Six new candidate members of the α/β twisted open-sheet family detected by sequence similarity to flavodoxin

RITA GRANDORI AND JANNETTE CAREY

Chemistry Department, Princeton University, Princeton, New Jersey 08544

(RECEIVED May 18, 1994; ACCEPTED September 22, 1994)

Abstract

Strong sequence similarity has been reported among WrbA (the Trp repressor-binding protein of *Escherichia coli*); Ycp4, a protein of unknown function from the budding yeast *Saccharomyces cerevisiae*; P25, the pap1-dependent protein of the fission yeast *Schizosaccharomyces pombe*; and the translation product of a partial cDNA sequence from rice seedling root (*Oryza sativa*, locus Ricr02421a; here referred to as RicR). Further homology search with the profile method indicates that all the above sequences are related to the flavodoxin family and, in turn, allows detection of the recently proposed flavodoxin-like proteins from *E. coli*, MioC and the hypothetical protein YihB. We discuss sequence conservation with reference to the known 3-dimensional structures of flavodoxins. Conserved sequence and hydrophobicity patterns, as well as residue-pair interaction potentials, strongly support the hypothesis that these proteins share the α/β twisted open-sheet fold typical of flavodoxins, with an additional α/β unit in the WrbA family. On the basis of the proposed structural homology, we discuss the details of the putative FMN-binding sites. Our analysis also suggests that the helix-turn-helix motif we identified previously in the C-terminal region of the WrbA family is unlikely to reflect a DNA-binding function of this new protein family.

Keywords: FMN-binding; profile analysis; residue-pair interaction potential; response regulators; structural homology; WrbA family

The recently identified WrbA family of proteins exhibits high sequence conservation (~50% identity) in very distantly related organisms (Grandori & Carey, 1994a), a fact not yet accounted for by the available information about the functions of these proteins. WrbA is induced during the stationary phase of *Escherichia coli* and enhances the affinity of the tryptophan repressor for one of its operator sequences (Yang et al., 1993). The fission yeast protein P25 is the target of a complex transcriptional control system (Toda et al., 1992) that involves the AP-1-like factor PAP1 as positive regulator and *CRM1* as negative regulator. Overexpression of P25 in cold-sensitive *crm1* mutants correlates with the acquisition of staurosporine resistance and aberrant chromatin structure, but it is not known whether P25 mediates these responses; P25 also may be involved in brefeldin A resistance (Turi TG, 1993, Genbank entry X73558). No functional characterization is available for Ycp4, the member from budding yeast, although its sequence contains a C-terminal extension, relative to the other members of the family, bearing

the signals for protein farnesylation and carboxy-methylation (Grandori & Carey, 1994a). RicR appears to be a development-specific product in rice and its mRNA is expressed in seedling roots (Minobe Y, 1994, Genbank entry D23815). Only a partial sequence from cDNA is available.

Sequence motif analysis originally revealed the presence of a helix-turn-helix (HTH) signature in the C-terminal region of WrbA (Grandori & Carey, 1994a) that suggested a possible DNA-binding function for this family. Koonin and coworkers (Koonin & Bork, 1994; Koonin et al., 1994) noted a conserved N-terminal sequence motif characteristic of flavodoxins that suggests an FMN-binding function for the family. Here we report results of sequence analysis of the WrbA and flavodoxin families by means of different methods based on the complementary approaches of sensitive detection of sequence similarity (Grib-skov et al., 1987, 1990) and homology-independent fold recognition (Sippl & Weitckus, 1992; Sippl, 1993a). We analyze the structural and functional relevance of the sequence similarities to flavodoxins for the WrbA family and 2 other new flavodoxin-like proteins from *E. coli*. The results suggest that this family shares the flavodoxin fold, and that the HTH motif is unlikely to be part of an independent DNA-binding domain.

Reprint requests to: Rita Grandori or Jannette Carey, Chemistry Department, Princeton University, Princeton, New Jersey 08544; e-mail: rita@chemvax.princeton.edu or carey@chemvax.princeton.edu.

Results

Homology search

By screening the Swissprot database with a profile derived (see Materials and methods) from the aligned sequences of WrbA, Ycp4, and P25, we detected weak but significant sequence similarities to flavodoxins (around 25% identity, *Z* scores 3–11). The large number of available flavodoxin sequences offers a good starting set for the reciprocal screen. Using a profile derived from the flavodoxin family, the 3 sequences rank (*Z* scores 4.9–7.3) among or above known FMN-binding proteins (*Z* scores 4.3–8.4) whose homology to flavodoxins has been already discussed: the NADPH-cytochrome P450 reductase (Porter et al., 1990), the nitric-oxide synthase (Bredt et al., 1991), the sulfite reductase α -component (Ostrowski et al., 1989), and the flavodoxin-like protein of *Rhodobacter capsulatus* (Jouanneau Y, 1990, Genbank entry X54054; Jouanneau et al., 1990).

Database search with the flavodoxin profile also detects 2 other candidate flavodoxin-like proteins from *E. coli*, MioC (Løbner-Olesen & Boye, 1992) (*Z* score 11) and the hypothetical protein YihB (Daniels et al., 1992) (*Z* score 7.8), showing less than 20% sequence identity with each other and with the WrbA family. MioC scores as the best non-flavodoxin sequence of the database. It has been implicated in the control of *oriC* initiation on minichromosomes (de Wind et al., 1987), but its mechanism of action is not known. No experimental characterization is available for YihB; its sequence is reported to match the flavodoxin signature ((L,I,V)(L,I,V,F,Y)(F,Y)X(S,T)X₂(A,G)XT-X₃A X₂(L,I,V)) in the PROSITE dictionary of protein patterns (Bairoch, 1992). This signature has been deduced from sequence conservation among flavodoxins in their N-terminal region spanning the first loop of the FMN-binding site. MioC and the WrbA family (Fig. 1), as well as other known FMN-binding proteins (data not shown), show 1 or 2 mismatches to this consensus sequence. A more flexible method for motif search based on position-dependent weight matrices detects significant conservation of an N-terminal sequence pattern among flavodoxins, other FMN-binding proteins, MioC, YihB, and the WrbA family (Koonin & Bork, 1994; Koonin et al., 1994). FASTA (Pearson & Lipman, 1988) does not detect (best 100 scores) sequence similarity to flavodoxins or flavodoxin-like proteins for Ycp4, P25, and YihB.

Sequence comparison

To evaluate the structural and functional relevance of the detected sequence similarities, we referred to 6 known structures of flavodoxins (Burnett et al., 1974; Smith et al., 1983; van Mierlo et al., 1990; Watt et al., 1991; Fukuyama et al., 1992; Rao et al., 1992). The common feature of the flavodoxin fold is a central 5-stranded parallel hydrophobic β -sheet sandwiched between 2 pairs of amphipathic α -helices (Fig. 2). The fifth β -strand is interrupted by 4 residues in 2 short-chain flavodoxins (Burnett et al., 1974; Watt et al., 1991) or by a long loop insertion in 3 long-chain flavodoxins (Smith et al., 1983; Fukuyama et al., 1992; Rao et al., 1992). A poorly conserved 3₁₀-helix is located between β 2 and β 3. The fold correlates with a highly conserved, repetitive sequence pattern in which hydrophobic residues cluster in β -strands and have a 3–4-residue periodicity in α -helices; the correlation of this pattern with solvent accessibil-

ity in the 3-dimensional structure (Bowie et al., 1990) led to detection of the almost identical CheY fold (Stock et al., 1989a; Volz & Matsumura, 1991) in the absence of any significant sequence similarity.

Figure 1 shows multiple sequence alignment of flavodoxins and the 6 candidate flavodoxin-like proteins with comparison to the consensus sequence derived from the CheY superfamily (Stock et al., 1989b; Volz, 1993). The 6 new sequences show conservation of the hydrophobicity pattern typical of this class of fold. Of 34 conserved hydrophobic positions (# in Fig. 1) in flavodoxins, 31 are conserved in MioC as well as in YihB, and 23 in the WrbA family. Regions corresponding to β 2 and β 5 of flavodoxins and CheY are less hydrophobic, consistent with their role as the external strands of the sheet, and this feature is shared by the other sequences in Figure 1. A quantitative comparison of the hydropathic features of the examined sequences is presented in Figure 3.

Within the group of 6 candidate flavodoxin-like proteins, general sequence similarity to flavodoxins is highest in regions corresponding to the known elements of secondary structure of flavodoxins (Fig. 1). We also note high conservation of residues following β 1 and β 3, and preceding β 4 and β 5 (the available sequence of RicR is truncated at the loop following β 4). Five of the 7 perfectly conserved residues (! in Fig. 1) of flavodoxins (G16, A22, G60, F102, G132; hereafter, residue numbering refers to positions in the alignment, indicated below it) are conserved in the new sequences (with the exception of G/I60 in MioC, and F/I102 in MioC and YihB). Their chemical nature and locations in the structure of known flavodoxins suggest that these 5 residues are probably identical due to structural constraints. Two of these residues are glycines located in tight turns (G16, G132). Interestingly, G132 and the highly conserved A133 correspond to a nearly invariant GA sequence immediately before β 5 in the CheY superfamily. Comparison of 79 members of the CheY class (Volz, 1993) showed that the residue corresponding to position 133 of Figure 1 (A102 in CheY) shows covariance with the residue corresponding to position 136 (Y105 in CheY), resulting in complementary changes in the size of the 2 side chains. In flavodoxins an alanine at position 133 and an aliphatic residue at position 136 are most frequently observed, whereas in the WrbA family a glycine is present at position 133 together with a bulky aromatic residue at position 136. The remaining 2 invariant residues of flavodoxins (T15 and T18) that are not conserved in the new sequences are involved in FMN binding.

Gaps in the alignment of Figure 1 cluster in regions that correspond to loops in the flavodoxin structures. Although the flavodoxin profile used here had reduced gap penalty factors in regions presenting gaps among flavodoxin sequences, the location of insertions and deletions within the WrbA family is the same when it is analyzed separately from flavodoxins (not shown). Due to the high sequence conservation within the WrbA family, this observation provides an independent criterion supporting the assignment of loop regions. Moreover, loops following strand β 1 and helix α 3 that appear to be constrained in flavodoxins also show no length variability in the WrbA family.

We note high sequence conservation in a 24-residue insertion in WrbA, Ycp4, and P25 after β 4. These sequences may form an additional α/β motif, according to the periodicity of hydrophobic residues and prediction of secondary structure (Garnier et al., 1978) (Fig. 1), reminiscent of the N-terminal insertion of

Fig. 1. Multiple sequence alignment of 16 flavodoxins, MioC, YihB, the WrbA family, and the consensus sequence of the CheY superfamily (†, Stock et al., 1989a; Volz, 1993; see Materials and methods). Sequences are named by their Swissprot identifier, except for RicR. The PDB identifier is indicated for the 5 crystal structures referred to in the text. The residues involved in the elements of secondary structure (according to DSSP; Kabsch & Sander, 1983) are underlined on each flavodoxin sequence with known 3-dimensional structure (†, van Mierlo et al., 1990; see Materials and methods). The 5 β -strands and 4 α -helices of the flavodoxin fold are numbered on the top line proceeding from the N-terminus; the poorly conserved 3_{10} -helix between $\beta 2$ and $\beta 3$ is not indicated. Symbols above the alignment indicate sequence conservation inside the flavodoxin family: 100% conserved identities (!); highly conserved positions (#, hydrophobic; :, others) defined by the categories G, A, P, T, S, C, D, E, N, Q, R, K, H, I, V, L, M, F, Y, W (A, R, and H were not considered to mismatch otherwise conserved hydrophobic positions). Highly conserved positions were defined as those where at least 14 flavodoxin sequences present residues of the same category. These positions are highlighted in bold in flavodoxins, as well as in the new sequences that match. Positions that display FMN-side-chain interactions in all the experimental structures of flavodoxins are indicated by Δ (contacts with the isoalloxazine ring) or $\diamond$ (contacts with the phosphate group). Dots under the first group of sequences indicate residues of the flavodoxin signature (according to PROSITE). Asterisks indicate the N- and C-termini. Numbers at the beginning or end of a sequence indicate the first or last residue shown. Numbers within a sequence indicate length of insertions that are not shown. Below the alignment is indicated the insertion after $\beta 4$ in the WrbA family and the secondary structure prediction (Garnier et al., 1978) for each sequence. Organism codes: Entag, *Enterobacter agglomerans*; Klepn, *Klebsiella pneumoniae*; Azovi, *Azotobacter vinelandii*; Anasp, *Anabaena PCC 7120*; Synp7, *Anacystis nidulans*; Synp2, *Agmenellum quadruplicatum*; Ecoli, *Escherichia coli*; Chocr, *Chondrus crispus*; Desgi, *Desulfovibrio gigas*; Dessa, *D. salexigens*; Desvh, *D. vulgaris*; Desde, *D. desulfuricans*; Megel, *Megasphaera elsdenii*; Clomp, *Clostridium MP*; Cloab, *C. acetobutylicum*; Rhoca, *Rhodobacter capsulatus*; Rhoca, *Rhodobacter capsulatus*; human, *Homo sapiens*; rat, *Rattus norvegicus*; yeast, *Saccharomyces cerevisiae*; Schpo, *Schizosaccharomyces pombe*.

Fig. 2. Stereo drawing showing the C α trace and FMN atoms of clostridial flavodoxin (black) and the C α trace of superimposed CheY (gray) generated from the PDB file 4FXN and 3CHY by the program INSIGHTII. Residues S7, M56, W90 contacting FMN from the 3 different loops of the FMN-binding site in 4FXN are labeled and their side chains are shown, as well as the side chains of the 3 aspartic acids defining the acidic pocket at the active site of 3CHY.

Fig. 3. Hydrophilicity plots. Values (according to Kyte & Doolittle, 1982) are averaged over a window of 7 residues. Sequences are gapped as in the alignment of Figure 1; breaks in the plots correspond to gaps in the sequences. Numbers on the x axis correspond to the positions in the alignment.

an additional β -strand in ras relative to CheY and flavodoxin (Stock et al., 1991). The boundaries of this insertion are ambiguous, being shifted by a few residues in the comparison of each sequence to the flavodoxin profile. In Figure 1, we set the boundaries of this insertion to point out the presence of a tyrosine conserved in the WrbA family that can be aligned to the aromatic residue interacting with FMN in flavodoxins (position 111). Another long insertion in the WrbA family follows β 2, but in this region the different members of the family do not show high sequence similarity. YihB presents a long insertion after α 2 of the same length and at the same position as the nitric-oxide synthase from brain compared to the NADPH-cytochrome P450 reductase (Bredt et al., 1991) and to flavodoxins (not shown).

The most ambiguous region of the alignment is the one that corresponds to α 4 of flavodoxins. The corresponding region of the WrbA family contains the previously identified HTH motif (Grandori & Carey, 1994a). Individual comparison to the flavodoxin profile generates inconsistent alignments for WrbA, Ycp4, and P25. In Figure 1 we show the frame of alignment that is best for WrbA.

Compatibility test

As an additional test of the hypothesis that the 6 proteins examined here have a flavodoxin fold, we used a recently developed knowledge-based potential function for residue-pair interactions (Sippl & Weitckus, 1992; Sippl, 1993a) to evaluate the unoptimized structural models implied by the alignment in Figure 1. These potentials are derived from empirical frequency distributions of $C\beta$ - $C\beta$ distances on the basis of Boltzmann's principle in order to describe the mean force field acting in a natively folded structure. The potentials are used by the program PROSAIL to derive a position-dependent energy profile and the global energy value of a structural model for a given sequence. Because the potentials are derived from $C\beta$ - $C\beta$ distances, the method evaluates compatibility between amino acid sequence and main-chain topology independently of assumptions about the conformation of the side chains. The method has been shown to be sensitive to deliberately misfolded proteins and incorrectly modeled regions within experimentally determined native structures (Sippl, 1993a). Such regions correlate with positive peaks of the energy function, whereas correct structures typically give all-negative functions fluctuating between -2 and -0.5 kcal/mol.

Figure 4A shows the energy profile for residue-pair interactions in a native flavodoxin structure (4FXN) and in each of the 6 test sequences threaded onto the flavodoxin fold on the basis of the alignment in Figure 1. Although the test profiles are all shifted to higher energy than the flavodoxin profile, in no case do they show regions of unfavorable energy values. A noteworthy feature of the energy profiles is the remarkable similarity in overall shape to the flavodoxin control. Figure 4B shows pair-energy profiles for positive and negative control sequences using 4FXN as the structural template. The positive control of CheY threaded onto the flavodoxin fold on the basis of structural superposition (Fig. 2) shows an energy profile comparable to those of the 6 models in Figure 4A. As negative control sequences we used the histone-like protein of *Bacillus stearothermophilus* (hu) and the chloramphenicol acetyl transferase III from *E. coli* (cat3), which are known to fold distinctively from flavodoxin (Tanaka et al., 1984; Leslie, 1990), but which are still selected from the sequence database by the flavodoxin profile

(Z scores 4.91 and 2.6, respectively). Their sequences were threaded onto the flavodoxin fold on the basis of their alignment with the flavodoxin profile (not shown). The derived pair-energy profiles are almost all-positive functions, fluctuating mainly between 0 and 0.5 kcal/mol. The controls shown in Figure 4C use hypothetical structural templates for WrbA different from flavodoxins. We considered annexin V from chicken (Bewley et al., 1993, PDB: 1ALA) and endonuclease V from T4 bacteriophage (Morikawa et al., 1992; PDB: 1END), which are selected by screening the sequence database with the profile derived from the WrbA family (Z scores 3.6 and 2.58, respectively). The WrbA sequence folded as 1ALA or 1END on the basis of sequence-profile alignment (not shown) gives almost all-positive pair-energy profiles, whereas the native sequences of the templates are all-negative, indicating that the structures do not present strained or faulty parts.

The global energy value calculated for a structural model can be used to derive the associated Z score by comparing the global energies of alternative structural templates for the same sequence. The program PROSAIL tests all the possible conformations generated by sliding the input sequence along a hypothetical polypeptide of known structural domains. Z scores range between -5 and -10 for native proteins of about 150 residues, whereas incorrectly folded proteins typically give values between -2 and 5 kcal/mol (Sippl, 1993a). Table 1 reports the Z scores from the global energy values of the 6 sequences folded in the flavodoxin conformation, and those of the controls described above. Z scores for $C\beta$ - $C\beta$ interactions cluster near -5 , whereas the positive control of CheY, which is known to have the flavodoxin fold, has a value of only -3.25 . The Z scores derived from surface potentials (Sippl, 1993b) and from their combination with pair-interaction potentials are also shown in Table 1. For each

Table 1. Z scores for the global energy values of 4FXN, the 6 test sequences folded as 4FXN, and positive and negative controls^a

	Pair	Surf	Comb
4FXN	-9.00	-8.35	-12.21
MioC/4FXN	-5.67	-6.18	-8.24
YihB/4FXN	-5.10	-6.36	-8.41
WrbA/4FXN	-5.30	-4.76	-7.11
Ycp4/4FXN	-4.99	-5.53	-7.38
P25/4FXN	-4.18	-4.65	-6.18
RicR/4FXN	-4.77	-6.09	-7.76
CheY/4FXN	-3.25	-5.01	-6.10
hu/4FXN	-0.55	-2.79	-2.29
cat3/4FXN	-1.23	-2.27	-2.27
1END	-7.65	-4.75	-9.21
WrbA/1END	-2.26	-1.65	-2.70
1ALA	-9.01	-6.69	-10.83
WrbA/1ALA	-1.29	-2.17	-2.38

^a Native structures are indicated only by their PDB identifier, whereas models are indicated by the sequence name followed by the PDB identifier of the structure used as a template. Values were calculated from database screening with the program PROSAIL (Sippl, 1993a). Z scores from pair interaction energy (pair), surface energy (surf), and combined energy (comb) (Sippl, 1993b). In the calculation of combined energy, surface energy values were weighted by the factor $\sigma_{\text{pair}}/\sigma_{\text{surf}}$ as a function of the standard deviations of pair- and surface-energy distributions.

Fig. 4. Position-dependent energy profiles calculated from residue-pair ($C\beta$ - $C\beta$) interactions with the program PROSAIL (Sippl, 1993a). Data are smoothed with a window size of 50 residues. Pair energy is expressed in units of E/kT , where E is energy in kcal/mol, k is Boltzmann's constant, and T is absolute temperature. Residue numbers correspond to the known structures. **A:** 4FXN and the 6 sequences examined here threaded onto the 4FXN structure according to the alignment of Figure 1: —, 4FXN (native sequence); ·····, MioC; - - - -, YihB; — · — ·, WrbA; ·····, P25; — — —, Ycp4; - - - -, RicR. **B:** positive and negative control sequences threaded onto the 4FXN structural template: —, 4FXN (native sequence); ·····, CheY; — · — ·, hu; - - - -, cat3. **C:** Positive and negative control sequences threaded onto structural templates other than flavodoxin: —, 1ALA (native sequence); — · — ·, WrbA/1ALA; - - - -, 1END (native sequence); ·····, WrbA/1END.

tested sequence in the frame of alignment given in Figure 1, 4FXN results as the best match by screening the PROSAIL library of protein conformations. However, because the program does not allow for insertions and deletions, the input sequences were already deleted of the regions corresponding to insertions relative to 4FXN, and residues of 4FXN corresponding to gaps in the aligned sequences were omitted from the calculation by the designation "unknown." Thus, these data can be interpreted as a strong indication for structural significance of the proposed alignment but not as confirmation that the proposed conformation is the optimal one. Nevertheless, we were able to use the same method to test alternative frames of alignment of the 6 test sequences relative to flavodoxin. Due to the longer sequence of YihB and the WrbA family, there are some ambiguities in the positions of insertions, particularly in the less conserved C-terminal region. Alternative alignments obtained with different gap penalties would assign the sequence YihB (81)PSAFYSV(87) to $\beta 4$ and (126)LRYPRYRWY(134) to $\beta 5$ (YihB numbering) and the region of the WrbA family (136)VIVPIAYAA(144) to $\beta 5$

(WrbA sequence and numbering). The alignment shown in Figure 1 implies the best energy profiles.

The FMN-binding site

Flavodoxins bind FMN with very high affinity ($K_a \sim 10^9 \text{ M}^{-1}$) (Mayhew & Tollin, 1992). The FMN molecule is oriented with the isoalloxazine ring toward the surface and the ribityl tail pointing toward the interior of the protein, positioning the phosphate group close to the N-terminus of $\alpha 1$. Three loops of the protein define the FMN-binding site. We inspected residue conservation in the 6 new sequences at the critical positions for FMN-protein interactions to evaluate the possibility of predicting an FMN-binding site.

In the known flavodoxin structures, the loop after $\beta 1$ interacts with the ribityl tail. Three positions are involved in side-chain-FMN H bonds in all known structures. Residues at these positions interact with the phosphate group of FMN and are highly conserved in flavodoxins ([S,T]13; T15; T18), but only

one (S13) is conserved in MioC and in the WrbA family, and two (T13, T18) in YihB. In particular, an aliphatic residue is present at position 15 in MioC and at position 18 in MioC and the WrbA family. Mutagenesis studies on *Klebsiella pneumoniae* flavodoxin (Drummond et al., 1993) showed that, despite high conservation of these functional residues in flavodoxins, substitution of T18 (Fig. 1 numbering) by apolar side chains can preserve significant levels of activity. Aliphatic substitutions at this position are also shown to be tolerated by the natural occurrence of alanine in other known FMN-binding proteins (Grandori & Carey, 1994b; Koonin & Bork, 1994). This evidence indicates that the hydroxyl function at position 18 is dispensable and that alternative H bond donors or acceptors are thus likely. Information of this kind is not available to help in evaluating the possible effects of the substitution T/L15 displayed by MioC.

The loop after $\beta 3$ in flavodoxins provides a conserved hydrophobic residue that interacts with the isoalloxazine ring from the protein side (position 64). This function could be provided by the corresponding histidine or arginine residue in the new sequences. The conserved aromatic residue at position 65 in YihB and the WrbA family is an alternative possibility that cannot easily be discriminated in that context of local sequence conservation.

The loop after $\beta 4$ of flavodoxins hosts the almost invariant aromatic residue that brackets the isoalloxazine ring from the solvent side (position 111). All the new sequences present a tyrosine in that region of the alignment, although in a context of locally high sequence divergence. Assignment to position 111 of this aromatic residue with the common FMN-binding function in the WrbA family and flavodoxins implies, according to our analysis, that this function would be exerted from topologically nonequivalent loops, because Y111 in the WrbA family follows the long insertion that could form an additional β -strand after $\beta 4$. No aromatic or aliphatic residues are conserved in the loop we assign to follow $\beta 4$ in the WrbA family. Thus, the WrbA family seems to display significant divergence from flavodoxins in the third loop that interacts with FMN in known structures, making the prediction of an FMN-binding site for this group of sequences less certain. However, superposition of known 3-dimensional structures of flavodoxins indicates conspicuous differences in the tilt of the isoalloxazine ring plane ($\sim 20^\circ$), suggesting tolerance toward variability at this superficial site (Fukuyama et al., 1992), as also reflected by the different length among flavodoxins of the loop following $\beta 4$ that hosts the bracketing aromatic residue (Fig. 1).

The present analysis also supports prediction of an FMN-binding site in MioC and YihB and suggests that the 2 residues bracketing the isoalloxazine ring are H57 and Y95 in MioC (MioC numbering), and R58 (or Y59) and Y128 in YihB (YihB numbering). S10 of MioC and T9 and T14 of YihB are predicted to provide H bonds with the phosphate group of FMN.

Discussion

The 6 examples analyzed here belong to the so-called "twilight zone" of sequence similarity, where structural homology is not implied with high probability (Chothia & Lesk, 1986; Sander & Schneider, 1991). It is in these cases that we can gain the most insights about protein structure principles from distant evolutionary relationships. Convergent evidence deriving from dif-

ferent methods and the use of groups of related sequences can provide support for structural prediction from analysis of even quite distantly related sequences (Benner, 1992; Doolittle, 1992; Rost et al., 1993).

The data reported here suggest that MioC, YihB, WrbA, Ycp4, P25, and probably RicR share the common core of the 3-dimensional structure of flavodoxins. An interesting implication is that this conservation of secondary and tertiary structure may be compatible with a change in quaternary structure in the WrbA family because experimental data suggest a homodimeric structure for WrbA (Yang et al., 1993) and Ycp4 (Toda et al., 1992). As well, CheY has been observed to self-associate under a wide range of conditions (A. Robertson, pers. comm.). Our analysis indicates 2 regions of major structural divergence from flavodoxins (the long insertions after $\beta 2$ and $\beta 4$) that could be tested for a role in dimerization in the WrbA family.

The proposed structural homology to flavodoxins supports a possible FMN-binding function for the 6 proteins examined here, as has been proposed already on the basis of the conserved N-terminal motif (Koonin & Bork, 1994; Koonin et al., 1994). However, inspection of the putative FMN-binding sites at critical positions of FMN-side-chain interactions provides further support to this hypothesis only for MioC and YihB and points out poor conservation of the third loop of the FMN-binding site in the WrbA family. In this case, the analysis yields only the best candidate contact residues and suggests that experimental characterization of an FMN-binding site in the WrbA family may reveal interesting differences from the flavodoxins. It is relevant to note that the α/β open-sheet fold correlates with a highly conserved position of the active site, at the crevice between the loops following $\beta 1$ and $\beta 3$ (Branden & Tooze, 1991). Our analysis allows identification of these regions in the 6 new sequences independently of assumptions about the nature of the ligand. In the corresponding regions of CheY, a cluster of conserved aspartic acid residues forms the binding site for a divalent metal ion and contains the target site for phosphorylation (Stock et al., 1991). This acidic pocket has no equivalent in the 6 proteins examined here (Fig. 1).

Aside from the N-terminal region, 2 other sequence motifs common to the flavodoxin and WrbA families have been reported by Koonin and Bork (1994) that are in conflict with our results (as already discussed elsewhere; Grandori & Carey, 1994b). The major discrepancy is in the assignment of one sequence region of the WrbA family to either $\beta 3$ or $\beta 4$, which carries implications for the configuration of the active site. We think that their original assignment was weakened by an incorrect alignment among flavodoxins in the region of $\beta 3$ and $\beta 4$, which share a repetitive sequence pattern of residue types. Another difference between the proposed alignments implies a shift in the region corresponding to the fifth β -strand.

The data presented here suggest that the region containing the putative HTH motif of the WrbA family (Grandori & Carey, 1994a) corresponds to the C-terminal helix of flavodoxins, though with some ambiguities about the register of the alignment. This fact, together with the overall sequence similarity to the flavodoxins, suggests that the C-terminal region of the WrbA family is unlikely to be part of an independent domain involved in DNA binding. Although the available experimental evidence concerning the function of WrbA (Yang et al., 1993) might favor the hypothesis of a DNA-binding site, it seems more likely that the HTH signature reflects structural rather than functional

features, a situation reminiscent of the similarity between the HTH motif and part of the globin fold (Subbiah et al., 1993).

Our results may be relevant for studies of brefeldin A resistance. P25 may be implicated in this process in fission yeast (Turi TG, 1993, Genbank entry X73558; Turi et al., 1994) and thus Ycp4 would be a good candidate to play a similar role in budding yeast. Recent isolation of a mutant strain of *Saccharomyces cerevisiae* permeable to brefeldin A (Vogel et al., 1993) will allow use of this organism as an experimental system, and the availability of a structural model could provide new input for experimental design.

On the basis of published studies on the transcriptional control exerted by Pap1 on P25 in *Schizosaccharomyces pombe* (Toda et al., 1992), we searched the region upstream of the Ycp4 open reading frame for possible AP-1-like sites that might suggest a similar regulation of the homologue in budding yeast. We found two AP-1-like sites (TTAGTAA and TTACTAA, starting at -351 and -450 from the beginning of translation, respectively) that match the consensus sequence T(T/G)A(G/C)T (A/C)A derived from Toda et al. (1992) and Schnell et al. (1992). This finding suggests that Ycp4 could be regulated by PAR1, the homologue of PAP1 in *S. cerevisiae* (Schnell et al., 1992). Interestingly, AP-1-dependent regulation has been shown to be sensitive to redox conditions (Abate et al., 1990), and the known targets of PAR1 in *S. cerevisiae* are involved in defense against oxidative stress (Schnell et al., 1992). It would be relevant to further characterize the biochemical properties of the members of the Wrba family and to investigate whether these represent a new class of regulatory proteins that respond to the oxidative state of the cell.

Materials and methods

All protein sequences were derived from the SwissProt database (release 27, entries indicated in Fig. 1), with the exception of the translated sequence from rice cDNA (Genbank: D23815). Flavodoxin structures correspond to Protein Data Bank (PDB) files 1FLV, 1OFV, 2FCR, 1FX1, 4FXN, and respective DSSP files (Kabsch & Sander, 1983). For the NMR structure of flavodoxin from *Megasphaera elsdenii*, which is not deposited in the database, we referred to the secondary structure assignment, according to DSSP, as reported by van Mierlo et al. (1990).

The multiple sequence alignment was generated by the program PILEUP of the GCG package of the Wisconsin Genetics Group (Devereux et al., 1984) and edited by hand on the basis of structural alignment among flavodoxins and alignment of the nonflavodoxin sequences to the profile of the flavodoxin family. The consensus sequence for the CheY superfamily was added to the alignment on the basis of superposition of the 3-dimensional structures of CheY (PDB: 3CHY) and flavodoxin (PDB: 4FXN), and visual inspection of corresponding residues. We referred to the published consensus sequence for the CheY superfamily (Volz, 1993) with the additional indication of the positions that define the conserved hydrophobic core of the protein as reported by Stock et al. (1989b).

PROFILEMAKE and related programs (GCG) were employed for profile analysis (Gribskov et al., 1987, 1992). In the absence of structural information for the most divergent member of the flavodoxin family (flav_cloab), its sequence was aligned to a profile derived from the other flavodoxin sequences and then included in the family. The profiles of the flavodoxin

and Wrba families were derived from alignments of each group of sequences as indicated in Figure 1, with exponential weighting of nonobserved characters, and with the default gap penalty factors for regions with gaps. In deriving these profiles, RicR was not included in the Wrba family, and the sequence of the loop insertion in the middle of the fifth β -strand in the long-chain flavodoxins was not included for the flavodoxin family.

The program PROSAIL, employed for energy analysis of the structural models, was obtained from the Center of Applied Molecular Engineering (CAME), Salzburg, Austria (Sippl, 1993a, 1993b; PROSAIL, PROtein Structure Analysis, version 2.0, instruction manual, 1994, copyright by CAME, Salzburg, Austria).

Note added in proof

After submission of this paper we detected 2 new members of the Wrba family whose sequences have been recently deposited in the database (GenBank: X78224, X78225) corresponding to minor allergens of *Cladosporium herberum* and *Alternaria alternata*.

Acknowledgments

We thank Jeff Stock for his constant and valued contribution to this project, Michael Rozycki for critical reading of the manuscript, Robert Bruccoleri and Jiri Novotny for helpful discussions, Manfred Sippl and Andrej Šali for constructive comments, encouragement, and technical advice in the application of PROSAIL, and Hannes Floeckner for providing coordinates of virtual C β atoms for glycines of 4FXN.

References

- Abate C, Patel L, Rauscher FJ III, Curran T. 1990. Redox regulation of Fos and Jun DNA-binding activity in vitro. *Science* 249:1157-1161.
- Bairoch A. 1992. PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res* 20:2013-2018.
- Benner SA. 1992. Predicting de novo the folded structure of proteins. *Curr Opin Struct Biol* 2:402-412.
- Bewley MC, Boustead CM, Walker JH, Waller DA, Huber R. 1993. Structure of chicken annexin V at 2.25-Å resolution. *Biochemistry* 32:3923-3929.
- Bowie JU, Clarke ND, Pabo CO, Sauer RT. 1990. Identification of protein folds: Matching hydrophobicity patterns of sequence sets with solvent accessibility patterns of known structures. *Proteins Struct Funct Genet* 7:257-264.
- Branden C, Tooze J. 1991. *Introduction to protein structure*. New York: Garland Publishing, Inc.
- Bredt DS, Hwang PM, Glatt CE, Lowenstein C, Reed RR, Snyder SH. 1991. Cloned and expressed nitric oxide synthase structurally resembles cytochrome P-450 reductase. *Nature (Lond)* 351:714-718.
- Burnett RM, Darling GD, Kendall DS, LeQuesne ME, Mayhew SG, Smith WW, Ludwig ML. 1974. The structure of the oxidized form of clostridial flavodoxin at 1.9 Å resolution. *J Biol Chem* 249:4383-4392.
- Chothia C, Lesk A. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823-826.
- Daniels DL, Plunkett G III, Burland VD, Blattner FR. 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* 257:771-778.
- Devereux J, Haeblerli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387.
- de Wind N, Parren P, Stuitje AR, Meijer M. 1987. Evidence for the involvement of the 16 kD gene promoter in initiation of chromosomal replication of *Escherichia coli* strains carrying a B/r-derived replication origin. *Nucleic Acids Res* 15:4901-4914.
- Doolittle RF. 1992. A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci* 1:1563-1577.
- Drummond M, Huff S, Green A. 1993. Roulette mutagenesis of the FMN-

- binding site of *Klebsiella pneumoniae* flavodoxin. *Eur J Biochem* 217:395–400.
- Fukuyama K, Matsubara H, Rogers LJ. 1992. Crystal structure of oxidized flavodoxin from a red alga *Chondrus crispus* refined at 1.8 Å resolution. *J Mol Biol* 225:775–789.
- Garnier J, Osguthorpe DJ, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120:97–120.
- Grandori R, Carey J. 1994a. Two highly homologous putative DNA-binding proteins in yeast and *E. coli*. *Trends Biochem Sci* 19:72; 19:107.
- Grandori R, Carey J. 1994b. FMN or DNA binding? [reply]. *Trends Biochem Sci* 19:235.
- Gribskov M, Lüthy R, Eisenberg D. 1990. Profile analysis. *Methods Enzymol* 183:146–159.
- Gribskov M, McLachlan AD, Eisenberg D. 1987. Profile analysis: Detection of distantly related proteins. *Proc Natl Acad Sci USA* 84:4355–4358.
- Jouanneau Y, Richaud P, Grabau C. 1990. The nucleotide sequence of a flavodoxin-like gene which precedes two ferredoxin genes in *Rhodobacter capsulatus*. *Nucleic Acids Res* 18:5284.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Koonin EV, Bork P. 1994. FMN or DNA binding? *Trends Biochem Sci* 19:234–235.
- Koonin EV, Bork P, Sander C. 1994. Yeast chromosome III: New gene functions. *EMBO J* 13:493–503.
- Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132.
- Leslie AGW. 1990. Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 Å resolution. *J Mol Biol* 213:167–186.
- Løbner-Olesen A, Boye E. 1992. Different effects of MioC transcription on initiation of chromosomal and minichromosomal replication in *Escherichia coli*. *Nucleic Acids Res* 20:3029–3036.
- Mayhew SG, Tollin G. 1992. General properties of flavodoxins. In: Müller F, ed. *Chemistry and biochemistry of flavoenzymes, vol III*. Boca Raton, Florida: CRC Press. pp 427–466.
- Morikawa K, Matsumoto O, Tsujimoto M, Katayanagi K, Ariyoshi M, Doi T, Ikehara T, Ohtsuka E. 1992. X-ray structure of T4 endonuclease V: An excision repair enzyme specific for a pyrimidine dimer. *Science* 256:523–526.
- Ostrowski J, Barber MJ, Rueger DC, Mileer BE, Siegel LM, Kredich NM. 1989. Characterization of the flavoprotein moieties of NADPH-sulfite reductase from *Salmonella typhimurium* and *Escherichia coli*. *J Biol Chem* 264:15796–15808.
- Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85:2444–2448.
- Porter TD, Beck TW, Kasper CB. 1990. NADPH-cytochrome P-450 oxidoreductase gene organization correlates with structural domains of the protein. *Biochemistry* 29:9814–9818.
- Rao ST, Shaffie F, Yu C, Satyshur KA, Stockman BJ, Markley JL, Sundaralingam M. 1992. Structure of the oxidized long-chain flavodoxin from *Anabaena* 7120 at 2 Å resolution. *Protein Sci* 1:1413–1427.
- Rost B, Schneider R, Sander C. 1993. Progress in protein structure prediction? *Trends Biochem Sci* 18:120–123.
- Sander C, Schneider R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins Struct Funct Genet* 9:56–68.
- Schnell N, Krems B, Entian KD. 1992. The *PAR1 (YAP1/SNQ3)* gene of *Saccharomyces cerevisiae*, a *c-jun* homologue, is involved in oxygen metabolism. *Curr Genet* 21:269–273.
- Sippl MJ. 1993a. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct Funct Genet* 17:355–362.
- Sippl MJ. 1993b. Boltzmann's principle, knowledge based mean fields and protein folding. *J Comput Aided Mol Design* 7:473–501.
- Sippl MJ, Weitckus S. 1992. Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins Struct Funct Genet* 13:258–271.
- Smith WW, Patridge KA, Ludwig ML, Petsko GA, Tsernoglou D, Tanaka M, Yasunobu KT. 1983. Structure of oxidized flavodoxin from *Anacystis nidulans*. *J Mol Biol* 165:737–755.
- Stock AM, Mottonen JM, Stock JB, Schutt CE. 1989a. Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature (Lond)* 337:745–749.
- Stock JB, Lukatt GS, Stock AM. 1991. Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. *Annu Rev Biophys Biophys Chem* 20:109–136.
- Stock JB, Ninfa AJ, Stock AM. 1989b. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* 53:450–490.
- Subbiah S, Laurents DV, Levitt M. 1993. Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Curr Biol* 3:141–148.
- Tanaka I, Appelt K, Dijk J, White SW, Wilson KS. 1984. 3-Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature* 310:376–381.
- Toda T, Shimanuki M, Saka Y, Yamano H, Adachi Y, Shirakawa M, Kyogoku Y, Yanagida M. 1992. Fission yeast pap1-dependent transcription is negatively regulated by an essential nuclear protein, crm1. *Mol Cell Biol* 12:5474–5484.
- Turi TG, Webster P, Rose JK. 1994. Brefeldin A sensitivity in *Schizosaccharomyces pombe*. *J Biol Chem* 269:24229–24236.
- van Mierlo CPM, Lijnzaad P, Vervoort J, Müller F, Berendsen HJC, de Vlieg J. 1990. Tertiary structure of two-electron reduced *Megasphaera elsdenii* flavodoxin and some implications, as determined by two-dimensional ¹H-NMR and restrained molecular dynamics. *Eur J Biochem* 194:185–198.
- Vogel JP, Lee JN, Kirsch DR, Rose MD, Sztul ES. 1993. Brefeldin A causes a defect in secretion in *Saccharomyces cerevisiae*. *J Biol Chem* 268:3040–3043.
- Volz K. 1993. Structural conservation in the CheY superfamily. *Biochemistry* 32:11741–11753.
- Volz K, Matsumura P. 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution. *J Biol Chem* 266:15511–15519.
- Yang W, Liuying N, Somerville RL. 1993. A stationary-phase protein of *Escherichia coli* that affects the mode of association between the *trp* repressor protein and operator-bearing DNA. *Proc Natl Acad Sci USA* 90:5796–5800.
- Watt W, Tulinsky A, Swenson RP, Watenpaugh KD. 1991. Comparison of the crystal structures of a flavodoxin in its three oxidation states at cryogenic temperatures. *J Mol Biol* 218:195–208.