et al. [11] in FAD-binding proteins of the GR family. The characteristic sequence is TxxxxhϕhhGD, where ϕ is an aromatic residue. The hydrophobic residues belong to the fifth β-strand of the FAD-binding domain, found near the C-terminus of the protein. The invariant G provides for a turn at the end of the β-strand, and the terminal D H-bonds to the O3* (and sometimes O4*) of the ribityl in the flavin moiety of the cofactor. The initial T participates in the formation of a "greek key" motif found just before the fifth strand.

The Rossman fold is the prevalent, although not unique[12,13] way adopted by proteins to bind dinucleotides. In the case of flavoproteins, numerous families have been described, and in most cases a classical DBM can be found in the N-terminal part of the sequence. However, outside of this region, sequence comparison programs in general have failed to identify sequence similarities between these families. The three dimensional (3-D) structure of several flavoproteins is known at atomic resolution, and their folding pattern appears much more similar than would have been suggested by comparison of their sequences. The question of whether all these flavoproteins have a common origin or have convergently acquired their common structural similarities and DBM sequence has hitherto remained largely unanswered.

In a previous study, we presented the cloning and sequencing of a complementary deoxyribonucleic acid (cDNA) for a flavoprotein, the catalytic subunit of the periplasmic L-amino acid oxidase (LAO) of the green alga *Chlamydomonas reinhardtii*.[14] Sequence comparison has revealed the presence of a short "GG motif," found in LAOs and in a wide variety of other flavoprotein families. In addition, we identified another, hitherto overlooked sequence signature, the "ATG motif," present in a large collection of flavoproteins and in many NAD(P)H-binding proteins. Finally, we systematically examined the presence in flavoproteins of the fingerprint described by Eggink et al.[11] Our analysis lends support to the hypothesis of a common evolutionary origin for these flavoprotein families, rooted among dehydrogenases.

MATERIALS AND METHODS

Sequences and 3-D structures were retrieved from databases by using a web interface (www.infobiogen.fr/srs5). Homology searches were performed with the BLASTP or PSI-BLAST programs[15] on the NCBI server. Multiple alignments were generated with MACAW[16] or CLUSTALW[17] (www.genome.ad.jp/SIT/CLUSTALW.html) or were retrieved from the PUMA database (www-c.mcs.anl.gov/home/compbio/PUMA/). Secondary structure predictions were generated with PSIpred (http://globin.bio.warwick.ac.uk/psipred/). 3-D structures were visualized with RASWIN.[18] Interatomic distances and torsion angles were computed with the CCP4 program suite.[19]

RESULTS

GG Motif

Besides the classical DBM, the *C. reinhardtii* LAO shows only weak sequence similarity to the other LAO sequences, and we propose to classify it in a family different from the other LAOs. However, all LAOs share the sequence RxGGRxx(S/T) (hereafter referred to as the GG motif), located exactly four residues downstream of the DBM. This sequence was also found, at the same distance from the DBM, in many other families of flavoproteins (listed in Table I). Suspecting that the GG motif was indicative of homology, we attempted to identify other conserved sequences between these proteins. We found that four of these families share sufficient similarity to be considered homologous, i.e., of probable common ancestry: non-algal LAOs, corticosteroid-binding proteins (CBP), monoamine oxidases (MAO), and tryptophan 2-monooxygenases (TMO). Selected regions of a MACAW alignment are presented in Figure 1 (the entire file is available from the author upon request, as are all the other alignments used in this study). The line above the human MAO-B in Figure 1 is a secondary structure prediction for this sequence, generated with the program PSIpred. Other sequences yielded similar predictions for the conserved regions (not shown). The *C. reinhardtii* LAO and the two molluscan antibacterial proteins were also included in the alignment, even though the similarity to the other protein

TABLE I. Flavoproteins With a GG Motif

Protein family:	{n} Sequence identifiers	References	GG motif	Exceptions
L-amino acid oxidase (LAO, EC 1.4.3.2)	{1} U78797 (*C. reinhardtii*)	14	RxGGRxxS	
L-amino acid oxidase (LAO, EC 1.4.3.2)[a]	{6} A38314 (*N. crassa*); O93364 (*Crotalus*); BC542A (*B. cereus*); O34363 (*B. subtilis*); Z48565 (*Synechococcus* sp. PCC6301); MMU70430 (*M. musculus* FIG1)	54–57	RxGGRxx(S/T)	
Antibacterial protein[b]	{2} ACHC_ACHFU; D83255	58, 59	RxGGRxx(S/T)	
Corticosteroid-binding protein (CBP)[c]	{2} CBP1_CANAL; FMS1_YEAST	60, 61	RxGGRxxT	
Mono-amine oxidase (MAO, EC 1.4.3.4)	{17} Vertebrate MAOs: AOFA_HUMAN; AOFB_HUMAN; AOFA_RAT; AOFA_BOVIN; S45812; X15609; AOF_ONCMY Other: Z78198 (*C. elegans*); e327500 (*A. thaliana*) AOFN_ASPNG (*A. nudans* MAO-N); PUO_MICRU (*M. rubens* putrescine oxidase); AOFH_MYCTU (*M. tuberculosis*); Y782_SYNY3 (*Synechocystis* sp.); AB010716 (*M. luteus* tyramine oxidase); AJ223391 (*A. nicotinovorans* 6-hydroxy-L-nicotine oxidase)	50, 62, 63	RxGGRxx(S/T)	AOFB_RAT : CxGGRxxT Z35602 (C. elegans R13G10.2) : RxGGRxxD Y782_SYNY3 (S. sp. PCC6803) : RxGGRxxG
Tryptophan 2-mono-oxygenase (TMO, EC 1.13.12.3)	{9} TR2M_AGRRA; S28687; TR2M_AGRT4; 499582; TR2M_AGRT3; TR2M_PSESS; A20966; A53376;	64, 65	RxGG(R/K)xxS	TR2M_AGRVI : VxGGRxxT
Carotenoid desaturase[d]	{27} −CRTD_RUBGE; CRTD_RHOCA; CRTD_RHOSH; CRTI_RHOCA; CRTI_RHOSH; PDEH_STRGR; e224562; CRTI_ERWHE; CRTI_AGRAU; X97985; S32171; S43324; CRTJ_MYXXA; CRTI_NEUCR; CRTI_NEUCR; CRTI_PHYBL; CRTI_MYXXA; −CRTI_SYNY3; CRTI_ARATH; CRTI_MAIZE; CRTI_SOYBN −S66625; ZCDS_SYNY3 −crtN: B55548	66–70	xxGG(R/K)	YZ25_MYCTU; d1011212; 2749982 : xxGGA ZCDS_ARATH : xxGAK P49_STRLI; e318945; e1237742 : xxGGG
Protoporphyrinogen oxidase (EC 1.3.3.4)[e]	{11} HEMG_BACSU; g699193; PPOX_MYCTU; SCHEM14PO; YAM7_SCHPO; PPOX_HUMAN; PPOX_MOUSE; ATHPPOX; NTY13465	71, 72	RxGGxxx(S/T)	d1022768: HxGGxxxS NTY13466: KxGGxxxS

[a]Homology between the *Synechococcus* and *Neurospora* proteins was overlooked in[55] and [14] because of sequencing errors. The Fig 1 protein has been classified as a MAO but has more similarity to LAOs.[56,57]
[b]Achacin and aplysianin-A have not been recognized previously as flavoproteins. Their antibacterial effect could be the result of H_2O_2 production.
[c]FMS1, the yeast CBP, acts as a multicopy suppressor of fenpropimorph resistance (Genbank entry), suggesting a role in sterol biosynthesis.[73]
[d]Known as phytoene, methoxyneurosporene, hydroxyneurosporene, ζ-carotene, or dehydrosqualene desaturases (dehydrogenases).
[e]The *S. cerevisiae* enzyme binds FAD covalently. It oxidizes protoporphyrinogen IX, whereas the *B. subtilis* sequence is that of a protoporphyrinogen IX/coproporphyrinogen III bispecific enzyme.

families is somewhat lower and homology is less certain. Carotenoid desaturases (not shown) show a weak similarity to these sequences in the middle (hhhxxP/S) and last (FAGE) similarity blocks (see Table V). With protoporphyrinogen oxidases, similarity was limited to the DBM and GG motif.

Expanding the Search: The "GG Doublet"

In the GG motif, the two central glycines appear as the only strictly invariant residues in the families listed in Table I. In a library of 394 sequences of nucleotide-binding proteins (mostly flavoproteins), extracted from NBRF by using various versions of the DBM consensus, we found 50 sequences with a GG doublet 6 residues after the end of the DBM. The protein families in which this GG doublet was significantly conserved are presented in Table II. A few additional observations are detailed below.

On the basis of sequence comparison, fumarate reductases can be divided into two families sharing significant similarities (Fig. 2). The GG doublet was conserved in the soluble fumarate reductases (SFR) that bind FAD noncovalently,[20] but not in the other family, comprising the soluble thiol-fumarate reductases from archaebacteria[21] and the membrane-bound succinate dehydrogenases of eubacteria and mitochondria[22] that bind FAD covalently.[23] In the latter family, the corresponding region contained the His residue that substitutes FAD (replaced by Cys in archaebacteria). However, the presence of a GG doublet was not strictly correlated with noncovalent binding of FAD: L-aspartate oxidases that can reduce fumarate and are related in sequence to succinate dehydrogenases[24] bind FAD noncovalently.[25] Still, their sequence lacks a GG doublet (Fig. 2). Several steroid dehydrogenases showing dispersed similarity to SFRs have been included in the alignment of Figure 2. Here, the GG doublet was found at a variable distance from the DBM (+5 to +8), but other regions of similarity indicate a common origin. The boxed sequences in Figure 2 are conserved only between fumarate reductases, succinate dehydrogenases, and aspartate oxidase: they may delineate the substrate-binding domain.

Also listed in Table II are several families of complex flavoproteins that share the particularity of containing two dinucleotide-binding domains, each with its DBM. In general, the first domain is involved in binding FAD, the

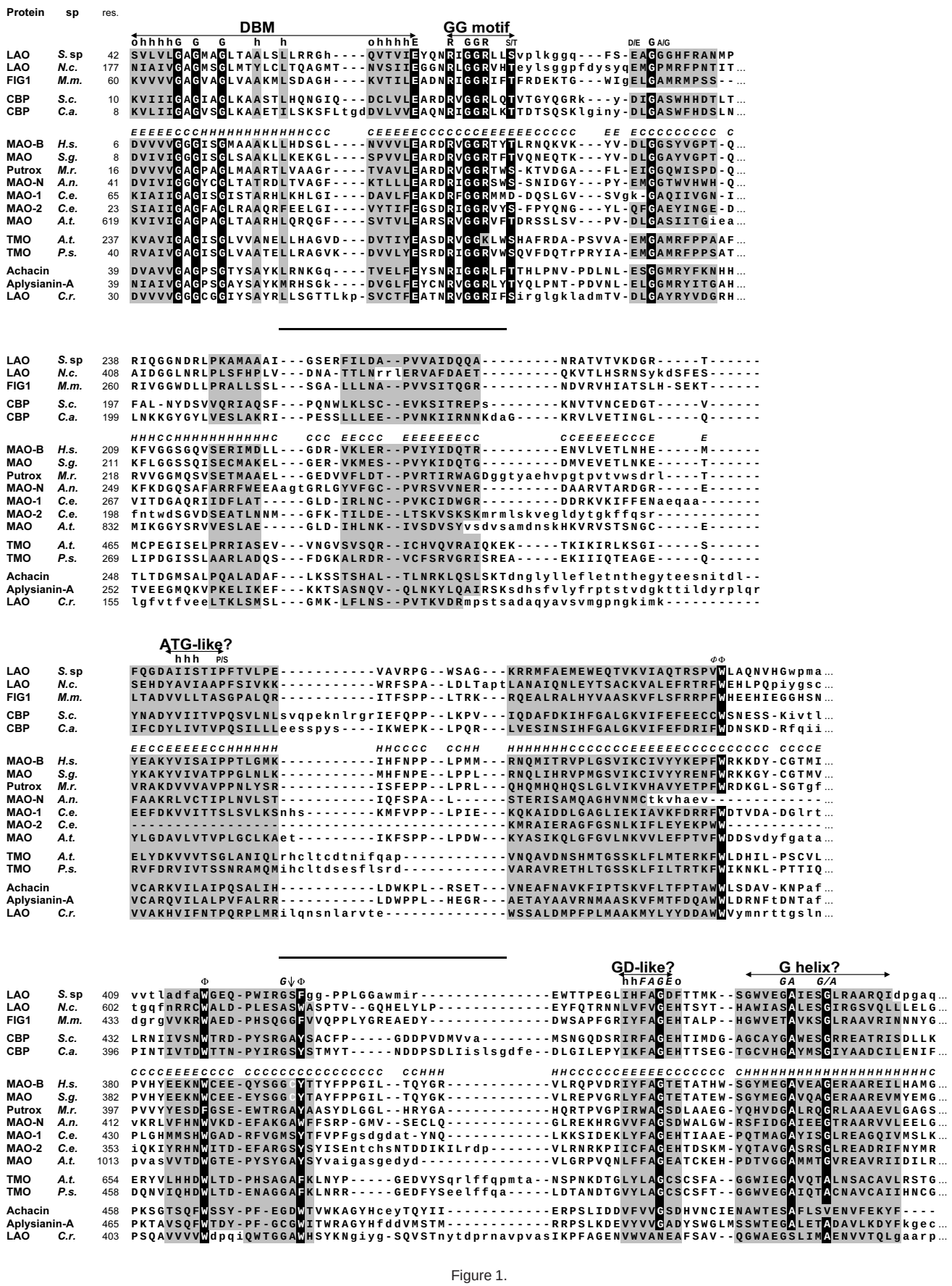

Figure 1.

second one NAD(P)H. The GG motif was always found after the first DBM (Table II). An alignment of several of these families is presented in Figure 3.

In particular, a series of proteins related to the *baiH* gene product of *Eubacterium* sp.[26] present this type of organization in their C-terminal region. The 3-D structure of a protein in this family, trimethylamine dehydrogenase (TMD)[27,28] allowed us to visualize the position of the GG doublet in the molecule (Fig. 4). In this protein, ADP, not FAD, occupies the first DBM.[29] The GG doublet (residues 425–426) is located at the base of an α-helix that points toward the second dinucleotide-binding domain. The dihedral angles of these Gly (Table III) are in a domain of the Ramachandran plot that is not allowed for other types of residue.[30] In addition, the first Gly is in contact with central residues in the DBM, whereas the second one is in close proximity to the pyrophosphate and ribose of the ADP cofactor (Table III). The packing is so dense that the introduction of side chains at these positions appears impossible without extensive rearrangement of the backbone. The next residue, H427, also plays an important structural role, because it makes main chain and side chain contacts with atom O2A of the cofactor. These strong structural constraints may explain the excellent conservation of this GG doublet in the family.

The folding pattern of the two C-terminal domains of TMD closely resembles that of the FAD and NADPH domains of GR.[27] The similarity is striking in the regions that surround the ADP/FAD cofactors. In a structure-based alignment, the GG doublet in TMD was shown perfectly superposed to two Gly residues in GR.[29] This doublet is found not six, but five residues downstream of the DBM, which explains why GR did not appear in our initial screen for proteins with a GG doublet. In GR, as in TMD, the first Gly is close to the DBM, and the second one approaches the ribose AO3* atom of the cofactor (Table III). The residue immediately following the GG doublet, T57, makes main chain atom contact with the AO2 atom of FAD (as does H427 in TMD), plus side chain contact with AO1. The following α-helix runs alongside the ribityl of the FAD and contains the two redox-active Cys. In the structure-based alignment of GR and TMD presented in Figure 3, the GG doublet appears as one of the few cases of sequence identity between the two proteins.

A number of proteins share significant sequence and structural similarity with GR and are usually grouped in the superfamily flavoprotein pyridine nucleotide reductases (a more appropriate term than the often used "flavoprotein disulfide reductases"). The GG doublet was perfectly conserved in these proteins (Table II), which is remarkable in view of their limited sequence similarity. The spacing between the end of the DBM and the GG doublet varied considerably between different families, and even within a family. Still, in the Families of Structurally Similar Proteins (FSSP) database,[31] where a given structure is structurally aligned with all other structures in PDB, we found that nearly all the flavoprotein pyridine nucleotide reductases had a GG doublet superposed to that of GR (Table II). The only exceptions were NADH peroxidase and NADH oxidase of *Enterococcus faecalis*.[32] In these proteins, the α-helix that follows the DBM starts farther away from the cofactor and contains modified Cys residues.[32] In flavocytochrome *c*-sulfide dehydrogenase (FCSD) of *Chromatium vinosum*, which does not bind pyridine nucleotides but is structurally related to GR,[33,34] the GG doublet was also missing, probably correlated with the covalent binding of FAD via Cys residues in the α-helix following strand 2.

When we examined the available 3-D structures of flavoproteins with a single dinucleotide-binding domain (5 families), Rab GDP dissociation inhibitor (Rab-GDI)[35] was the only one with a bona fide GG doublet. Although this protein does not bind any dinucleotide cofactor, a sequence resembling a DBM can be found near its N-terminus,[36] and its 3-D structure is strikingly similar to that of a flavoprotein.[37] In particular, a stretch of 12 residues (D-34 to S-45), including the GG doublet, was easily superimposable with the GR structure. Another interesting case was that of cholesterol oxidase (COX)[38] and glucose oxidase (GOX).[39] With use of the FSSP database, these proteins were found to contain two consecutive Gly residues that could be superimposed with the GG doublet of GR. Surprisingly, this sequence was found at +72 and +51, respectively, after the DBM, i.e., at a position completely different from the usual GG doublet. Still, the orientation, dihedral angles, and distance to the DBM are comparable with those in TMD and GR (Table III). As in GR and TMD, the GG doublet is close to the FAD, and the residue immediately following makes contact with the pyrophosphate (OA1). Based on their similar 3-D structure, COX and GOX have been grouped in a superfamily of flavoproteins, called GMC oxidoreductases.[40,41] We have found perfect conservation of these two Gly residues in the GMC superfamily (not shown).

A Novel Conserved Sequence in GR-Related Proteins: The "ATG Motif"

We wondered whether other discrete conserved sequence motifs could be identified in flavoproteins, in addition to the DBM and GG motifs. We initiated our

Fig. 1. Three regions of a MACAW sequence alignment of MAOs and related sequences. The italicized line above the MAO-B sequence is the secondary structure prediction generated by PSIpred (C for random coil, E for β-strand, H for α-helix). Residues that can be aligned with sequences from another family appear in upper case, with gray shading denoting the best conserved regions. Conserved residues appear on black background and are identified on the upper line (in italics when conservation is not absolute; h stands for hydrophobic, o for polar Φ for aromatic). The arrow points to the Cys residue that covalently binds the flavin in vertebrate MAOs.[50] **LAO** from *Chlamydomonas reinhardtii* (U78797), *Neurospora crassa* (A38314), and *Synechococcus* sp. PCC6301 (Z48565); **CBP** from *Saccharomyces cerevisiae* (FMS1_YEAST) and *Candida albicans* (CBP1_CANAL); **MAO** from *Homo sapiens* (AOFB_HUMAN), *Salmo garnieri* (AOF_ONCMY) and *Aspergillus niger* (S55273), and related sequences from *Coenorhabditis elegans* (R13G10.2 and F55C5.6) and *Arabidopsis thaliana* (ATFCA5: protein e327500); **Putrox**: putrescine oxidase from *Micrococcus rubens* (PUO_MICRU); **FIG1** from *Mus musculus* (U70429); **TMO** from *Agrobacterium tumefaciens* (A20966) and *Pseudomonas syringeae* (TR2M_PSESS); Achacin from *Achatina fulica* (ACHC_ACHFU) and Aplysianin-A from *Aplysia kurodai* (D83255).

TABLE II. Flavoproteins With a GG Doublet†

Protein family	{n} Sequence identifiers	References	GG	Exceptions
I) One dinucleotide-binding domain:		74–77	+6	
THI4 enzyme[a]	{9} THI4_YEAST; THI4_SCHPO; THI4_ARATH; TH41_MAIZE; TH42_MAIZE; THI4_ALNGL; THI4_FUSOX; THI4_FUSSH; THI4_METJA			
dTDP-4-dehydrorhamnose reductase[b] (EC 5.4.99.9)	{5} GLF_MYCGE; GLF_MYCPN; GLF8_KLEPN; GLF1_KLEPN; GLF_ECOLI	78, 79	+6	
Soluble fumarate reductase (SFR)[c]	{9} e257742 (Z78020); OSM1_YEAST; YEF7_YEAST; JC5123; CAB16560; CE01D10; FRDA_SHEPU; CAB38558.1; CAB37062	20, 22, 80	+6	
3-Oxosteroid 1-dehydrogenases	{4} 3O1D_COMTE; e280795 (Z82098); U59422; D37969	81	+6	e266575: GG at +8
δ4,5-alpha steroid dehydrogenase	{1} L23428	82	+5	
Rab GDP dissociation inhibitor (Rab-GDI)[d]	{21} GDIA_BOVIN; GDIA_RAT; C56956; GDIA_HUMAN; GDIB_RAT; GDIB_MOUSE; GDIB_HUMAN; CELRABGDI; S36746; X93166; U62866; X94983; D83531; AF012823; AF016896; AF016897; RAE1_HUMAN; RAE2_HUMAN; RAE1_RAT	35–37	+6	RAEP_YEAST : GD YD4C_SCHPO : GA
II) Two dinucleotide-binding domains:		83–88	+6	
Glutamate synthases (GOGAT)[e]	{6} A46602; B29617; Z83864; X89221; Z49889; JQ1977;			
GOGAT-like[f]	{5} U20981; ECAE000333 (yffG); ECAE000372 (f644); AE000303 (o462); U73807	83–88	+6	
Dimethylaniline monooxygenase (FMO)[g]	{16} FMO2_RABIT; L08449 etc.. + CEC01H6; CELC46H11; CEK08C7(.2); CEK08C7(.5); CEK08C7(.7) YHX6_YEAST	89–91	+6	
Cyclic monooxygenases (CMO)[h]	{10} YZ20_MYCTU; Y4ID_RHISN; STCW_EMENI; AB010439; Z80108; AL021287; AL021942; AL021309; Z83864	92, 93	+6	CYMO_ACISP : AG
BaiH—Trimethylamine dehydrogenase[i]	{10} BAIH_EUBSP; NADO-_THEBR; 2649317 (AE001017); FADH_ECOLI; CAA15852 (AL010186); CAA71086 (Y09960); CAA76082 (Y16136); DHTM_METME; X89575	26, 53, 94–96	+6	BAIC_EUBSP : AG
Glutathione reductase (GR)	{9} GSHR_ARATH; GSHR_CAEEL; GSHR_ECOLI; GSHR_HAEIN; GSHR_HUMAN; GSHR_PEA; GSHR_PSEAE; GSHR_SPISP; GSHR_YEAST		+5/+15	
Dihydrolipoamide dehydrogenase	{41} DEECLP; DEHULP etc . . . + related: S42920; U10552; S70187; E35156		+5/+12	
Thioredoxine reductase	{21} U67594; AE000058; X79603; U82978; TRXB_COXBU; TRXB_ECOLI; TRXB_CLOLI; R34K_CLOPA; TRXB_EUBAC; TRXB_HAEIN; TRXB_MYCGE; TRXB_MYCLE; TRXB_MYCTU; TRXB-_NEUCR; TRXB_PENCH; TRXB_STRCL; TRXB_STRCO; TRXB_YEAST; YHQ6_YEAST; U63713; S44027		+5/+11	
Alkyl hydroxyperoxide reductase	{5} DHNA_BACSP, DHNA_BACSU; U82598; AHPF_SALTY; JC2311		+3	
Mercuric reductases	{11} MERA_BACSR; MERA_PSEAE; MERA_SERMA; MERA_SHIFL; MERA_STAAU; MERA_STRLI; MERA_THIFE; RDPSHA, ECOMERTET; D90903; NCNR1MER		+5	
Trypanothione reductase	{5} TYTR_CRIFA; TYTR_LEIDO; TYTR_TRYBB; TYTR_TRYCO; TYTR_TRYCR		+14	

†Figures in the GG column indicate its distance to the end of the DBM.

[a]Known in yeasts as MOL1 or NMT2, in other fungi as STI35, in plants as THI1. They are involved in thiazole biosynthesis.

[b]Products of the *rfbD* genes from bacteria, involved in O-antigen biosynthesis.

[c]The *Shewanella putrefasciens* flavocytochrome c binds FAD noncovalently and has an additional cytochrome c-like N-terminal domain.

[d]Rab-GDIs from mammals, yeast, *C. elegans* and fruit fly, and choroideremia-related gene products CHM and CHML. These proteins do not bind dinucleotides, but a DBM-like sequence is found near their N-terminus, and the 3-D structure is similar to that of a flavoprotein. The region of the GG double (from D-34 to S-45) can be superposed with that of GR.

[e]NADH- and NADPH-dependent GOGATs (EC1.4.1.14 and .13, respectively) are made up of either one or two polypeptide chains. The C-terminal part (or β-subunit) contains the two DBMs.

[f]Formate dehydrogenase (subunit β) of *Moorella thermoacetica* and the product of two *E. coli* open reading frames, ORF644 and YffG (for YffG, the EMBL AE000333 entry appears correct, whereas the Swissprot YFFG_ECOLI and Genbank ECOAEG530A entries apparently contain sequence errors that disrupt the alignment).

[g]NO-forming (EC 1.14.13.8), also known as flavin-containing monooxygenases. Microsomal enzymes involved in the detoxification of xenobiotics. Their second DBM binds NADPH. A FMO homologue (presumably YHX6_YEAST) has been described in yeast.

[h]Fungal and bacterial sequences related to cyclohexanone monooxygenase and steroid monooxygenase. Show homology to FMOs.

[i]The *baiH* and *baiC* gene products of *Eubacterium;* NADH oxidases of *Thermoanaerobium brockii* and *Archaeoglobus fulgidus;* 2,4-enoyl-CoA reductase of *E. coli* (the product of the *fadH = ygiL* gene) and *Mycobacterium tuberculosis;* 2-enoate reductase from *Clostridium tyrobutyricum* and *Moorella thermoacetica;* trimethylamine dehydrogenase (TMD) (EC 1.5.99.7) of *Methylotrophus methylotrophus* and its close relative dimethylamine dehydrogenase (DMD) of *Hyphomicrobium* X.

search with flavoprotein pyridine nucleotide reductases, for which ample structural and sequence data were available. In the 87 sequences examined, we found that the region of the fourth β-strand in the FAD-binding domain obeyed the following consensus:

oohhhAT**G**

where o stands for a polar or charged, and h for a hydrophobic, residue (o can be D, E, K, R, H, N, S, or P, rarely A; h can be I, V, L, F, Y, or A). In the human GR sequence, this corresponded to residues 150–157 (PHILIATG). In the various structures available, the β-strand starts at the first residue of the consensus or slightly before and always ends at the fifth. The last three residues form a connection with a β-strand leading into the NADPH-binding domain. The sixth and seventh positions could be occupied not only by A or T (dominant) but also by S, V, or I. However, the final G was perfectly conserved. In GR, it is located in close proximity to O3 of the cofactor, and its ϕ and ψ angles are not compatible with the presence of a side chain at this position (Table III). In the following, this consensus sequence will be referred to as the ATG motif.

Extending our search to the NAD(P)H domain of these proteins, we found that the sequence corresponding to the fourth β-strand of the Rossman fold and the following loop in that domain also fitted the ATG motif. In human GR, this second ATG motif (DCLLWAIG) corresponded to residues 283–290, again just at the end of the NADPH-binding domain. The second residue of the motif could be hydrophobic instead of polar. The sixth residue was almost always an A as in the FAD-binding domain, whereas the seventh residue was more often I than T. Again, the final G was perfectly conserved, and its ϕ and ψ angles are such that no other residue can be accommodated (Table III). The sequence similarity between the two ATG motifs is remarkable, especially in view of the limited overall sequence conservation between the two domains in each family and between the different families.

ATG and GD Motifs Are Present in Most Flavoproteins With Two Dinucleotide-Binding Domains

In flavoprotein pyridine nucleotide reductases and in rubredoxin reductase (EC 1.18.1.1), Eggink et al.[11] described a consensus sequence encompassing the fifth β-strand of the FAD binding site. This motif can be written: TxxxxhϕhhGD, and we refer to it as the GD motif. In addition, some sequence similarity has also been recognized in GR immediately downstream,[42] corresponding to the fifth α-helix of the FAD-binding domain. A G residue is well conserved in the middle of the helix, where it comes closest to the fifth β-strand. The fourth position before that G is occupied by A, which in the available 3-D structures faces the G of the GD motif, and a hydrophobic residue is usually found immediately afterward. We call this AhxxG sequence the "G helix."

We found that the GD motif and the G helix could be recognized, together with the two ATG motifs, in all flavoprotein pyridine nucleotide reductases and in nearly all the other protein families that present two dinucleotide binding-domains (Table IV). Whenever possible, the 3-D structure was examined to confirm assignation of the motifs (see Fig. 4). Figure 3 presents an alignment of representative sequences from four of these families (TMD, GOGAT, FMO, and CMO). Once these motifs have been aligned, other regions of sequence similarity appear: for those families in which no 3-D structure is available, this may allow a few precise structural predictions. The other flavoproteins with two dinucleotide-binding domains are listed in Table IV together with the motifs observed. When all these sequences were taken into consideration, the definition of the motifs had to be broadened somewhat. In the first ATG motif, some variability was found in the penultimate and antepenultimate positions, although the final G was perfectly conserved. In the second ATG motif, the second residue was more often hydrophobic than polar. The GD motif often was missing the initial T (correlated in available 3-D structures with a loss of the greek key), and the second residue of the hydrophobic stretch was not necessarily aromatic as in GR. For the G helix, it could be recognized only by virtue of its proximity to the GD motif. Its second residue was not necessarily hydrophobic, and the final G could be replaced by A.

As can be seen in Table IV, the two ATG motifs are present in nearly all the families examined, the only exceptions being linked to loss of dinucleotide-binding. A clear GD motif (and G helix) can be recognized in all but four families, of which three may have a GD-like sequence. The good conservation of these motifs points to a common structural organization for all flavoproteins with two dinucleotide-binding domains, suggestive of a common origin. Their general organization can be described as follows (the parentheses are for sequence motifs that are not present in all families): DBM_{FAD}-(GG)-ATG_{FAD}-$DBM_{NAD(P)H}$-$ATG_{NAD(P)H}$-(GD_{FAD})-(G helix).

Are ATG and GD Motifs Also Present in Proteins With a Single FAD-Binding Domain?

We wondered whether the ATG and GD motifs could also be identified in flavoproteins with a single dinucleotide-binding domain (Table IV). When the available 3-D structures were examined, only *p*-hydroxybenzoate hydroxylase (PHH)[43] and the related phenol hydroxylase[44] were found to show a bona fide GD motif, as noted in Eggink et al.,[11] with the final D liganding O3*. The other flavoproteins all lacked the final GD residues in the 5th strand and used a different bonding pattern for the ribityl. In COX and GOX, O3* is bound by water molecules, not by a D residue as in the GD motif. In D-amino acid oxidase (DAO), the flavin moiety adopts a different conformation,[45] turning its O3* in the opposite direction, away from the fifth strand.

In contrast, the ATG motif appeared much more widespread, at least among the flavoprotein families for which 3D-structures were available. It was found at the expected

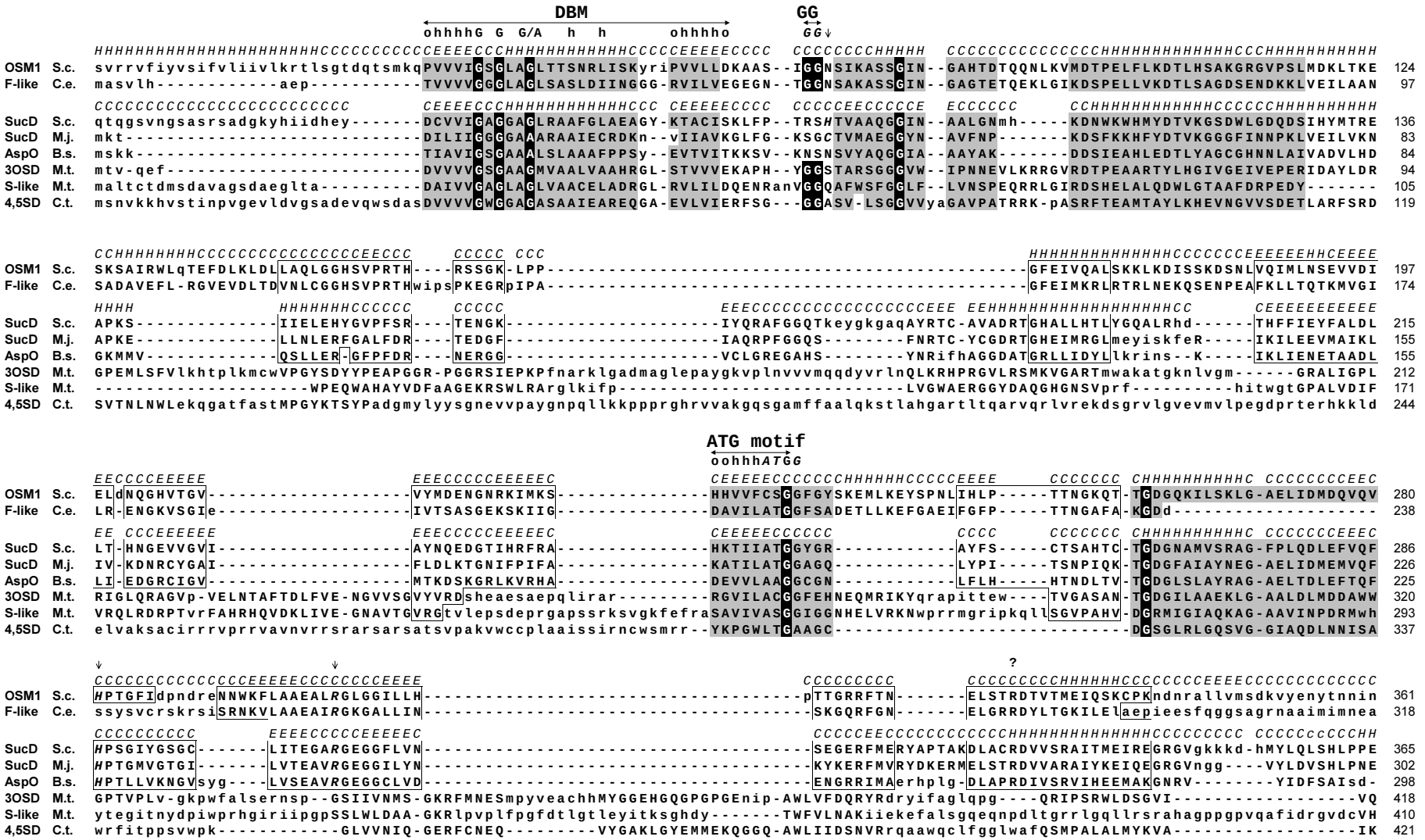
DBM
GG
ATG motif
OSM1 S.c.
F-like C.e.
SucD S.c.
SucD M.j.
AspO B.s.
3OSD M.t.
S-like M.t.
4,5SD C.t.

Figure 2.

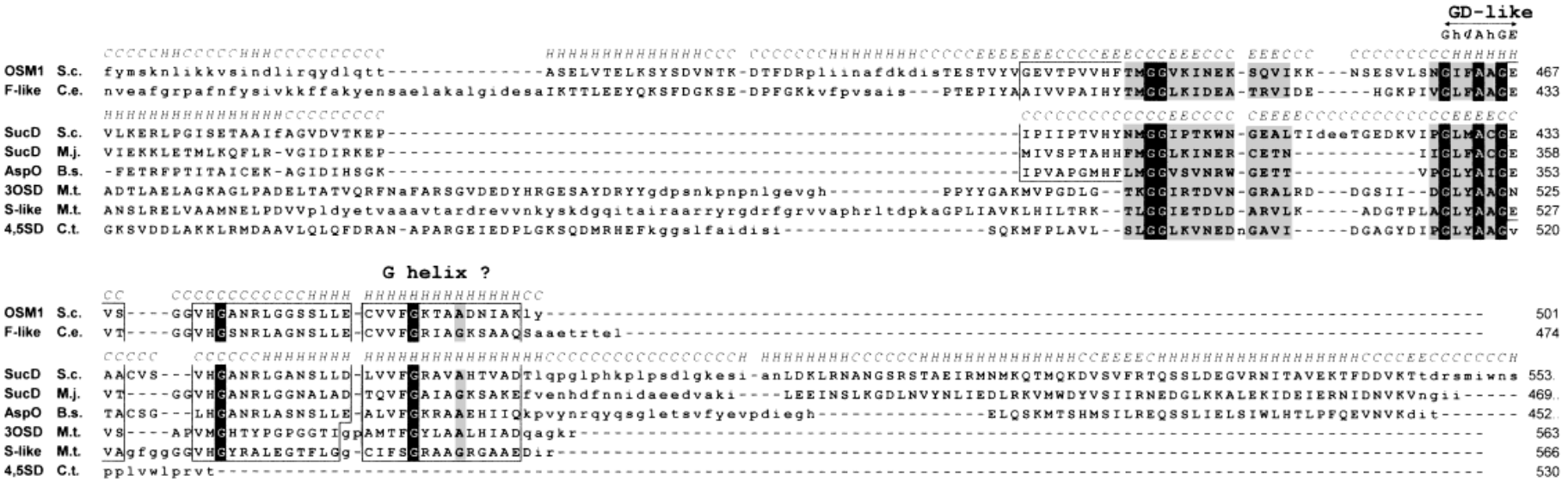

Fig. 2. (Continued.) MACAW sequence alignment of SFR and related sequences. The alignment was generated with 15 sequences, but only 8 were retained for clarity. The DBM, GG, ATG, and GD-like motifs are shown. Arrows and italics indicate the residues involved in covalent binding of FAD[23] or fumarate binding.[51] ? indicates the only other fully conserved Arg, possibly the second active site Arg postulated in.[51] Regions conserved between all families are shaded in gray (in black for fully conserved residues). Regions common to fumarate reductases, succinate dehydrogenases, and aspartate oxidases are boxed. The italicized lines above the OSM1 and SucD sequences are the secondary structure predictions generated by PSIpred. **OSM1** (OSM1_YEAST, res. 4 to end); **F-like**: an unidentified protein from *C. elegans* (AAC46539.1) with similarity to soluble fumarate reductases but lacking one of the two basic residues implicated in fumarate binding; **SucD**: succinate dehydrogenases from *S. cerevisiae* (DHSA-_YEAST, mature) and *Methanococcus jannaschii* (U67462); **AspO**: aspartate oxidase from *Bacillus subtilis* (NADB_BACSU); **3OSD**: 3-oxosteroid 1-dehydrogeneses from *M. tuberculosis* (MTCY03C7.19c); **S-like**: steroid dehydrogenase-like protein from *M. tuberculosis* (MTCY369.29); **4,5SD**: δ4, 5-alpha steroid dehydrogenase from *Comamonas testosteroni* (L23428).

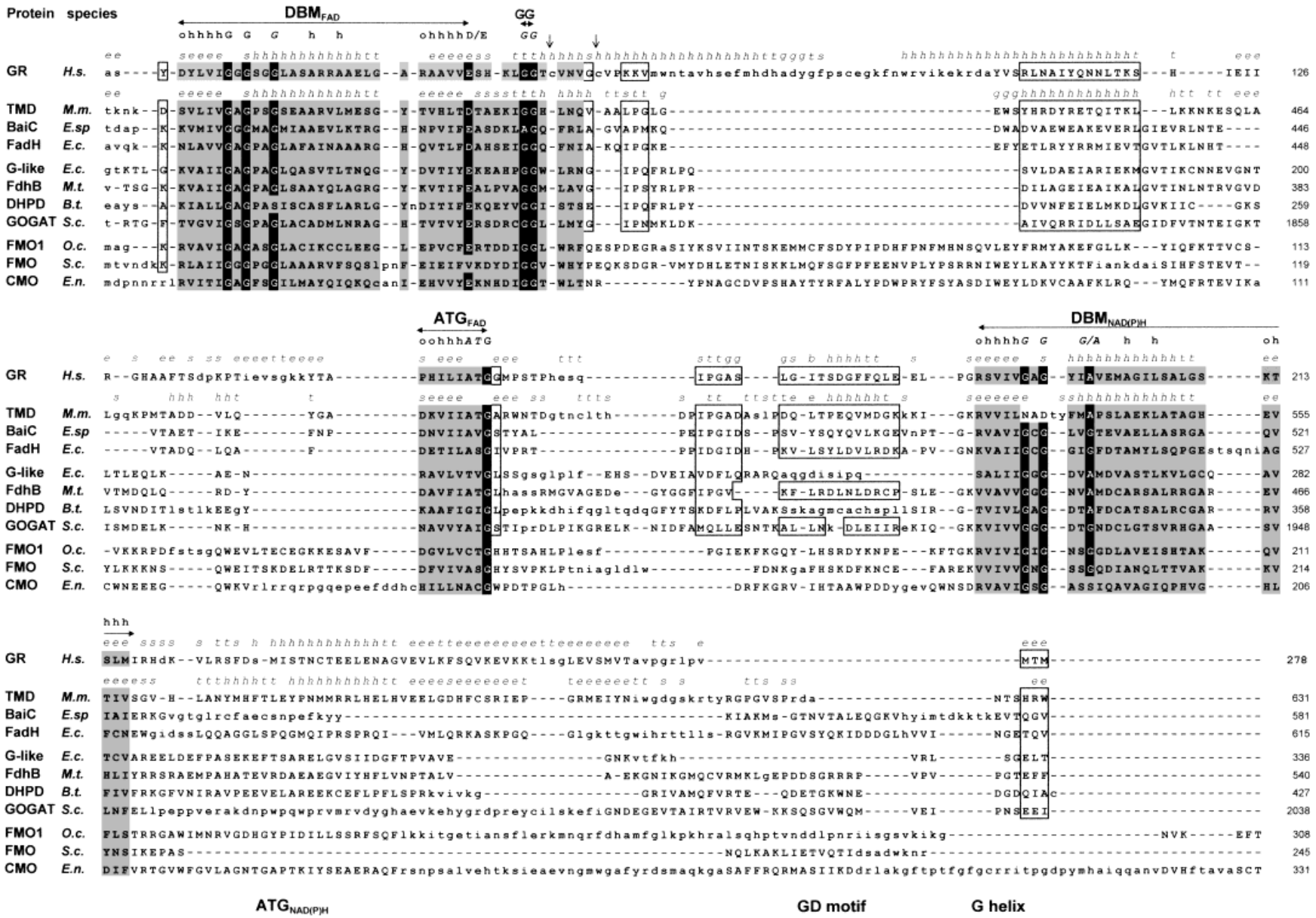
Protein species
DBM$_{FAD}$
GG
ATG$_{FAD}$
DBM$_{NAD(P)H}$
GR H.s.
TMD M.m.
BaiC E.sp
FadH E.c.
G-like E.c.
FdhB M.t.
DHPD B.t.
GOGAT S.c.
FMO1 O.c.
FMO S.c.
CMO E.n.
ATG$_{NAD(P)H}$
GD motif
G helix

Figure 3.

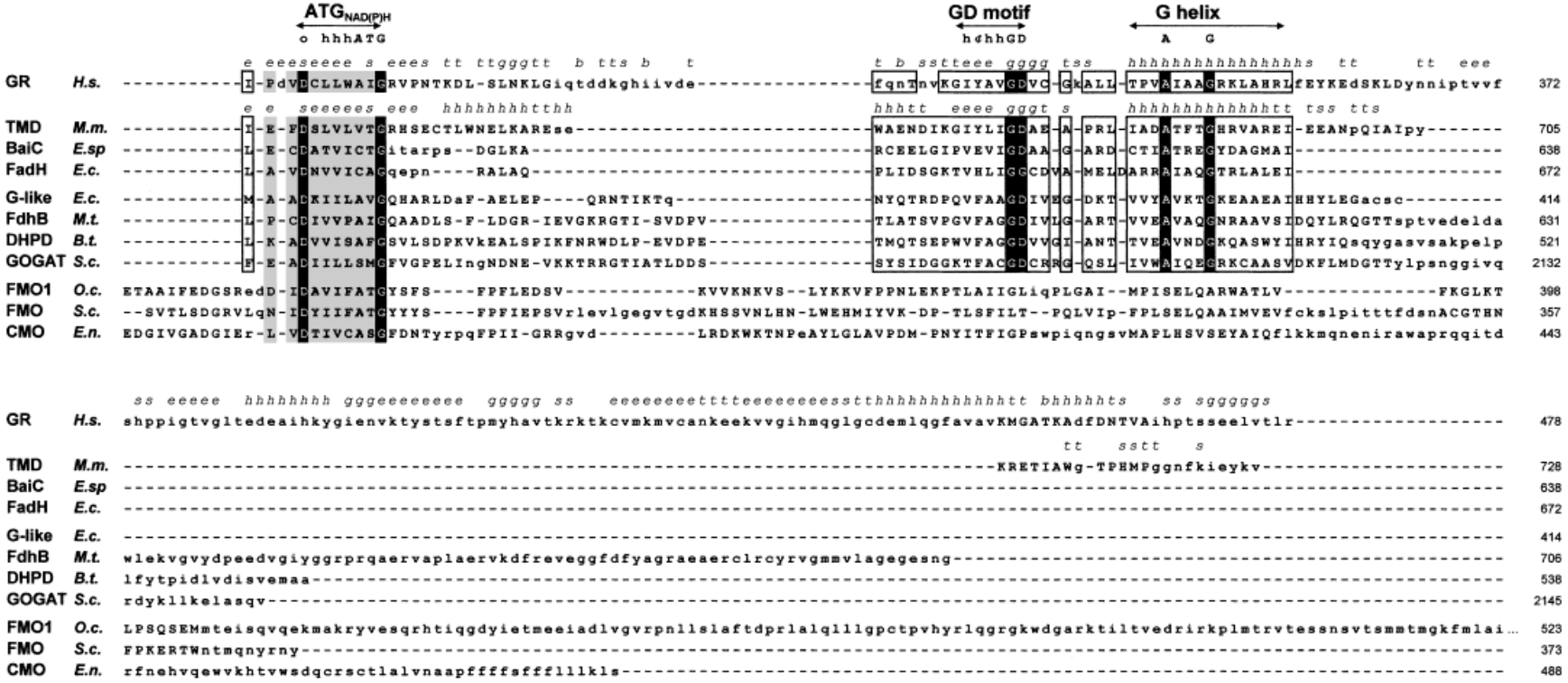

Fig. 3. (Continued.) Sequence alignment, using MACAW, of flavoproteins with two dinucleotide-binding domains. The alignment was generated with 27 sequences, but only 10 have been retained for clarity. The GR sequence was added based on structural superposition with TMD in the FSSP entry for GR. The lines in italics above the GR and TMD sequences indicate their secondary structures, according to FSSP: *e* for extended (β-strands), *h* and *t* for α-helices, g for 3-turns, *s* for bends. Arrows point to the redox-active Cys in GR. **GR** (GSHR_HUMAN, res. 19 to end); **TMD** (DHTM_METME, res. 386 to end); **BaiC** (BAIC_EUBSP, res. 369 to end; note that the sequence has been corrected by adding a nucleotide immediately before the published stop codon,[52] to restore homology to the other family members); **FadH**: 2–4 enoyl CoA reductase (FADH_ECOLI, res. 371 to end; note that the sequence in[53] contains a frameshift that masks the GD motif); **GOGAT-like** protein of *E. coli* = ORF462 (AE000303 o462, res. 120 to end); **FdhB**: formate dehydrogenase β subunit (U73807, res. 371 to end); **DHPD**: dihydropyrimidine dehydrogenase (BTU20981, res. 183–539); **GOGAT** from *S. cerevisiae* (SCGLUTSYN, res. 1778 to end); **FMOs** from *Oryctolagus cuniculus* (1C1: FMO5_RABIT) and *S. cerevisiae* (YHX6_YEAST); **CMO:** stcW, a putative sterigmatocystin biosynthetic gene from *Emericella nidulans* (STCW_EMENI).

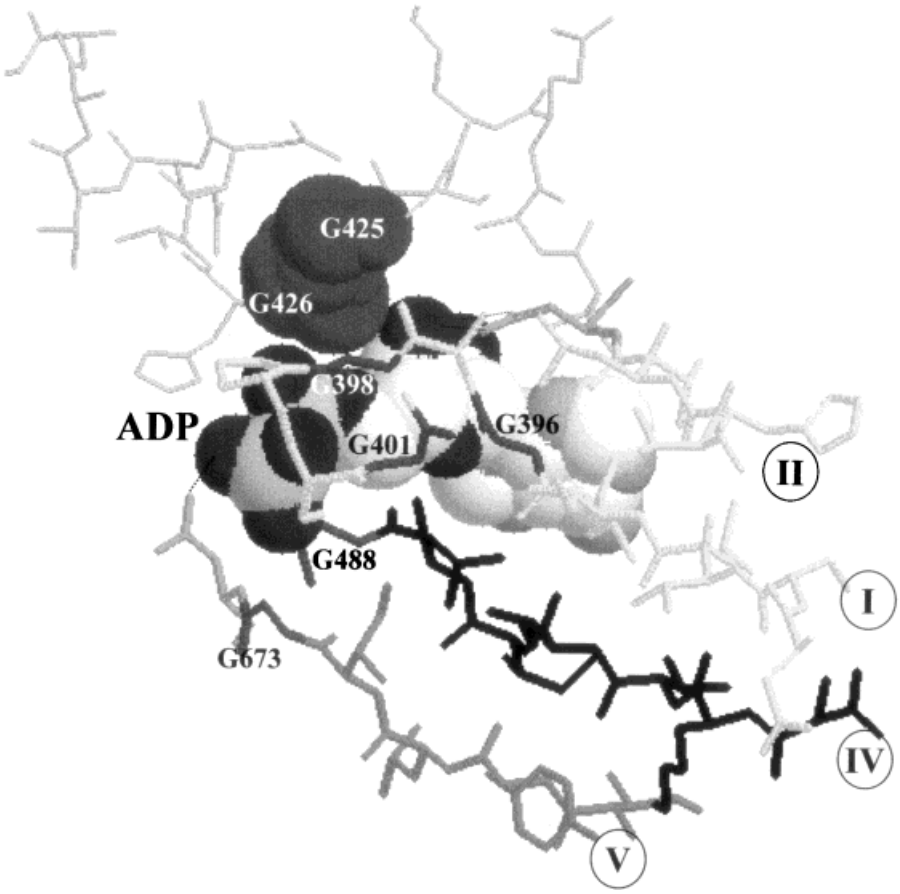

Fig. 4. Position of the GG doublet in TMD. A space-filling representation is used for the GG doublet and ADP cofactor. Other important structures are represented as wireframe: the connection between DBM and GG doublet and the following helix (thin), the DBM (most of the helix removed), the ATG and GD motifs. β-strands are marked with roman numerals at their N-terminus. Dotted lines indicate H-bonds between ADP and residues D419 and D674.

position in COX, GOX, and the entire GMC family, in DAO[45] and in PHH and the related phenol 2-monooxygenases (PheA/TfdB). A notable exception was Rab-GDI (which as mentioned above does not bind FAD). When we examined flavoprotein families for which no structural data were available, our sequence alignments sometimes revealed putative ATG motifs (Table IV and Fig. 2). In SFR, the PSIpred program consistently predicted the corresponding region as a β-strand, and three other strands were predicted in the preceding 40 residues that may constitute the three-stranded antiparallel β-sheet connecting helix III and strand 4. In most families, however, the ATG motif could not be recognized unambiguously. In combination with secondary structure predictions, only tentative assignments could be made. For example, in the alignment proposed for MAO-related sequences, a hydrophobic stretch followed by P might correspond to the fourth β-strand (ATG-like in Fig. 1), whereas another region can be proposed to form the fifth β-strand (GD-like) and following helix. Similarly, a central conserved region of carotenoid desaturases, centered around a hydrophobic sequence, may be interpreted as forming the fifth strand (residues 424–429 in CRTI_SYNY3).

DISCUSSION

Evolution of Dinucleotide-Binding Proteins: The Rossman Fold, the "Pre-DBM" and the DBM

Most of the flavoproteins whose 3-D structure has been solved were found to adopt a similar folding pattern based on the Rossman fold. This fold is also found in NAD(P)H-binding proteins and in proteins that bind other cofactors or do not bind any. A list of the available 3-D structures is presented in Table V, organized according to the FSSP classification. Figure 5 presents a proposal for a classification of Rossman-type proteins, intended to organize them from simplest to most complex on the basis of their topology, cofactor-binding properties, and recognized sequence motifs. It may also be viewed as a tentative scheme for the evolution of dinucleotide-binding proteins, but it should be stressed that other pathways are possible: the building of a solid evolutionary tree is beyond the scope of this study.

As described in the introduction, the Rossman fold in its simplest form (i.e., as it appears in glucose 6-phosphate dehydrogenase) is a six-stranded parallel β-sheet, with α-helices connecting the strands in a right-handed fashion.

TABLE III. Angle and Distance Data for the GG, ATG, and GD Motifs in TMD, GR, and COX

Motif	Molecule	Res No	Φ	Ψ	Distance to DBM (or ATG)	Distance to cofactor
GG	TMD	G 425	135	−22	N at 2.85 Å from A 397 O C at 3.72 Å from G 398 CA	
		G 426	56	−133	N at 3.74 Å from A 397 O N at 3.67 Å from G 398 CA	CA at 3.46 Å from O3* CA at 3.74 Å from O5* CA at 3.39 Å from O2A
	GR	G 55	104	−24	N at 2.90 Å from G 28 O C at 3.90 Å from G 29 CA	
		G 56	60	−151	N at 3.66 Å from G 28 O CA at 3.76 Å from G 29 CA	N at 3.89 Å from AO3* CA at 3.15 Å from AO3* CA at 3.86 Å from A05* CA at 3.79 Å from A02 C at 3.88 Å from A02
	COX	G 113	117	−12	N at 3.10 Å from S 19 O N at 3.05 Å from S 19 OG CA at 3.07 Å from S 19 O C at 3.17 Å from S 19 O O at 3.76 Å from G 20 CA	
		G 114	52	−136	N at 3.51 Å from S 19 O CA at 3.95 Å from G 20 CA	N at 3.49 Å from AO3* CA at 3.88 Å from AO5* CA at 3.04 Å from AO3* CA at 3.72 Å from AC3* C at 3.87 Å from AO1
1st ATG	TMD	G 488	87	−162		N at 3.71 Å from C5* CA at 3.77 Å from C5* CA at 3.97 Å from O2B CA at 3.40 Å from O1A CA at 3.27 Å from O3A
	GR	G 157	65	−153		N at 3.65 Å from AC5* N at 3.68 Å from A01 CA at 3.34 Å from O3 CA at 3.66 Å from OP2
	COX	G 290	95	164		N at 3.87 Å from AC5* CA at 3.60 Å from O3 CA at 3.68 Å from OP2
2nd ATG	TMD	G 643	156	173		
	GR	G 290	163	−173		
GD	TMD	G 673	68	−153	CA at 3.67 Å from S 400 OG	CA at 3.59 Å from O2B C at 3.74 Å from O2B
		D 674	−64	−21	CB at 3.76 Å from G 488 C (ATG) C at 3.50 Å from G 488 O (ATG)	N at 2.96 Å from O2B OD2 at 2.63 Å from O3B
	GR	G 330	83	−158		CA at 3.52 Å from OP2
		D 331	−55	−31	CB at 3.92 Å from G 157 CA (ATG)	N at 2.97 Å from OP2 OD1 at 3.32 Å from C3* OD2 at 2.76 Å from O3* OD2 at 3.19 Å from C5*
	COX	D 474	−124	−159	OD2 at 2.78 Å from G 290 N (ATG)	CB at 3.43 Å from OP2
		G 475	−68	−11		N at 2.95 Å from OP2

Many variations on this theme can be found, i.e., structures that can be derived simply from the Rossman fold by deleting or adding structural elements without reorganizing the existing connectivity of the strands. For example, the final strand is missing in flavoproteins, and their third helix is replaced by an antiparallel β-sheet (except in DAO). The same holds for the NAD(P)H-binding domains of flavoprotein pyridine nucleotide reductases, which in addition lack the fifth helix and have an additional β-strand at their N-terminus.

In Rossman-type proteins, G residues are often found at the C-terminus of the first (central) β-strand. This is true even for proteins that do not bind cofactors or bind cofactors other than dinucleotides at that site, such as aspartate carbamoyltransferase, vp39 methyltransferase or carboxyl methylesterase. The tendency to accumulate G residues in that area becomes particularly strong in NADH-binding proteins. For example, aldehyde dehydrogenase has the sequence GSTVG at this position, whereas glucose 6-phosphate dehydrogenase has GGTGDLA. For a large series of NAD(P)H-binding proteins, a more elaborate consensus sequence (type 3 in first column of Table V) can be derived for this region: G(G/A/x)x(G/x)xx(G/A), with the Gs at positions 1, 4, and 7 occupying positions equivalent

TABLE IV. Occurrence of Sequence Motifs in Flavoproteins[†]

Protein family	{n} domain:	Example	GG FAD	ATG FAD	ATG NAD(P)H	GD FAD	G helix FAD
II) Two dinucleotide-binding domains:							
Pyridine nucleotide reductases*[a]	{92}	GSHR_HUMAN	+56	+157	+290	+331	+346
NADH oxidase/NADH peroxidase*	{26}	NAPE_ENTFA	–	+112	+243	+281	+306
BaiH-Trimethylamine dehydrogenase*	{10}	DHTM_METME	+426	+[b]488	+[c]641	+[d]572	+686
Glutamate synthase (GOGAT)	{11}	SCGLUTSYN	+1819	+1876	+[e]2046	+[d]2088	+2103
Dimethylaniline monooxygenase	{16}	RABDIANMON	+40	+149	+[f]329	?[g] (369)	–
Cyclic monooxygenases (CMO)	{10}	STCW_EMENI	+46	+148	+352	?[h] (399)	–
Rubredoxin/putidaredoxin/ferredoxin reductases	{22}	RURE_PSEOL	–	+[i]107	+238	+275	+309
Nitrate and nitrite reductases	{20}	NIRB_ECOLI	–	+110	+[j]242	+[k]281	+[l]304
NADH-ubiquinone reductase/ndhA	{14}	NDI1_YEAST	–	+117	+280	+[d]323	+[l]400
Adrenodoxin and ferredoxin reductases	{6}	ADRO_BOVIN	–	+121	+366	?[m] (405)	–
Lysine:N6-, ornithine:N5-hydroxylases	{5}	IUCD_ECOLI	–	+150	+331	–	–
Flavocytochrome *c* sulfide dehydrogenase*	{9}	1fcd	–	+135	–[n]	+[d′]324	+342
I) One dinucleotide-binding domain:							
GMC oxidoreductases*[o]	{27}	GOX_ASPNG	?124	+312		–	–
D-amino acid oxidase*	{7}	OXDA_PIG	–	+[p]183		–	–
p-hydroxybenzoate hydroxylase*	{13}	PSEPOBA	–	+[q]160		+[d]286	–
Rab-GDI*	{21}	GDIA_BOVIN	+41	–		–	–
Fumarate reductase, steroid dehydrogenases	{9}	OSM1_YEAST	+72	+230		?[r]467	+494
dTDP-4-dehydrorhamnose reductase	{5}	GLF_ECOLI	+38	?[s] (231 or 239)		–	–
THI4	{9}	THI4_YEAST	+104	+[t]236		?[u] (291)	? (303)
LAO, MAO, TMO, CBP	{36}	A38314	+213	?[v] (475)		?[w] (647)	? (663)

[†]Families for which 3-D structures are known are indicated by *. The figures in the table indicate the position of the last residue of the motif in the example sequence.

[a]GR, mercuric reductase, dihydrolipoamide dehydrogenase, thioredoxine reductase, alkyl-hydroxyperoxide reductase and trypanothione reductase, see Table II.

[b]Can have Y as the 2nd residue, or T as the 3rd.

[c]In DMD, the final G is replaced by S; in DMD as in TMD, this domain does not bind NAD(P)H.[97]

[d]The initial T of GD motif is missing or [d′]: replaced by S.

[e]Charged residues at positions 3 or 5 pointing toward the antiparallel β-sheet.

[f]This sequence had been noted in[98] but was considered not significant because both flavin-binding sites were thought to lie in the α-subunit. The *S. cerevisiae* enzyme lacks the initial T.

[e]Sequence conservation had been noted in[90] but attributed to the formation of a stable C(α4)-hydroperoxyflavin intermediate.

[g]The closest match was a conserved hydrophobic stretch followed by GL. Absent in yeast FMO which has an complete DBM and appears to bind FAD weakly.[91]

[h]The closest match was a conserved hydrophobic stretch followed by GP.

[i]In *P. oleovorans* rubredoxin reductase, final G is replaced by P.

[j]Sometimes has R at the 6th position.

[k]Ends more often in E than in D.

[l]G often replaced by A.

[m]The closest match was a conserved hydrophobic stretch followed by GW.

[n]Final G replaced by P. The domain does not bind NAD(P)H, but rather another unknown cofactor.[33]

[o]Glucose, cholesterol, alcohol, choline, and methanol oxidases; glucose, alcohol, choline, sorbose, and cellobiose dehydrogenases; mandelonitrile lyase; versicolorine synthase.

[p]5th residue always an N.

[q]The unusual G and D residues at positions 5 and 7 are strictly conserved. In PHH and phenol hydroxylase, the D faces the N-end of a short α-helix in the back of the β-sheet, thus compensating for its dipole, and the G at the 5th position allows close approach to this helix.

[r]Ends in E, sometimes in N, A, or V.

[s]3rd or 4th residue is hydrophilic.

[t]5th residue is hydrophilic.

[u]Ends in M.

[v]Ends in P.

[w]Ends in D, E, S, T, C, or A.

to those of the DBM. These proteins include the RED superfamily[46] (dihydropteridine reductase, 3-α, 20-β-hydroxysteroid dehydrogenase and UDP-galactose 4-epimerase), carbonyl reductase, malate dehydrogenases, glucose 6-phosphate dehydrogenase, enoyl-acyl carrier protein reductases, cis-biphenyl-2,3-dihydrodiol-2,3-dehydrogenase, etc.. We call this type of sequence a "pre-DBM" because it is less defined and less compact than the classical DBM, but we have no indication that the latter has evolved form the former. As in the DBM, hydrophobic residues are generally found at the critical positions in the two β-strands and intervening α-helix. Some of these proteins (malate dehydrogenase, UDP-galactose 4-epimerase, dihydropicolinate reductase, etc.) also have a negatively charged

TABLE V. Sequence Motifs and Cofactor Binding in Rossman-Type Proteins[†]

DBM	ATG	Cofactor	FSSP branch	PDB no.	Name	Notes on structure
2	−[a]	NADH	55.1.1.1.1.1	1ad3	aldehyde dehydrogenase (N-term domain)	6th strand missing
2E	−[a]	HCA + FeMo	55.1.1.2.1.1	3min	nitrogenase molybdenum iron protein (C-term domain)	6th strand missing
2D[b]	−	S-Ade-homoCys	55.2.1.1.1.1	1v39	vp39 (polyA polymerase regulatory subunit)	6th strand missing
2D	−[a]	S-AdeMet	55.2.1.1.2.2	1xva	glycine N-methyltransferase	6th strand missing
3′AR	−	NAD(P)H	55.2.1.2.1.1	1dpg	glucose 6-phosphate dehydrogenase	
3E	−	NAD(P)H	55.2.1.2.2.1	1drw	dihydrodipicolinate reductase	additional 7th strand
4SR	+	NADPH	55.2.1.2.2.1	1dap	diaminopimelic acid dehydrogenase	additional 7th strand
2A	−	none (buried)	55.2.1.2.3.1	8atc	aspartate carbamoyltransferase (C-terminal domain)	N-term domain also Rossman fold: no DBM, an ATG
4E	−	NADH	55.2.1.2.4.1	1leh	leucine dehydrogenase	
4′D	−	NADH	55.2.1.2.4.2	2nad	formate dehydrogenase (C-term domain)	additional 7th strand (parallel)
4D	+	NADH	55.2.1.2.5.1	1gyp	glyceraldehyde-3-phosphate dehydrogenase	
4D	+	NADH	55.2.1.2.6.1	2ohx	alcohol dehydrogenase	
3′TR	+	NADPH	55.2.1.2.7.1	1cyd	carbonyl reductase	
3D	+	NADH	55.2.1.2.7.1	1dhr	dihydropteridine reductase	
2[c]G	+[d]	NADH	55.2.1.2.7.1	1eny	enoyl-acyl carrier protein reductase	
3′D	+	NADH	55.2.1.2.7.1	1hdc	3-α, 20-β-hydroxysteroid dehydrogenase	additional 7th strand
3D	+	NADH	55.2.1.2.7.1	1xel	UDP-galactose 4-epimerase	
3′LR	+	NADPH	55.2.1.2.7.1	1fds	17-β-hydroxysteroid-dehydrogenase	
3E	−[e]	NADH	55.2.1.2.8.1	1bmd	malate dehydrogenase (T. flavus)	
3D	+	NADH	55.2.1.2.8.1	2cmd	malate dehydrogenase (E. coli)	
3D	+	NADH	55.2.1.2.8.1	1hyh	1-2-hydroxyisocaproate dehydrogenase	
4D	+	NADH	55.2.1.2.8.1	1ldg	1-lactate dehydrogenase	
4NR	−	NADPH	55.2.1.2.9.1	2pgd	6-phosphogluconate dehydrogenase	additional anti-parallel 7th strand
4LR	−	NADPH	55.2.1.2.9.2	1yveI	acetohydroxy acid isomeroreductase	
2[f]Q	+	none	55.2.2.1.1.1	1chd	cheB carboxyl methylesterase (C-terminal domain)	additional strand + helix after 2nd strand, order reads: 6, 5, 4, 1, 2, 3, 2′
2G	−	vit.B12	55.2.2.2.1.1	1bmt	methionine synthase	3rd strand missing
3T	−	CoA	55.2.2.2.2.1	1scu	succinate-CoA ligase (N-terminal)	2nd helix replaced by β-strand
1E	−	none	55.2.2.2.3.1	3chy	CheY	3rd strand missing
2Y	−	NAD(P)H	55.2.2.3.3.1	1iso	isocitrate dehydrogenase	3rd and 6th strands missing
2H	+	ATP	55.2.2.4.1.1	1bnc	biotin carboxylase	
1G	−	ATP	55.2.3.1.1.1	1gpb	glycogen phosphorylase b	
1N	−	none	55.2.3.2.1.1	1nba	N-carbamoylsarcosine amidohydrolase (A)	
2S	−	none	55.3.1.1.1.1	1esc	serine esterase	3rd strand missing
2	−	NADH	55.5.1.1.1.1	1ndh	cytochrome b_5 reductase	6th strand missing
2	+	ADP	55.5.2.1.1.1	1php	3-phosphoglycerate kinase (N-terminal domain)	additional strand + helix after 2nd strand 6th strand missing; order: 6, 5, 1, 2, 4, 3
3′I	−	none	55.6.1.1.1.1	1pdo	mannose permease phosphotransferase	3rd and 6th strands missing
2	−	none	55.7.1.1.1.1	1tpt	thymidine phosphorylase	6th strand missing
*4D	+	*ADP	55.8.1.1.1.1	2tmd	trimethylamine dehydrogenase (chain A, C-terminal part)	*no anti/ β-sheet, no greek key
§1S	−	§?				§4 strands + 1 (at N-terminus)
*4D	+	*FAD	55.8.1.1.1.1	1fcd	flavocytochrome c sulfide dehydrogenase, chain A	*No GG doublet, covalent FAD
§2S	−	§?				§4 strands + 1 (at N-terminus)
*4T	+	*FAD	55.8.1.1.1.1	1trb	thioredoxin reductase	*
§4HR	+	§NADPH				§4 strands + 1 (at N-terminus)
*4E	+	*FAD	55.8.1.1.1.1	3grs	glutathione reductase	*
§4VR	+	§NADPH				§4 strands + 1 (at N-terminus)
*4E	+	*FAD	55.8.1.1.1.1	1nhp	NADH peroxidase	*No GG doublet
§4D	+	§NADH				§4 strands + 1 (at N-terminus)
*4E	+	*FAD	55.8.1.1.1.1	3lad	dihydrolipoamide dehydrogenase	*
§4E	+	§NADH				§4 strands + 1 (at N-terminus)
4′	−	none	55.8.1.1.2.1	1gnd	guanine nucleotide dissociation inhibitor	GG doublet at base of a β-strand
4	+	FAD	55.8.1.1.2.2	1pbe	p-hydroxybenzoate hydroxylase	
4	+	FAD	55.8.1.1.2.2	1aa8	D-amino acid oxidase	
4	+	FAD	55.8.1.1.2.3	1gal	glucose oxidase	GG doublet at other location
4	+	FAD	55.8.1.1.2.3	3cox	cholesterol oxidase	GG doublet at other location

[†]Based on the FSSP classification (class 55 as of October 1997. This class also comprises proteins based on alternating β-strand and α-helices, but with a different topology that cannot be derived from the typical Rossman fold by a simple addition or deletion of strands. They have been omitted from the analysis). For the DBM, the nature of the central G-rich sequence is indicated by the first letter: 1 = no G; 2 = one G or more, but different from 3 or 4; 3 = GxxGxxG; 4 = Gx(G/A)xx(G/A). ′ indicates one mismatch. The last residue of the 2nd β-strand is also indicated, together with the following one when relevant to NADH/NADPH discrimination. When two dinucleotide-binding domains are present, they are marked with * and § for the N- and C-terminal domains, respectively.

[a]4th strand ends in G.

[b]D95 at end of β-strand 2 binds O2* and O3*.

[c]The other family members have a pre-DBM.

[d]Half the sequences had A at the last position, instead of G.

[e]Aligned with ATG motif: DYALLVGA.

[f]The sequence resembles a DBM: GASTGG, where the central S has been implicated in catalysis.[99]

residue at the end of the 2nd β-strand, which binds the ribose of the cofactor.

The distributions of the pre-DBM and DBM coincides in part with the FSSP classification (Table V), which is a good indication that they can be used to trace evolutionary links. Although both types of motifs are encountered in branches 55.2.1.2.2, .6, and .8, the pre-DBM is the rule in branch 55.2.1.2.7, and the DBM predominates in branches 55.2.1.2.4, .5, .6, and .9, and in branch 55.8 (the flavoproteins). Presumably, the latter branch has evolved from an NAD(P)H-binding ancestor that already presented a DBM (step G in Fig. 5).

The ATG Motif

Because of the pseudosymmetry axis present in the Rossman fold, the end of the fourth β-strand and connecting loop are subject to similar structural constraints that gave rise to the G-rich sequence of the DBM. Here again, the need to accommodate a substrate or a cofactor appears to have favored G over more bulky residues. Indeed, a bona fide ATG motif (loosely defined here as the sequence oxhhhxxG) can be recognized in many Rossman-type proteins (Table V). In many cases, the first residue was a D. The nature of the second residue was variable from one family to another, as had been observed in flavoprotein pyridine nucleotide reductases. At variance with the classical ATG motif of flavoproteins, the fifth residue was not always hydrophobic, because N, H, or even E could be found at that position. We did not endeavor to check the conservation of the motif in every member of these families, as we had done for flavoproteins. However, cursory examination of additional sequences and of published alignments showed excellent conservation of the first and last residues and of the hydrophobic nature of the 3rd and 4th.

Note that most of the proteins that present an ATG motif also bind NAD(P)H or FAD, which suggests that this sequence plays a role in the interaction with dinucleotides. However, the ATG motif was absent in several families of NAD(P)H-binding proteins, and no correlation was found between the ATG motif and the presence of NAD(P)H in the crystal structure nor the tightness of the connection with the 4th α-helix. Note also that the ATG motif was present in some families that lack a DBM or pre-DBM. The motif may have appeared independently several times, each time probably as a further specialization toward dinucleotide binding (for the sake of clarity, this is presented as a single step D, in Fig. 5). However, in the case of flavoproteins, its presence in both domains of pyridine nucleotide reductases and in other flavoproteins strongly suggests that it was an established feature of a common ancestor. In flavoproteins, the ATG motif has acquired a specific function, because it is always at the junction with the substrate-binding domain, and not within a domain as in NAD(P)H-binding proteins. Interestingly, several examples of secondary loss of the ATG motif are found in proteins that no longer bind dinucleotides (TMD, FCSD, and Rab-GDI). In these cases, the DBM is also altered, which underlines how both sequence characteristics are tightly linked to cofactor binding.

The GD Motif

As recognized by Eggink et al.,[11] the structural basis for conservation of the GD motif is in the bonding of the O3* of the flavin ribityl by the OD2 of its final D residue. Schulz[5] has noted the interesting symmetry between this D and the D/E residue that ends the 2nd β-sheet (see Fig. 4), similar to the symmetry we have described between the ATG and the DBM. The distribution of the GD motif, however, suggests a less ancient origin. It is found in most flavoproteins with two dinucleotide-binding domains, where it clearly indicates homology. However, its presence in PHH and maybe in SFR and other flavoprotein families may be due to convergence. In many cases, the critical D residue is missing or poorly conserved, which suggests a different type of interaction with the ribityl. In conclusion, the GD motif appears as a dispensable sequence feature in flavoproteins, acquired probably late in evolution (step H in Fig. 5), and possibly several times independently. When present, it is a useful tool for structural predictions, in combination with the ATG motif (see Figs. 1–3).

GG Motif Versus GG Doublet

We propose that the two conserved Gly residues (the GG doublet) found a few residues downstream of the first DBM in most flavoproteins with two dinucleotide-binding domains is an ancestral feature of these proteins, as are the ATG and GD motifs. Their absence in FCSD or NADH oxidase would be the consequence of secondary loss. When structural data are available, the GG doublet appears to play an important role in binding FAD (or ADP). It allows close contact between the first phosphate of the pyrophosphate (closest to the adenine moiety) and the N-terminal end of a structurally important α-helix. Stabilization of the cofactor is achieved via H-bonding of the next residue with O2A or O1A and via the favorable electrostatic interaction between the helix dipole and the negatively charged phosphate. A similar role has been described for the first α-helix of the Rossman fold.[47,46]

Concerning the GG motifs or GG doublets of flavoproteins with a single dinucleotide-binding domain, however, the lack of extended sequence similarity makes it difficult to distinguish between homology and sequence convergence. When the structures of Rab-GDI or GMC oxidoreductases are compared with that of GR, it appears that although the GG doublets are superimposable, the sequences that follow are completely unrelated. When considering families for which no structural data are available, secondary structure predictions may provide interesting clues. For MAO-related sequences (Fig. 1), carotenoid desaturases and protoporphyrinogen oxidases (i.e., all proteins with a GG motif) and for dTDP-4-dehydrorhamnose reductases and THI4 proteins, the prediction was that the two G's are immediately followed by a β-strand. Is this putative strand similar to the one found in Rab-GDI? Note that the latter also has conserved S/T four residues downstream, as in the GG motif. In contrast, the predic-

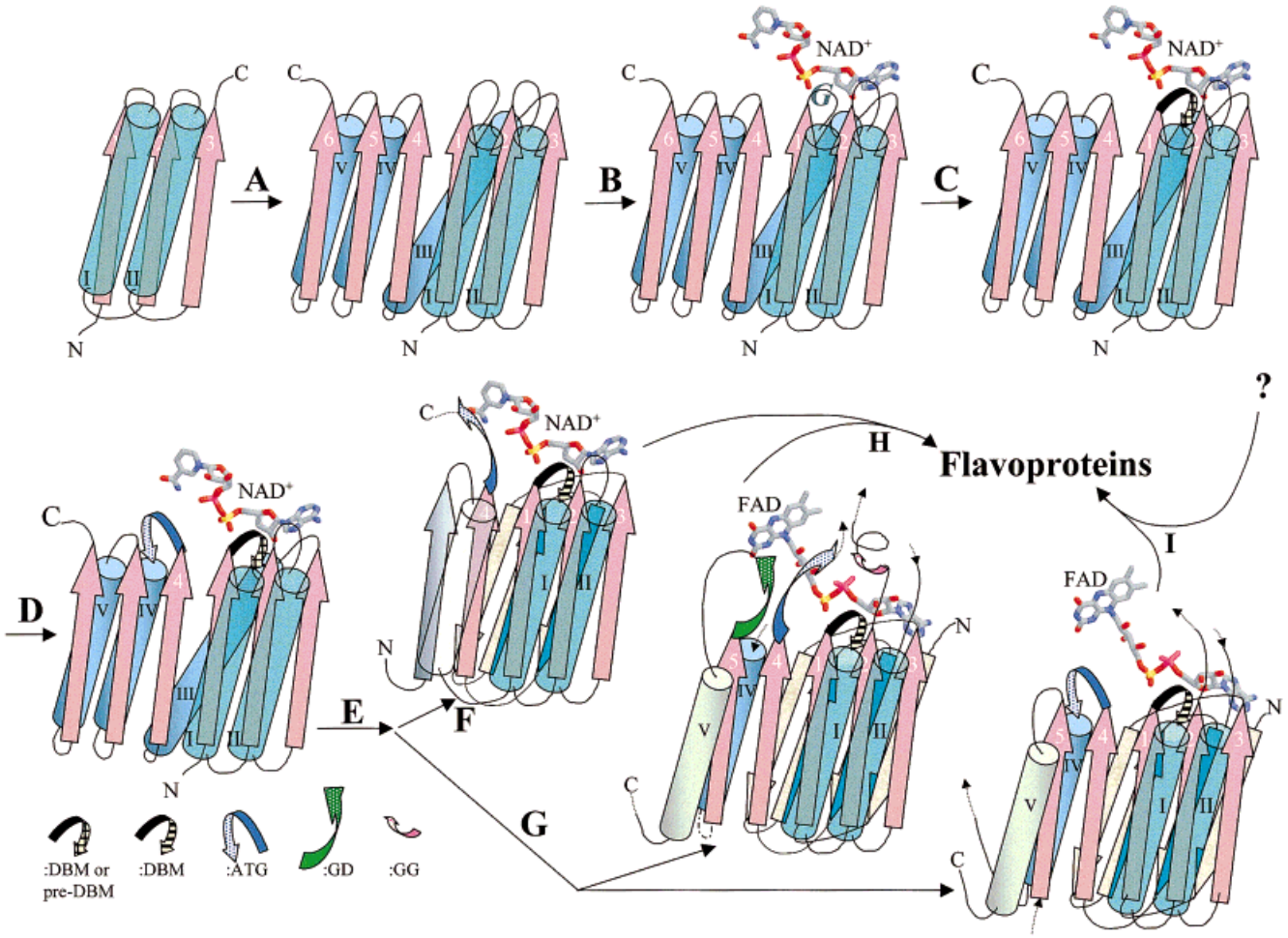

Fig. 5. A simplified classification, and possible evolutionary scheme, for Rossman-type proteins. β-strands are numbered in arabic, α-helices in roman numerals. **A:** Duplication of a hypothetical three-stranded β-sheet forms the basic Rossman fold (e.g., N-carbamoylsarcosine amidohydrolase). **B:** Acquisition of G residues between strand 1 and helix I (e.g., C-terminal domain of aspartate carbamoyltransferase), leading to binding of cofactor: NAD(P)H (e.g., enoyl-acyl carrier protein reductase), ATP (e.g., glycogen phosphorylase b), or other (e.g., glycine N-methyltransferase). **C:** Apparition of the DBM and pre-DBM motifs (e.g., leucine dehydrogenase, dihydrodipicolinate reductase). **D:** Acquisition of the ATG motif (e.g., alcohol dehydrogenase, dihydropteridine reductase). **E:** Replacement of helix II by a three-stranded antiparallel β-sheet; **F:** Loss of strands 5 and 6 and helices IV and V; addition of a strand and helix at the N-terminus (NAD(P)H-binding domain of flavoprotein pyridine nucleotide reductases). **G:** Acquisition of FAD-binding, loss of strand 6, helix V moves to front; acquisition of GD motif and GG doublet (only in the FAD domain of flavoprotein pyridine nucleotide reductases and a few others). **H:** insertion of NAD(P)H domain after strands 4 and 2 (flavoprotein pyridine nucleotide reductases). **I:** Insertion of substrate binding domain after helix IV and after strand 2 (other flavoproteins).

tion for SFR is that the GG doublet is followed by an α-helix, as in GR. Together with the presence of a classical ATG (and GD?) motif, this may be taken as an indication that SFR shares common ancestry with flavoprotein pyridine nucleotide reductases.

Origin and Evolution of Flavoproteins

Central to the building of Figure 5 is the notion that flavoproteins have evolved from an NAD(P)H-binding protein. Pyridine nucleotides are usually found as soluble electron/proton carriers in biological pathways, and their use in biological systems may therefore predate that of flavins, which generally serve as protein-bound intermediates in intraprotein redox reactions. The fact that NAD(P)H-binding proteins appear at the same time more diverse and simpler in structure than FAD-binding proteins also suggests an earlier appearance.

But where should we place the divergence between FAD- and NAD(P)H- binding proteins? Despite the limited number of structures available, it is obvious that the FAD-binding domain of simple flavoproteins (with only one Rossman-type domain) such as PHH, DAO, GOX, and COX (and even the non-flavoprotein GDI) closely resembles that of flavoprotein pyridine nucleotide reductases. All these proteins show, in addition to the DBM, several typical features (some already noted in Rao and Rossmann[2]) that are not present in other Rossman-type proteins: (a) an excursion into another (substrate-binding) domain after the 2nd strands; (b) a three-stranded antiparallel β-sheet between strands 3 and 4, often completed by

an additional strand at the N-terminus of the protein (this antiparallel β-sheet is absent in TMD and DAO, which have an α-helix instead); (c) another excursion into the substrate-binding domain after the 4th strand or 4th helix; (d) a 5th helix folding back not behind the β-sheet but in front (in our conventional orientation), i.e., next to the 1st helix; (e) a missing 6th strand, often replaced by an antiparallel β-strand found just before the 5th strand in the sequence. These structural similarities suggest that all these flavoproteins have evolved from a common ancestor, which was already quite differentiated from the basic Rossman-type protein.

Among these features of flavoproteins, it is remarkable that one of the most conspicuous, the antiparallel β-sheet between the 3rd and 4th strands, is also present, with the same connectivity, in the NAD(P)H-binding domain of flavoprotein pyridine nucleotide reductases. We have found no other example of a β-sheet connecting strands 3 and 4 in other NAD(P)H-binding proteins. In this domain, the incursion (a) described above is missing and the fifth strand and following helix in the central parallel β-sheet are missing (they are often replaced by an additional strand and helix at the N-terminus of the domain). In short, the NAD(P)H-domain appears as a truncated version of the FAD-domain. This, together with the presence of an ATG motif (see above) suggests that this domain has evolved from the same branch that gave the FAD-binding domain (step F in Fig. 5). In conclusion, we propose that flavoprotein pyridine nucleotide reductases have arisen by insertion into an ancestral FAD-binding domain (after the 4th strand) of an NAD(P)H-binding domain, itself evolved from the same ancestor. This scheme is somewhat more complex than the simple gene duplication event proposed before,[49] but it appears more able to explain the structural similarities between the available structures. In most flavoproteins, the bulk of the substrate-binding domain is inserted after the 4th β-strand or the 4th α-helix, the rest being found after the 2nd strand. When the structure of this additional domain is examined (i.e., using superimposition of Cα traces in FSSP), two main branches appear, probably reflecting different insertion events (steps H and I). In flavoprotein pyridine nucleotide reductases and related proteins, as stated above, the inserted domain is a classical dinucleotide-binding domain found immediately after the 4th strand (ATG). In PHH, GDI, and DAO, insertion occurs after the 4th helix. The inserted domain consists of a 7-stranded β-sheet, with mostly antiparallel orientation, whose complex topology is perfectly identical in the three proteins. As noted for COX,[38] the substrate-binding domain of GMC oxidoreductases adopts a similar topology, albeit with additional secondary structure elements. We believe that this is evidence for a common origin for PHH, GDI, DAO, and GMC oxidoreductases, and that other families of flavoproteins will turn out to be structurally and phyllogenetically related.

ACKNOWLEDGMENTS

The author thanks J.M. Camadro, Institut Jacques Monod, and D. Picot, Institut de Biologie Physico-Chimique, for their help with the sequence comparison and structural analysis programs, respectively.

REFERENCES

1. Rossmann MG, Moras D, Olsen KW. Chemical and biological evolution of nucleotide-binding protein. Nature 1974;250:194–199.
2. Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. J Mol Biol 1973;76:241–256.
3. Wierenga RK, Drenth J, Schulz GE. Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of p-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. J Mol Biol 1983;167:725–739.
4. Schulz GE, Schirmer RH, Sachsenheimer W, Pai EF. The structure of the flavoenzyme glutathione reductase. Nature 1978;273:120–124.
5. Schulz GE. Binding of nucleotides by proteins. Curr Opin Struct Biol 1992;2:61–67.
6. Ohlsson I, Nordstrom B, Branden CI. Structural and functional similarities within the coenzyme binding domains of dehydrogenases. J Mol Biol 1974;89:339–354.
7. Wierenga RK, Terpstra P, Hol WG. Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. J Mol Biol 1986;187:101–107.
8. Bork P, Grunwald C. Recognition of different nucleotide-binding sites in primary structures using a property-pattern approach. Eur J Biochem 1990;191:347–358.
9. Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold: implications for nucleotide specificity [published erratum appears in J Mol Biol 1993 Aug 5;232:1012]. J Mol Biol 1992;228:662–671.
10. Scrutton NS, Berry A, Perham RN. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 1990;343:38–43.
11. Eggink G, Engel H, Vriend G, Terpstra P, Witholt B. Rubredoxin reductase of *Pseudomonas oleovorans*: structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. J Mol Biol 1990;212:135–142.
12. Hecht HJ, Erdmann H, Park HJ, Sprinzl M, Schmid RD. Crystal structure of NADH oxidase from *Thermus thermophilus*. Nat Struct Biol 1995;2:1109–1114.
13. Benson TE, Filman DJ, Walsh CT, Hogle JM. An enzyme-substrate complex involved in bacterial cell wall biosynthesis. Nat Struct Biol 1995;2:644–653.
14. Vallon O, Wollman F-A. cDNA sequence of Mα, the catalytic subunit of the *Chlamydomonas reinhardtii* L-amino acid oxidase (Accession No. U78797): a new sequence motif shared by a wide variety of flavoproteins (PGR97–171). Plant Physiol 1997;115:1729.
15. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
16. Schuler GD, Altschul SF, Lipman DJ. A workbench for multiple alignment construction and analysis. Proteins 1991;9:180–190.
17. 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.
18. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. Trends Biochem Sci 1995;20:374.
19. Wild DL, Tucker PA, Choe S. A visual data flow environment for macromolecular crystallographic computing [published erratum appears in J Mol Graph 1995 Dec;13:385]. J Mol Graph 1995;13:291–298, 299–300.
20. Morris CJ, Black AC, Pealing SL, et al. Purification and properties of a novel cytochrome: flavocytochrome c from *Shewanella putrefaciens*. Biochem J 1994;302:587–593.
21. Heim S, Kunkel A, Thauer RK, Hedderich R. Thiol:fumarate reductase (Tfr) from *Methanobacterium thermoautotrophicum*—identification of the catalytic sites for fumarate reduction and thiol oxidation. Eur J Biochem 1998;253:292–299.
22. Pealing SL, Black AC, Manson FD, Ward FB, Chapman SK, Reid GA. Sequence of the gene encoding flavocytochrome c from

Shewanella putrefaciens: a tetraheme flavoenzyme that is a soluble fumarate reductase related to the membrane-bound enzymes from other bacteria [published erratum appears in Biochemistry 1993 Apr 13; 32:3829]. Biochemistry 1992;31:12132–12140.
23. Robinson KM, Rothery RA, Weiner JH, Lemire BD. The covalent attachment of FAD to the flavoprotein of *Saccharomyces cerevisiae* succinate dehydrogenase is not necessary for import and assembly into mitochondria. Eur J Biochem 1994;222:983–990.
24. Tedeschi G, Negri A, Mortarino M, et al. L-aspartate oxidase from *Escherichia coli*. II. Interaction with C4 dicarboxylic acids and identification of a novel L-aspartate: fumarate oxidoreductase activity. Eur J Biochem 1996;239:427–433.
25. Mortarino M, Negri A, Tedeschi, et al. L-aspartate oxidase from *Escherichia coli*. I. Characterization of coenzyme binding and product inhibition. Eur J Biochem 1996;239:418–426.
26. Franklund CV, Baron SF, Hylemon PB. Characterization of the baiH gene encoding a bile acid-inducible NADH:flavin oxidoreductase from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1993;175: 3002–3012.
27. Lim LW, Shamala N, Mathews FS, Steenkamp DJ, Hamlin R, Xuong NH. Three-dimensional structure of the iron-sulfur flavoprotein trimethylamine dehydrogenase at 2.4-A resolution. J Biol Chem 1986;261:15140–15146.
28. Barber MJ, Neame PJ, Lim LW, White S, Matthews FS. Correlation of x-ray deduced and experimental amino acid sequences of trimethylamine dehydrogenase. J Biol Chem 1992;267:6611–6619.
29. Lim LW, Mathews FS. Steenkamp DJ. Identification of ADP in the iron-sulfur flavoprotein trimethylamine dehydrogenase. J Biol Chem 1988;263:3075–3078.
30. Creighton TE: Proteins: structures and molecular properties. 1993.
31. Holm L, Sander C. Touring protein fold space with Dali/FSSP. Nucleic Acids Res 1998;26:318–321.
32. Ross RP, Claiborne A. Molecular cloning and analysis of the gene encoding the NADH oxidase from *Streptococcus faecalis* 10C1: comparison with NADH peroxidase and the flavoprotein disulfide reductases. J Mol Biol 1992;227:658–671.
33. Van Driessche G, Koh M, Chen ZW, et al. Covalent structure of the flavoprotein subunit of the flavocytochrome c: sulfide dehydrogenase from the purple phototrophic bacterium *Chromatium vinosum*. Protein Sci 1996;5:1753–1764.
34. Chen ZW, Koh M, Van Driessche G, et al. The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. Science 1994;266:430–432.
35. Schalk I, Zeng K, Wu SK, et al. Structure and mutational analysis of Rab GDP-dissociation inhibitor. Nature 1996;381:42–48.
36. Koonin EV. Human choroideremia protein contains a FAD-binding domain [published erratum appears in Nat Genet 1996 Apr;12:458]. Nat Genet 1996;12:237–239.
37. Wu SK, Zeng K, Wilson IA, Balch WE. Structural insights into the function of the Rab GDI superfamily. Trends Biochem Sci 1996;21: 472–476.
38. Vrielink A, Lloyd LF, Blow DM. Crystal structure of cholesterol oxidase from *Brevibacterium sterolicum* refined at 1.8 Å resolution. J Mol Biol 1991;219:533–554.
39. Hecht HJ, Kalisz HM, Hendle J, Schmid RD, Schomburg D. Crystal structure of glucose oxidase from *Aspergillus niger* refined at 2.3 Å resolution. J Mol Biol 1993;229:153–172.
40. Cavener DR. GMC oxidoreductases: a newly defined family of homologous proteins with diverse catalytic activities. J Mol Biol 1992;223:811–814.
41. Li J, Vrielink A, Brick P, Blow DM. Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. Biochemistry 1993;32:11507–11515.
42. Becker K, Muller S, Keese MA, Walter RD, Schirmer RH. A glutathione reductase-like flavoenzyme of the malaria parasite *Plasmodium falciparum*: structural considerations based on the DNA sequence. Biochem Soc Trans 1996;24:67–72.
43. Entsch B, van Berkel WJ. Structure and mechanism of para-hydroxybenzoate hydroxylase. FASEB J 1995;9:476–483.
44. Enroth C, Neujahr H, Schneider G, Lindqvist Y. The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. Structure. 1998;6:605–617.
45. Mizutani H, Miyahara I, Hirotsu K, et al. Three-dimensional structure of porcine kidney D-amino acid oxidase at 3.0 Å resolution. J Biochem (Tokyo) 1996;120:14–17.
46. Labesse G, Vidal-Cros A, Chomilier J, Gaudry M, Mornon JP. Structural comparisons lead to the definition of a new superfamily of NAD(P)(H)-accepting oxidoreductases: the single-domain reductases/epimerases/dehydrogenases (the 'RED' family). Biochem J 1994;304:95–99.
47. Hol WG, van Duijnen PT, Berendsen H. The alpha-helix dipole and the properties of proteins. Nature 1978;273:443–446.
48. Wierenga RK, De Maeyer MCH, Hol WG. Interaction of pyrophosphate moietites with α-helices in dinucleotide binding proteins. Biochemistry 1985;24:1346–1357.
49. McKie JH, Douglas KT. Evidence for gene duplication forming similar binding folds for NAD(P)H and FAD in pyridine nucleotide-dependent flavoenzymes. FEBS Lett. 1991;279:5–8.
50. Bach AW, Lan NC, Johnson DL, et al. cDNA cloning of human liver monoamine oxidase A and B: molecular basis of differences in enzymatic properties. Proc Natl Acad Sci USA 1988;85:4934–4938.
51. Schroder I, Gunsalus RP, Ackrell BA, Cochran B, Cecchini G. Identification of active site residues of Escherichia coli fumarate reductase by site-directed mutagenesis. J Biol Chem 1991;266: 13572–13579.
52. Mallonee DH, White WB, Hylemon PB. Cloning and sequencing of a bile acid-inducible operon from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1990;172:7011–7019.
53. He XY, Yang SY, Schulz H. Cloning and expression of the fadH gene and characterization of the gene product 2,4-dienoyl coenzyme A reductase from *Escherichia coli*. Eur J Biochem 1997;248: 516–520.
54. Niedermann DM, Lerch K. Molecular cloning of the L-amino-acid oxidase gene from *Neurospora crassa*. J Biol Chem 1990;265: 17246–17251.
55. Bockholt R, Masepohl B, Kruft V, Wittmann-Liebold B, Pistorius EK. Partial amino acid sequence of an L-amino acid oxidase from the cyanobacterium *Synechococcus* PCC6301, cloning and DNA sequence analysis of the aoxA gene. Biochim Biophys Acta 1995; 1264:289–293.
56. Raibekas AA, Massey V. Primary structure of the snake venom L-amino acid oxidase shows high homology with the mouse B cell interleukin 4-induced Fig1 protein. Biochem Biophys Res Commun 1998;248:476–478.
57. Chu CC, Paul WE. Fig1, an interleukin 4-induced mouse B cell gene isolated by cDNA representational difference analysis. Proc Natl Acad Sci USA 1997;94:2507–2512.
58. Obara K, Otsuka-Fuchino H, Sattayasai N, Nonomura Y, Tsuchiya T, Tamiya T. Molecular cloning of the antibacterial protein of the giant African snail, *Achatina fulica* Ferussac. Eur J Biochem 1992;209:1–6.
59. Takamatsu N, Shiba T, Muramoto K, Kamiya H. Molecular cloning of the defense factor in the albumen gland of the sea hare *Aplysia kurodai*. FEBS Lett 1995;27:373–376.
60. Malloy PJ, Zhao X, Madani ND, Feldman D. Cloning and expression of the gene from *Candida albicans* that encodes a high-affinity corticosteroid-binding protein. Proc Natl Acad Sci USA 1993;90:1902–1906.
61. Joets J, Pousset D, Marcireau C, Karst F. Characterization of the *saccharomyces cerevisiae* FMS1 gene related to *Candida albicans* corticosteroid-binding protein 1. Curr Genet 1996;30:115–120.
62. Ishizuka H, Horinouchi S, Beppu T. Putrescine oxidase of *Micrococcus rubens*: primary structure and *Escherichia coli*. J Gen Microbiol 1993;139:425–432.
63. Schilling B, Lerch K. Cloning, sequencing and heterologous expression of the monoamine oxidase gene from *Aspergillus niger*. Mol Gen Genet 1995;247:430–438.
64. Camilleri C, Jouanin L. The TR-DNA region carrying the auxin synthesis genes of the *Agrobacterium rhizogenes* agropine-type plasmid pRiA4: nucleotide sequence analysis and introduction into tobacco plants. Mol Plant-Microbe Interact 1991;4:155–162.
65. Emanuele JJ, Heasley CJ, Fitzpatrick PF. Purification and characterization of the flavoprotein tryptophan 2-monooxygenase expressed at high levels in *Escherichia coli*. Arch Biochem Biophys 1995;316:241–248.

66. Hugueney P, Romer S, Kuntz M, Camara B. Characterization and molecular cloning of a flavoprotein catalyzing the synthesis of phytofluene and zeta-carotene in *Capsicum chromoplasts*. Eur J Biochem 1992;209:399–407.
67. Albrecht M, Linden H, Sandmann G. Biochemical characterization of purified zeta-carotene desaturase from *Anabaena* PCC 7120 after expression in *Escherichia coli*. Eur J Biochem 1996;236: 115–120.
68. Ehrenshaft M, Daub ME. Isolation, sequence, and characterization of the *Cercospora nicotianae* phytoene dehydrogenase gene. Appl Environ Microbiol 1994;60:2766–2771.
69. Gari E, Toledo JC, Gibert I, Barbe J. Nucleotide sequence of the methoxyneurosporene dehydrogenase gene from *Rhodobacter sphaeroides*: comparison with other bacterial carotenoid dehydrogenases. FEMS Microbiol Lett 1992;72:103–108.
70. Wieland B, Feil C, Gloria-Maercker E. Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of *Staphylococcus aureus*. J Bacteriol 1994;176: 7719–7726.
71. Camadro JM, Labbe P. Cloning and characterization of the yeast HEM14 gene coding for protoporphyrinogen oxidase, the molecular target of diphenyl ether-type herbicides. J Biol Chem 1996;271: 9120–9128.
72. Dailey TA, Meissner P, Dailey HA. Expression of a cloned protoporphyrinogen oxidase. J Biol Chem 1994;269:813–815.
73. Ladeveze V, Marcireau C, Delourme D, Karst F. General resistance to sterol biosynthesis inhibitors in *Saccharomyces cerevisiae*. Lipids 1993;28:907–912.
74. Manetti AG, Rosetto M, Maundrell KG. nmt2 of fission yeast: a second thiamine-repressible gene co-ordinately regulated with nmt1. Yeast 1994;10:1075–1082.
75. Praekelt UM, Meacock PA. MOL1, a *Saccharomyces cerevisiae* gene that is highly expressed in early stationary phase during growth on molasses. Yeast 1992;8:699–710.
76. Choi GH, Marek ET, Schardl CL, Richey MG, Chang SY, Smith DA. sti35, a stress-responsive gene in *Fusarium* spp. J Bacteriol 1990;172:4522–4528.
77. Machado CR, de Oliveira RL, Boiteux S, Praekelt UM, Meacock PA, Menck CF. Thi1, a thiamine biosynthetic gene in *Arabidopsis thaliana*, complements bacterial defects in DNA repair. Plant Mol Biol 1996;31:585–593.
78. Zhang L, al-Hendy A, Toivanen P, Skurnik M. Genetic organization and sequence of the rfb gene cluster of *Yersinia enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. Mol Microbiol 1993;9:309–321.
79. Yao Z, Valvano MA. Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. J Bacteriol 1994;176:4133–4143.
80. Enomoto K, Ohki R, Muratsubaki H. Cloning and sequencing of the gene encoding the soluble fumarate reductase from *Saccharomyces cerevisiae*. DNA Res 1996;3:263–267.
81. Plesiat P, Grandguillot M, Harayama S, Vragar S, Michel-Briand Y. Cloning, sequencing, and expression of the *Pseudomonas testosteroni* gene encoding 3-oxosteroid delta 1-dehydrogenase. J Bacteriol 1991;173:7219–7227.
82. Florin C, Kohler T, Grandguillot M, Plesiat P. *Comamonas testosteroni* 3-ketosteroid-delta4(5alpha)-dehydrogenase: gene and protein characterization. J Bacteriol 1996;178:3322–3330.
83. Miller RE, Stadtman ER. Glutamate synthase from *Escherichia coli* : an iron-sulfide flavoprotein. J Biol Chem 1972;247:7407–7419.
84. Gregerson RG, Miller SS, Twary SN, Gantt JS, Vance CP. Molecular characterization of NADH-dependent glutamate synthase from alfalfa nodules. Plant Cell 1993;5:215–226.
85. Curti B, Vanoni MA, Verzotti E, Zanetti G. Glutamate synthase: a complex iron-sulphur flavoprotein. Biochem Soc Trans 1996;24: 95–99.
86. Gosset G, Merino E, Recillas F, Oliver G, Becerril B, Bolivar F. Amino acid sequence analysis of the glutamate synthase enzyme from *Escherichia coli* K-12. Protein Seq Data Anal 1989;2:9–16.
87. Cavicchioli R, Kolesnikow T, Chiang RC, Gunsalus RP. Characterization of the aegA locus of *Escherichia coli*: control of gene expression in response to anaerobiosis and nitrate. J Bacteriol 1996;178:6968–6974.
88. Albin N, Johnson MR, Diasio RB. cDNA cloning of bovine liver dihydropyrimidine dehydrogenase. DNA Seq 1996;6:243–250.
89. Lawton MP, Cashman JR, Cresteil T, et al. A nomenclature for the mammalian flavin-containing monooxygenase gene family based on amino acid sequence identities. Arch Biochem Biophys 1994;308: 254–257.
90. Atta-Asafo-Adjei E, Lawton MP, Philpot RM. Cloning, sequencing, distribution, and expression in *Escherichia coli* of flavin-containing monooxygenase 1C1: evidence for a third gene subfamily in rabbits. J Biol Chem 1993;268:9681–9689.
91. Suh JK, Poulsen LL, Ziegler DM, Robertus JD. Molecular cloning and kinetic characterization of a flavin-containing monooxygenase from *Saccharomyces cerevisiae*. Arch Biochem Biophys 1996; 336:268–274.
92. Chen YC, Peoples OP, Walsh CT. *Acinetobacter* cyclohexanone monooxygenase: gene cloning and sequence determination. J Bacteriol 1988;170:781–789.
93. Brown DW, Yu JH, Kelkar HS, et al. Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. Proc Natl Acad Sci USA 1996;93:1418–1422.
94. Liu XL, Scopes RK. Cloning, sequencing and expression of the gene encoding NADH oxidase from the extreme anaerobic thermophile *Thermoanaerobium brockii*. Biochim Biophys Acta 1993;1174: 187–190.
95. Boyd G, Mathews FS, Packman LC, Scrutton NS. Trimethylamine dehydrogenase of bacterium W3A1: molecular cloning, sequence determination and over-expression of the gene. FEBS Lett 1992; 308:271–276.
96. Yang CC, Packman LC, Scrutton NS. The primary structure of *Hyphomicrobium* X dimethylamine dehydrogenase: relationship to trimethylamine dehydrogenase and implications for substrate recognition. Eur J Biochem 1995;232:264–271.
97. Huang L, Rohlfs RJ, Hille R. The reaction of trimethylamine dehydrogenase with electron transferring flavoprotein. J Biol Chem 1995;270:23958–23965.
98. Pelanda R, Vanoni MA, Perego M, et al. Glutamate synthase genes of the diazotroph *Azospirillum brasilense*: cloning, sequencing, and analysis of functional domains. J Biol Chem 1993;268: 3099–3106.
99. Krueger JK, Stock J, Schutt CE. Evidence that the methylesterase of bacterial chemotaxis may be a serine hydrolase. Biochim Biophys Acta 1992;1119:322–326.