A structural model for the nucleotide binding domains of the flavocytochrome *b*_{–245} β -chain

WILLIAM R. TAYLOR,¹ DAVID T. JONES,^{1,2} AND ANTHONY W. SEGAL³

¹ Laboratory of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

² Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, United Kingdom

³ Department of Medicine, University College London, 5 University Street, London WC1E 6JJ, United Kingdom

(RECEIVED December 8, 1992; ACCEPTED June 30, 1993)

Abstract

NADPH is a system in phagocytic cells that generates O_2^- and hydrogen peroxide in the endocytic vacuole, both of which are important for killing of the engulfed microbe. Dysfunction of this oxidase results in the syndrome of chronic granulomatous disease, characterized by a profound predisposition to bacterial and fungal infections. A flavocytochrome *b* is the site of most of the mutations causing this syndrome. The FAD and NADPH binding sites have been located on the β subunit of this molecule, the C-terminal half of which showed weak sequence similarity to other reductases, including the ferredoxin-NADP reductase (FNR) of known structure. This enabled us to build a model of the nucleotide binding domains of the flavocytochrome using this structure as a template. The model was built initially using a novel automatic modeling method based on distance–matrix projection and then refined using energy minimization with appropriate side-chain torsional constraints. The resulting model rationalized much of the observed sequence conservation and identified a large insertion as a potential regulatory domain. It confirms the inclusion of the neutrophil flavocytochrome *b*_{–245} (Cb-245) as a member of the FNR family of reductases and strongly supports its function as the proximal electron transporting component of the NADPH oxidase.

Keywords: alignment; distance geometry; flavocytochrome; modeling; NADPH oxidase; phagocyte; prediction; protein structure

NADPH oxidase is a system in phagocytic cells that generates O_2^- and hydrogen peroxide in the endocytic vacuole, both of which are important for killing of the engulfed microbe (Segal, 1989). Dysfunction of this oxidase results in the syndrome of chronic granulomatous disease (CGD), characterized by a profound predisposition to bacterial and fungal infection. The identity of the molecules responsible for electron transport in this oxidase has been a puzzle for several decades. The first redox molecule to be identified was a low potential cytochrome *b* (Segal & Jones, 1978) called cytochrome *b*_{–245} because of its midpoint potential (or *b*₅₅₈, the wavelength of its α band of light absorption), which is situated in the plasma membrane and in the membranes of the specific cytoplasmic granules. An unusual feature of this cytochrome *b*

was that it was composed of proteins with molecular weights of 21 kDa and 76–92 kDa, referred to as the α and β subunits, respectively, (Parkos et al., 1987; Segal, 1987). The α subunit demonstrates very limited homology with the heme binding region of cytochrome *c* oxidase (Parkos et al., 1988), and radiation inactivation analysis revealed a vulnerable target associated with the heme of 21 kDa, almost exactly the size of the α subunit (Nugent et al., 1989). However, spectral analysis was indicative of a low-spin six-coordinate heme with axial imidazole or imidazolate ligands (Hurst et al., 1991). Because the α subunit only contains one invariant histidine, this would indicate an unusual ligand, or the association of the heme between two subunits, either between two α 's, or between an α and a β subunit.

The involvement of an FAD-containing flavoprotein in this system was suggested by the enhancement of activity of the oxidase in broken cell preparations by FAD (Babior & Kipnes, 1977), its inhibition by the analogue

Reprint requests to: William R. Taylor, Laboratory of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom.

5-deaza-FAD (Light et al., 1981), and the observation of an EPR signal resembling the FAD semiquinone in membranes from activated but not resting neutrophils (Kakinuma et al., 1986). Considerable difficulty was experienced in identifying the flavoprotein because the FAD is dissociated upon detergent solubilization of the membranes and because the system is inactive unless stimulated, an activation that is dependent upon the association of at least three cytoplasmic proteins, including the small GTP-binding protein p21rac, with the cytochrome *b* in the membrane (Abo et al., 1992). It transpired that the electron transport chain was in fact a flavocytochrome *b*, with the β subunit the site of FAD and NADPH binding (Segal et al., 1992). Membranes from neutrophils of patients with the X-linked form of CGD, in whom the lesion is on the gene coding for the β subunit that is missing in their cells, contain about 20% of the normal complement of FAD.

Additional support was given to the identification of the β subunit as a flavoprotein by the finding of amino acid sequence homology in the conserved nucleotide binding domains with several members of the family of ferredoxin-NADP reductases. A solution of the known crystal structure of a member of this family (Karplus et al., 1991), the ferredoxin reductase, FNR, has provided a template upon which to model the structure of the neutrophil protein. Unlike the other reductases in this family, the β chain of the flavocytochrome *b* is membrane bound, a property that is almost certainly conferred by the amino-terminal half of the molecule that is predominantly hydrophobic with a number of possible trans-membrane helices. This region of the protein is clearly important for anchoring the molecule in the membrane and for interacting with the α subunits. In this paper we concentrate upon the C-terminal (globular) half of the protein and construct the model using the ferredoxin-NADP reductase as a template.

Results and discussion

Sequence alignment

Several NAD(P)H reductases have been identified as being similar to the neutrophil flavocytochrome *b*₂₄₅ (Cb-245) sequence based on a few short sequence motifs (Segal

et al., 1992). Because the basis of this selection was neither rigorous nor complete, a more thorough scan was made on the current Leeds combined-sequence data bank (Bleasby & Wootton, 1990). Of principal interest was the relationship of Cb-245 and FNR—only sequences that contained motifs common to both these sequences were considered for further analysis.

In addition to the three motifs identified by Segal et al. (1992: A, B, and C in their Fig. 3), a further motif is conserved some 20 residues before motif B in both the Cb-245 and FNR sequences. This motif is not found in the cytochrome P-450 reductases or the sulfite reductases, and although both of these proteins are clearly members of the superfamily, they were excluded from further consideration. Of passing interest are the toluene reductases (Shaw & Harayama, 1992), which lack the N-terminal domain, beginning between motifs 2 and 3 (Fig. 1). The hierarchical condensation of the 32 sequences into families revealed the relationships shown in Figure 2.

The clustering produced two families, one based on the ferredoxin reductases, and the other containing the nitrate and cytochrome *b*₅ reductases. The FNR family contains two subgroups; the larger contains the FNR sequence itself, and the other comprises various distantly related proteins involved mainly in bacterial hydrocarbon metabolism. The nitrate reductase family, similarly, has a subgroup of distantly related proteins, including the Cb-245 sequence. However, given the uncertainties involved in such remote alignments, it is difficult to place much faith in the details of the final clustering order. It can only be said that, with the exception of a closer similarity to the ferric reductase FRE1 gene product of *Saccharomyces cerevisiae* (YSCFRE1A), the Cb-245 sequence is effectively equidistant from all those considered. To remove some of the bias associated with closely related entries among the selected sequences, the number considered was halved by eliminating all but one representative from the subfamilies (see Fig. 2 for details). The alignment and clustering of this reduced set produced much the same result but formed a more independent subgroup of the bacterial metabolic proteins (Fig. 2). The alignments of both the full and reduced set of proteins correctly aligned all motifs specified in Figure 1, thus providing independent confirmation of the initial selection of these motifs as characteristic. In addition, a further motif nearer the

<---- 1 ----> <-2-> <----- 3 -----> 4

[RH]x[YF][ST] x{30,100} GPxG x{10,40} [GAST]GxG[AVILFM] x{50,150} CG

Fig. 1. Regular expression for the ferredoxin reductase type domain. Motifs are identified by numbers. These differ slightly from the previous description: motif 2 has been added and motifs 1 and 3 have been made less specific. Amino acids are specified by their one-letter codes with x used to match any acid. Acids grouped by brackets are alternatives at one position, whereas pairs of numbers within braces are low and high multiplicative factors for the preceding element. The latter are used to specify variable gaps between motifs; thus, x{10,40} allows from 10 to 40 residues to be found between motifs 2 and 3. In FNR, the motif locations are as follows: 1: 93; 2: 149; 3: 169; 4: 272.

```

.....p----- BENC_ACICA
...p-----+b-----*XYLZ_PSEPU
..p-b-p-----+-----*MEMC_METCA
..|...b-----+-----p-*PSENAPDOXA
..|.....b-----b- PSENAPDOXB
..|.....p----- FENR_PEA
p-b-----p-+-----b-p-*CIPFNRA
|.....|.....b- N_1FNR
|.....|.....b- N_2FNR
|.....|.....b- FENR_SPIOL
|.....|...p----- FENR_SPISP
|.....b-p+b-----*SYONDFH
|.....b+-----*FENR_ANASO
|.....p-----+-----*C24B_HUMAN
b-p--b-----+-----*YSCFRE1A
..|...p-----+-----*DMPP_PSEPU
..b-p-b-----+-----*S15303
.....|.....p- A40495
.....|.....p-b- NC5R_RAT
.....|...p+---b-p-*B26616
.....|.....|.....b- JS0468
.....|.....|.....b- NC5R_HUMAN
.....|.....|.....b- BOVCYB5R
.....|.....|.....b- NC5R_BOVIN
.....|...p-b+p----- S16895
.....|.....|.....p- A35499
.....|.....|...b-p---b-*MZENAR
.....|.....|.....b-p--- NIA1_ORYSA
.....|.....|.....b--- S17453
.....|.....|.....p- NIA1_TOBAC
.....|.....|...p-b- NIA2_TOBAC
.....|.....|...p---b--- NIA_LYCES
.....b-p-b-p+b-----*NIA_SPIOL
.....|.....|...p----- NIA_CUCMA
.....|.....|...b+p-b-----*S15959
.....|.....|...b-p----- NIA1_ARATH
.....|.....|...b----- NIA2_ARATH
.....b---p+-----*NIA_ASPNI
.....b+-----*NIA_NEUCR

```

Fig. 2. Hierarchic condensation of the sequences extracted by the pattern in Figure 1. The sequences are identified by their data bank code. Each column (two characters) to the left of this represents an iterative cycle of the program MULTAL (Taylor, 1988). Vertical connections indicate that pairs of sequences (or families) were aligned during that cycle of the program. A representative sample of sequences was selected by retaining only one sequence (starred) from each subfamily that had formed before the fifth alignment phase (marked by a plus on the tree). The selected sequences are as follows:

C24B_HUMAN = Cytochrome *b*₂₄₅ heavy chain (P22 phagocyte B-cytochrome neutrophil cytochrome *b*, 91-kDa polypeptide) – human (Royer-Pokora et al., 1986)
YSCFRE1A = Essential for ferric reductase activity – *Saccharomyces cerevisiae* (Dancis et al., 1992)
NIA_ASPNI = Nitrate reductase (EC 1.6.6.1) – *Aspergillus nidulans* (Johnstone et al., 1990)
NIA_NEUCR = Nitrate reductase (NADPH) (EC 1.6.6.3) – *Neurospora crassa* (Okamoto et al., 1991)
B26616 = Cytochrome *b*₅ reductase (Yubisui et al., 1987)
NIA_SPIOL = Nitrate reductase (EC 1.6.6.1) (NR) – *Spinacia oleracea* (spinach) (Prosser & Lazarus, 1990)
S15959 = Nitrate reductase (NAD(P)H) (EC 1.6.6.2) – European white birch (Friemann et al., 1991)
MZENAR = NADH:nitrate reductase (AA at 1) (EC 1.6.6.1) – *Zea mays* (Gowri & Campbell, 1989)
DMPP_PSEPU = Hydroxylase P5 protein (EC 1.14.13.7) phenol 2-mono-oxygenase P5 component – *Pseudomonas* (Nordlund et al., 1990)
S15303 = Hypothetical protein 7.6 – *Salmonella typhimurium* (Jiang et al., 1991)
MEMC_METCA = Methane monooxygenase component C (EC 1.14.13.25) methane hydroxylase – *Methylococcus capsula* (Stainthorpe et al., 1990)
XYLZ_PSEPU = E 1,2-dioxygenase electron transfer component contains ferredoxin and ferredoxin-NAD(+) (Harayama et al., 1991)
PSENAPDOXA = DOXA locus PSENAPDOXA – *Pseudomonas putida* (M.J. Simon et al., unpubl. ms.)
FENR_ANASO = Ferredoxin-NADP reductase (EC 1.18.1.2) (FNR) – *Anabaena* sp. (Fillat et al., 1990)
SYONDFH = F N-terminal domain homologous to CpcD proteins of *Synecchococcus* sp. (Schluchter & Bryant, 1992)
N_1FNR = Ferredoxin:NADP⁺ oxidoreductase (ferredoxin reductase) (EC 1.18.1.2) – *Spinacia oleracea* (spinach) (Karplus et al., 1991)

amino-terminus was revealed, comprising G[QK][HFYS] (close to strand S1-5 in Fig. 3).

Alignment refinement

The alignment was refined in the light of specific structural constraints imposed by the known tertiary structure for one member of the family, FNR. It was felt that the most important aspects to be maintained were the hydrophobicity of buried positions and the avoidance of alignment gaps in the protein core. In the regions around the motifs described above, these constraints were easily satisfied. The resulting alignments were essentially similar to those obtained by Andrews et al. (1992), in which the HMP protein (a protein with a hemoglobin-like function) of *Escherichia coli* was aligned with other members of the FNR family of reductases, not including Cb-245. The more difficult regions were the two termini.

Amino-terminus alignment

The amino-terminus had the additional complication that the start of the reductase domain was often unclear. The problem was tackled by working back from the first motif and taking the first available alignment that fitted the structurally required hydrophobic residues. On the amino side of the first motif, a conserved pattern (*---*---*) of alternating hydrophobicity (*) was found with the two central positions corresponding to deeply buried residues in the β -strand of the known structure. A similar alternating pattern of hydrophobicity (*---*---*) was found beyond this, corresponding to the first β -strand, again with the central conserved positions aligned with the most buried residues. Between these there was a distinct preference for a positive charge (+ above). These two alignments were supported by examining the secondary structure predicted by the GOR method (Garnier et al., 1978) for each sequence.

[illegible]

Carboxy-terminus alignment

The C-terminus of the domain corresponded to the terminus of the protein, allowing less freedom in the alignment of the terminal β -strand. The most C-terminal region of hydrophobicity was taken as the final buried β -strand of the alternating β/α domain. This assignment was generally unambiguous, giving rise to a conserved terminal aromatic position (F or Y). For the Cb-245 sequence, however, two alignments seemed plausible against the FNR sequence. The first, with Cb-245 above and FNR below,

```
GVHFIF
WNVEVY'
```

conserves the terminal aromatic character and the bulk of the hydrophobic residues, whereas the second,

```
FNKENF
WNVEVY'
```

not only conserves the terminal aromatic residue but also two further positions (N and E). The latter was adopted, despite its hydrophilic nature, on the basis of this similarity and its additional advantage of reducing the size of the preceding gap.

The large insert

The only remaining region of uncertainty was in the placement of a sizeable insertion in both the Cb-245 reductase and its nearest relative FRE1 (YSCFRE1A). This large insert was confined to the edge of the alternating β/α domain, but it was not clear whether it should lie before or after the edge helix (in that domain). The former location was chosen on the basis of the alignment of the FRE1 sequence with its nearest partner in the alignment in Figure 3, based on the following similarity:

```
RPDLKELVRSEIK
RID*EELIRQHAG'
```

Although not strong, this assignment was supported by some further similarity in the inserted sequence in both

the Cb-245 and FRE1 (YSCFRE1A) sequences, consisting of a highly charged segment flanked by three residues of a more hydrophobic nature:

```
FAVHHDEEKDVIT
KGFDDKEESEITV'
```

Our choice was further supported by the general variation found in this (topological) location in other nucleotide binding proteins.

The GOR secondary structure prediction indicates an α -helical conformation for the large insertion (for both Cb-245 and FRE1). Despite the lack of hydrophobic positions expected for an amphipathic α -helix, such a structure might exist if some of the charges were compensated in a buried salt bridge. Further possibilities for this interesting region will be considered below.

Model construction

Application of the protocol described in the Methods produced a molecular model, the structure of which was derived automatically from the final sequence alignment. The model protein maintained all the essential features (motifs, secondary structure, hydrophobic core) of the known X-ray structure, which were implicit in the final alignment (Fig. 4; Kinemage 1). With the exception of the large insertion discussed above, most relative insertions and deletions were minor and occurred in well-exposed positions (as a consequence of the alignment strategy). Whereas this model is derivative, and its independent features cannot be relied upon absolutely, the major features of the overall structure closely mimicked those of the crystal structure. This indicated that the C-terminal half of Cb-245 could be included with confidence as a member of the FNR family of reductases, and as such is almost certainly the proximal electron-transporting molecule in the NADPH oxidase complex.

Features that were not dictated by the alignment were the details of side-chain packing and the conformation of inserted loops. At the distant level of sequence similarity on which the model was based, it cannot be expected that such details will have anything other than a qualitative correspondence to the real structure. These will, there-

Fig. 3 (*facing page*). The aligned sequences after refinement of gap location. Sixteen sequences are aligned, which fall into two distantly related blocks. The top group includes the flavocytochrome *b*₋₂₄₅ (C24B_HUMAN), whereas the lower contains the ferredoxin reductase of known structure (N_1FNR). Below the latter, relevant structural information is summarized. Buried: an asterisk indicates any residue with less than 10 Å² solvent-accessible surface area. Nucleotide: indicates residues contacting the nucleotide groups as follows: A, P, and F contact the adenine, phosphate, and flavin components of the FAD, respectively; N is a possible contact with the NADPH moiety. Secondary structure: the definitions extracted from the DSSP program of Kabsch and Sander (1983). Like letters are hydrogen-bonding partners in β -structure; @ characters indicate α -helix. In addition to allowing easy comparison with the alignment of Andrews et al. (1992), the original FNR secondary structure identifiers (Karplus et al., 1991) and a simplified representation of the secondary structure is also given, where strands are represented by β characters and α -helix by α characters.

Fig. 4. The overall fold of the final model is indicated by a bold line connecting α -carbon positions. The bound nucleotides are indicated by finer lines, as is the movement of the large insert required to allow NADPH binding.

fore, not be described in any detail except where they appear to have implications for nucleotide binding and function.

Remodeling the large insert

The most remarkable difference in the Cb-245 model and the FNR structure was the addition of the large insertion on the edge of the alternating β/α domain (remote from the C-terminus; Kinemages 1, 2). On the basis of its predicted secondary structure, this segment was restrained toward an α -helical conformation during the modeling process, resulting in a rather large regular helix (20 residues) lying directly over the nucleotide binding cleft. This location arose because the point of insertion of the structure is in a loop facing the binding cleft, thus constraining the helix to both begin and end in this region. From an acquired knowledge of the structure of other β/α nucleotide binding domains, such a prominent structure appeared very unusual and would clearly have dramatic implications for the binding of both nucleotides.

Despite its solid prediction as an α -helix, it was considered probable that the large insert might contain a bend. The most likely location for this would be the highly charged sequence: HHDEEKD (residues 493–499¹). However, any model based on longitudinally packed secondary structures (parallel or antiparallel) places the charged sequence remote from the binding site, where, by

comparison with the other loops on this face, considerable sequence variation would be expected, including insertions and deletions. By contrast, when compared to the FRE1 sequence (the other protein with an equivalent large insert), the charged segment contains the only matching positions. In an attempt to reconcile these conflicting observations, the model was modified “by hand,” producing a structure that preserves the bulk of the insert as helical but with some unwinding of the N-terminus, allowing it to be moved partially clear of the binding cleft. In this position, the conserved residues face the cleft in the region of the NADP binding location, whereas the N-terminal end of the insert is extended and could easily hydrogen bond to the edge of the β -sheet. After 100 cycles of minimization, no bad contacts remained and some of the negative charges in the insert formed salt bridges across the cleft.

A particularly interesting aspect of the oxidase is that it is maintained in an inactive state until attachment of opsonized particles to receptors on the plasma membrane triggers ingestion and activation of electron transport. This process is associated with phosphorylation of the flavocytochrome (Garcia & Segal, 1988) and the binding of two cytoplasmic proteins, p47phox and p67phox, that are required for activation (Clark et al., 1990). We postulate that the additional sequence (413–503) might prevent access of the NADPH to this cleft in the inactive state, and that activation, with access of the NADPH to the FAD, could be induced by its displacement, possibly by direct binding of one (or both) of the cytosolic factors following phosphorylation of the cytochrome. We would anticipate that the charged sequence (HHDEEKD) is implicated directly in mediating these events, perhaps even containing the site of phosphorylation. Given its predicted (and modeled) α -helical structure, some charges could face the binding cleft, whereas others participate in a recognition site for the control factors. It was suggested by the model that some of the negative charges could occupy the position of the NADPH phosphates in the inactive state, thus directing the positive residues toward a regulatory site.

It is of particular interest that a patient with X-linked CGD has been described in whom aspartic acid at residue 499 (second Asp in the charged segment) is mutated to glycine (Roos et al., 1993). The D $\rightarrow$ G mutation leaves both subunits of the flavocytochrome with intact redox centers, which is consistent with its location outside the core of the protein, in a position where it could be critical to function but not structure. Demonstration of the effect of the mutation on the binding of the cytosolic factors upon activation and on access of NADPH to its binding site will provide valuable information regarding the function of this potentially regulatory region of the molecule. The iron uptake systems in yeast are regulated in response to the availability of iron. This occurs at the level of transcription of proteins involved in cellular iron metabolism (Klausner et al., 1993). Whether, in addition,

¹ All residue numberings refer to Leeds sequence data bank entry CB24_HUMAN.

electron transport by the ferric reductase (FRE1) can be activated in a similar manner to the NADPH oxidase remains to be determined.

Nucleotide interactions

Although the conformation of the side chains of surface residues in the model cannot be relied upon, those in the core will be more constrained, and those in the binding sites should be most conservative. This latter expectation will be realized only if the mode of nucleotide binding remains unchanged because then the interacting side chains have a fixed set of potential interactions in contrast to mutually mobile side-chain-side-chain interactions. It is probable that the FAD moiety has remained largely unchanged in its location, and its binding site will consequently be examined in detail. By contrast, the NADP binding site in the region remote from the FAD is likely to be substantially altered as a consequence of the large insert in that region (Fig. 5; Kinemage 2).

FAD binding site

In the FNR structure, the principal interaction of the flavin is with two tyrosines (95 and 315), which pack flat on either side of the isoalloxazine rings (Karplus et al., 1991: their Fig. 3). Both residues correspond to phenylalanines in the model (residues 339 and 570), consistent with the lack of interaction of the tyrosine hydroxyl groups with the flavin. As mentioned above, the C-terminal residue is strongly conserved as Phe or Tyr (with one exception), whereas the equivalent residue to Tyr 96 (FNR) is Phe 339, which is conserved in motif 1 (Fig. 1).

All the residues in motif 1 (HPFT) interact with the flavin: the histidine interacts extensively with the sugar component and also the nearest phosphate, both the Pro and Phe pack against the isoalloxazine group, while the threonine forms a hydrogen bond to the O4 atom of the central ring. The latter is a more specific interaction than that

formed by the equivalent Ser 96 in the FNR structure, which has a main-chain but no side-chain hydrogen bond to the nucleotide and may simply have formed because the model lacks the buried water molecules found in the crystal structure. A further bond to the isoalloxazine N5 atom is made by Tyr 46 in the model, which has probably only formed because the added bulk of the tyrosine (previously Ser) places the hydroxyl group further into the binding site. This is consistent with the mainly hydrogen-bond donors found at this position (Y, S, H).

The internal side of the flavin binding pocket is formed by Ile 410 in motif 3 (GIG). This position is more typically threonine, and together with the Tyr-Phe changes described above, results in a more hydrophobic ring environment than found in the FNR structure. The phosphates of the FAD interact with the amide groups at the terminus of the small helix in the β -sheet domain. This interaction is facilitated by Gly 369, which is totally conserved except in the FRE1 sequence, where it is a leucine. In this sequence, a shift of six positions is required to bring the nearest Gly into alignment (at little other cost), and if the FRE1 sequence were to be modeled, such a shift might be considered.

The adenine group forms few interactions in the FNR structure, packing only against a tyrosine, which is equivalent to Asp 359 in the model. This unusual substitution might be explained by the proximity of the carboxyl group of Asp to the adenine sugar with which hydrogen bonding seems probable.

NADP binding site

Little can be said with any certainty about NADP binding, not only because of the changes in the model, but also because there is no NADP bound in the crystal structure of FNR. However, an adenine-derived inhibitor is bound in what is taken to be the binding site of the adenine component of the NADP (Karplus et al., 1991), and because the nicotinamide component of the dinucleotide must in-

Fig. 5. The final model with both NADPH and FAD bound. The residues indicated constitute the main sequence motifs and interact with the nucleotides deep in the binding cleft at the active site where reduction occurs. Their numbering is that of the full β subunit as deposited in the sequence data bank (C24B_HUMAN).

teract with the flavin, the orientation of the cofactor is roughly determined.

Dihydrofolate reductase provides a model for NADP binding (Filman et al., 1982), and the coordinates of the NADP were taken into the Cb-245 model with the adenine ring in the location suggested by the inhibitor and the nicotinamide group lying beside the flavin. Following Karplus et al. (1991: their Fig. 4), the nicotinamide ring was stacked against the flavin ring such that the atoms involved in hydride transfer (flavin N4, nicotinamide C4) were overlaid at a distance of 3.3 Å when viewed normal to the ring planes.

With the NADP in this conformation, it was immediately apparent that the location of the large insert left insufficient space for the adenine group. The helical portion of the insert was accordingly swung back (by rotation of 30° about a line joining the C α atoms of residues 489–508) to create some space, breaking the salt bridges that had formed across the cleft. The complex was then subjected to 150 cycles of minimization.

In the refined model, the adenine of the NADP was buried below the helical portion of the large insert. The 2'-phosphate interacted closely with Lys 21, but as this is not a conserved position, better candidates that are within range are Arg 445 and Lys 507, both of which are commonly positive in character in the alignment. The 5'-phosphate, by contrast, has a clear bonding partner in Arg 355, which is equivalent to that postulated by Karplus et al. (1991) for the NADP bound in FNR. Arginine 355 lies between the adenine and the nicotinamide halves and interacts equally with both phosphates. As would be expected for such a key interaction, the position is conserved either as Arg or Lys in the sequence alignment.

Few specific interactions were made with the nicotinamide sugar, although Lys 372 was well placed to bond both with the sugar and phosphate groups. The nicotinamide group itself maintained its initial position stacked against the flavin rings, and the only change in this region of the binding site was a small outward displacement of the C-terminal Phe to avoid a steric clash with the nicotinamide group.

Orientation of the molecule to the membrane

The flavocytochrome *b* of the NADPH oxidase lies in the plasma membrane between its substrate in the cytosol and the phagocytic vacuole into which it secretes superoxide.

The model we have constructed of the globular portion of the β subunit of the cytochrome must be largely exposed to the solvent and accessible to NADPH from the cytoplasm, yet still be linked to its membrane-bound N-terminal half. In addition, there is also the need to place the isoalloxazine group of the FAD moiety within a reasonable distance (under 10 Å) of the heme held in the membrane by the α subunits to allow electron transfer. We have not attempted to specifically model the N-terminal

region of the β subunit. This region is hydrophobic and appears to comprise between four and six potential transmembrane helices. Its structure will, undoubtedly, be dependent upon interactions with the transmembrane helical regions of the α subunits, and the resulting complex will be important for orientating the hemes to interact with the FAD upon activation. The intact complex is sufficiently large with perhaps 10 helices (say, 2 each from the α and 4–6 from the β subunit), to form a channel through which the superoxide is directed into the exterior.

To gain an impression of the orientation of the globular portion of the β -chain relative to the membrane complex, we extracted the L and M subunits of the photosynthetic reaction center (Deisenhofer et al., 1985) (10 transmembrane helices) also retaining the two central chlorophyll molecules to represent the two hemes held by the α subunits of the cytochrome. The N-terminus of the globular domain was connected to the C-terminus of one of the membrane subunits, imitating the intact β subunit of the flavocytochrome. Maintaining this link made it possible to bring the isoalloxazine group of the flavin to within 8 Å of both hemes without incurring any serious steric hindrance. Together, these considerations constrained the C-terminal (globular) half of the β subunit to lie with the C-terminal edge of the sheet in the β/α domain packing against the membrane component as shown in Figure 6. The C-terminus was deeply buried alongside the flavin, as was the GPFG sequence comprising motif 2 (Fig. 1). This location may offer some explanation for the high conservation of this motif, though the GP part of this motif is undoubtedly conserved as a consequence of the *cis*-proline conformation observed in the FNR structure.

There have been two reports relating to the structure of the β subunit of Cb-245. The first indicated that the C-terminus was exposed on the cytosolic face of the plasma membrane (Rotrosen et al., 1990). This was on the basis that synthetic peptides to this region of the molecule inhibited activation of the oxidase in the cell-free assay, a system in which sodium dodecyl sulfate or arachidonic acid can activate electron transport by a mixture of membranes and cytosol from resting cells. These peptides inhibited translocation of the cytosolic proteins to the membranes upon activation in this system, and antibodies raised to them appeared to react with the inner surface of the plasma membrane. However, in the orientation of our model (Fig. 6), the C-terminus must be inaccessible from the cytosol. It is possible, however, in the artificial conditions used by Rotrosen et al. (1990), that some fraction of the subunits dissociate, possibly as a transient hinge movement of the globular portion of the β subunit away from the membrane-bound component, thus exposing the terminus to antibody binding.

The second investigation (Imajoh-Ohmi et al., 1992) of the structure of this molecule used immunostaining of intact and sonicated neutrophils and cytoplasts (bags of cytosol contained in membranes that are prepared by the

Fig. 6. Schematic model of the complete membrane-bound complex. The orientation of the globular domain of the β subunit is determined by remaining linked to its membrane-bound N-terminal half while allowing the flavin to approach the two hemes, which are probably held in the membrane by the two α subunits. The membrane component is based on a part of the bacterial photosynthetic reaction center, and the two hemes occupy the location of the central symmetric pair of chlorophylls. It must be stressed that this diagram is not intended as an accurate depiction of the complete B-245 complex but is merely intended to show that the model is at least compatible with the formation of such a complex.

extrusion of the cytoplasmic granules and nuclei by centrifugation on density gradients). The antibodies to the β subunit were raised to the C-terminus, which did not appear to be exposed to the exterior, and to residues 150–172 in the hydrophobic half, which did. They also conducted proteolytic cleavage on intact cells using papain and found that the β subunit was cleaved to yield an 18-kDa C-terminal fragment, which they interpreted as indicating that the region around residues 369–398 is extracellular. This region lies between the NADPH and FAD binding sites, the portion of the clearest and most conserved sequence similarity throughout the whole superfamily. The only reconciliation of these latter data with our model is that the protein has been denatured completely, allowing the passage of a hydrophobic segment (normally in the protein core) across the membrane where it has been bound by antibodies.

Methods

Database search

From the original alignment it was clear that four motifs could be identified. These were combined in a joint pat-

tern of the regular expression type (Fig. 1), which was used to scan the current release (version 18) of the combined sequence data banks (Bleasby & Wootton, 1990).

Multiple sequence alignment

We used the UWGCG program FINDPATTERNS (Devereux et al., 1984), modified so that the region spanning the combined motifs was extracted along with a 100-residue flanking sequence at either end (if present), in a format suitable for multiple sequence alignment by the program MULTAL (Taylor, 1988).

Structural break points

All deletions in the X-ray sequence were made such that the distance between the break points was a minimum. Similarly, insertions were placed at the extremities of exposed turns, which were measured both by solvent accessibility as calculated by the program DSSP (Kabsch & Sander, 1983) and by reversal in local chain direction. This property (s) was defined for an insertion between residues i and j (in a set of atomic coordinates $\mathbf{x}$) as:

$$s = (\mathbf{x}_i \cdot [\mathbf{x}_i - \mathbf{x}_{i-1}]) + (\mathbf{x}_j \cdot [\mathbf{x}_j - \mathbf{x}_{j+1}]),$$

where $\cdot$ indicates the scalar (dot) product of two vectors. When the molecule has its center of gravity on the origin, the measure s scores most highly for turns both directed away from the core and remote from the center.

A novel modeling method was used that has previously only been applied to predicted structures where structural motif constraints were available (Taylor, 1991, 1993). However, by declaring all parts of the FNR structure with a correspondence in the Cb-245 sequence to be a "motif," the same fast building method can be used. The method of Taylor (1991, 1993) progresses by alternating cycles of distance-matrix projection and real-space regularization. This maintains the global fold and all known interactions, while allowing the surface and loops to adapt. The method has the advantage that it is completely automatic and operates on a more schematic representation (α -carbons) than equivalent modeling programs, which require at least all main-chain atoms, and often side-chain atoms, to be specified. In addition to the motif regions, the large insert between residues 483 and 503 was assigned an ideal α -helix as a structural template on the basis of its predicted secondary structure.

α -Carbon to main-chain conversion

Many methods have been developed to convert an α -carbon model to a main-chain representation based on local energy minimization (Purisima & Scheraga, 1984) or fitting overlapping fragments (Claessens et al., 1989; Holm & Sander, 1991; Levitt, 1992). With the model resulting

from the preceding methods, the accuracy (and corresponding computational complexity) of these approaches is hardly justified, particularly in loop regions. Instead, the simple observation of Levitt (1978) was used, which relates local virtual bond geometry to secondary structure type. Although the method was originally used to define secondary structure, it can be used equally in reverse to construct quite an accurate backbone in secondary structure regions.

Side-chain addition

Given backbone coordinates and residue identities, the CHARMM program (Brooks et al., 1983) can produce a full atomic model. CHARMM can also then be used to adjust the structure to minimize steric clashes. This was done initially with all backbone atoms held in fixed positions, then with only α -carbons fixed, and finally with all atoms free to move. For each stage, 300, 200, and 100 hundred cycles of minimization were allowed, respectively. Although this approach is rather crude compared to more sophisticated methods based on rotamer preference (Ponder & Richards, 1987; Holm & Sander, 1991; Desmet et al., 1992), the resulting models are well packed and under casual inspection cannot be distinguished from X-ray models. The more rigorous tests of Morris et al. (1992) were also applied, which showed the model to have good stereochemistry and no significant bad contacts. In addition to the initial unconstrained minimization procedure, to allow functionally important side chains to retain the conformations observed in the X-ray structure, a final 1,000 cycles of minimization were performed using specific torsional constraints. Residues aligned with buried positions in the X-ray structure (positions marked with * in Fig. 2) were given a CHARMM torsional constraint weight of 50, and residues further presumed to be interacting with the FAD, NADP, or phosphate (marked with F, N, or P), a weight of 100. Other residue side chains were not torsionally constrained.

Manual adjustments

Detailed changes, both on the α -carbon and full atomic models, were made using the program Quanta (Molecular Simulations, Inc.). This program was also used to introduce the nucleotide into the final model by superposing the known X-ray structure with bound nucleotide onto the model to generate corresponding positions in the coordinate frame of the model.

FAD binding site

In the crystal structure of FNR (Karplus et al., 1991), the FAD molecule is extended with the flavin component buried deeply in the cleft between the two domains and the adenine component interacting loosely with the large hair-

pin loop, which extends above the cleft. This nucleotide conformation was taken directly into the Cb-245 model and minimized using CHARMM for 100 cycles.

Acknowledgments

D.T.J. was supported by an SERC CASE studentship. W.R.T. and A.W.S. acknowledge the support of the MRC and the Wellcome Trust, respectively. We thank P.A. Karplus for the refined coordinates of the ferredoxin reductase and J.H. Nugent for useful discussion. T.B.L. Kirkwood is thanked for comments on the manuscript.

References

- Abo, A., Boyhan, A., West, I., Thrasher, A.J., & Segal, A.W. (1992). Reconstitution of neutrophil NADPH oxidase activity in the cell-free system by four components: p67-phox, p47-phox, p21rac1, and cytochrome b245. *J. Biol. Chem.* 267, 16767-16770.
- Andrews, S.C., Shipley, D., Keen, J.N., Findlay, J.B.C., Harrison, P.M., & Guest, J.R. (1992). The haemoglobin-like protein (HMP) of *Escherichia coli* has ferrisiderophore reductase activity and its C-terminal domain shares homology with ferredoxin NADP reductases. *FEBS Lett.* 302, 247-252.
- Babior, B.M. & Kipnes, R.S. (1977). Superoxide-forming enzyme from human neutrophils: Evidence for a flavin requirement. *Blood* 50, 517-524.
- Bleasby, A.J. & Wootton, J.C. (1990). Construction of validated, non-redundant composite protein sequence databases. *Protein Eng.* 3, 153-159.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4, 187-217.
- Claessens, M., van Cutsem, E., Lasters, I., & Wodak, S. (1989). Modelling the polypeptide backbone with "spare-parts" from known protein structures. *Protein Eng.* 2, 335-345.
- Clark, R.A., Volpp, B.D., Leidal, K.G., & Nauseef, W.M. (1990). Two cytosolic components of the human neutrophil respiratory burst oxidase translocate to the plasma membrane during cell activation. *J. Clin. Invest.* 85, 714-721.
- Dancis, A., Roman, D.G., Anderson, G.J., Hinnebusch, A.G., & Klausner, R.D. (1992). Ferric reductase of *Saccharomyces cerevisiae*: Molecular characterization, role in iron uptake, and transcriptional control by iron. *Proc. Natl. Acad. Sci. USA* 89, 3869-3873.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Ångströms resolution. *Nature* 318, 618-618.
- Desmet, J., Demaeyer, M., Hazes, B., & Lasters, I. (1992). The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356, 539-542.
- Devereux, J., Haeblerli, P., & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387-395.
- Fillat, M.F., Edmondson, D.E., & Gomezmoreno, C. (1990). Structural and chemical-properties of a flavodoxin from *Anabaena* PCC-7119. *Biochim. Biophys. Acta* 1040, 301-307.
- Filman, D.J., Bolin, J.T., Matthews, D.A., & Kraut, J. (1982). Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Ångströms resolution: II. Environment of bound NADPH and implications for catalysis. *J. Biol. Chem.* 257, 3663-3672.
- Friemann, A., Brinkmann, K., & Hachtel, W. (1991). Sequence of a cDNA-encoding the bi-specific NAD(P)H-nitrate reductase from the tree *Betula pendula* and identification of conserved protein regions. *Mol. Gen. Genet.* 227, 97-105.
- Garcia, R.C. & Segal, A.W. (1988). Phosphorylation of the subunits of cytochrome b-245 upon triggering of the respiratory burst of human neutrophils and macrophages. *Biochem. J.* 252, 901-904.

- Garnier, J., Osguthorpe, D.J., & 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.
- Gowri, G. & Campbell, W.H. (1989). cDNA clones for corn leaf NADH:nitrate reductase and chloroplast NAD(P)⁺:glyceraldehyde-3-phosphate dehydrogenase. *Plant Physiol.* 90, 792–798.
- Harayama, S., Rekik, M., Bairoch, A., Neidle, E.L., & Ornston, L.N. (1991). Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal BENABC and *Pseudomonas putida* TOL PWWO plasmid XylXYZ, genes encoding benzoate dioxygenases. *J. Bacteriol.* 173, 7540–7548.
- Holm, L. & Sander, C. (1991). Database algorithm for generating protein backbone and side-chain coordinates from a C α trace: Applications to model building and detection of co-ordinate errors. *J. Mol. Biol.* 218, 183–194.
- Hurst, J.K., Loehr, T.M., Curnutte, J.T., & Rosen, H. (1991). Resonance Raman and electron paramagnetic resonance structural investigations of neutrophil cytochrome *b*558. *J. Biol. Chem.* 266, 1627–1634.
- Imajoh-Ohmi, S., Tokita, K., Ochiai, H., Nakamura, M., & Kanegasaki, S. (1992). Topology of cytochrome *b*558 in neutrophil membrane analyzed by anti-peptide anti-bodies and proteolysis. *J. Biol. Chem.* 267, 180–184.
- Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romana, L.K., & Reeves, P.R. (1991). Structure and sequence of the *rfb* (O-antigen) gene-cluster of *Salmonella* serovar *typhimurium* (strain-LT2). *Mol. Microbiol.* 5, 695–713.
- Johnstone, I.L., McCabe, P.C., Greaves, P., Gurr, S.J., Cole, G.E., Brow, M.A.D., Unkles, S.E., Clutterbuck, A.J., Kinghorn, J.R., & Innis, M.A. (1990). Isolation and characterization of the *cRNA*-*nia*-*niad* gene-cluster for nitrate assimilation in *Aspergillus nidulans*. *Gene* 90, 181–192.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637.
- Kakinuma, K., Kaneda, M., Chiba, T., & Ohnishi, T. (1986). Electron spin resonance studies on a flavoprotein in neutrophil plasma membranes: Redox potentials of the flavin and its participation in NADPH oxidase. *J. Biol. Chem.* 261, 9426–9432.
- Karplus, P.A., Daniels, M.J., & Herriott, J.R. (1991). Atomic structure of ferredoxin-NADP⁺ reductase: Prototype for a structurally novel flavoenzyme family. *Science* 251, 60–66.
- Klausner, R.D., Rouault, T.A., & Harford, J.B. (1993). Regulating the fate of mRNA: The control of cellular iron metabolism. *Cell* 72, 19–28.
- Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry* 17, 4277–4285.
- Levitt, M. (1992). Accurate modelling of protein conformation by automatic segment matching. *J. Mol. Biol.* 226, 507–533.
- Light, D.R., Walsh, C., O'Callaghan, A.M., Goetzl, E.J., & Tauber, A.I. (1981). Characteristics of the cofactor requirements for the superoxide-generating NADPH oxidase of human polymorphonuclear leukocytes. *Biochemistry* 20, 1468–1476.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G., & Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins Struct. Funct. Genet.* 12, 345–364.
- Nordlund, I., Powlowski, J., & Shingler, V. (1990). Complete nucleotide-sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain-CF600. *J. Bacteriol.* 172, 6826–6833.
- Nugent, J.H.A., Gratzner, W., & Segal, A.W. (1989). Identification of the haem binding subunit of cytochrome *b*245. *Biochem. J.* 264, 921–924.
- Okamoto, P.M., Fu, Y.H., & Marzluf, G.A. (1991). NIT-3, the structural gene of nitrate reductase in *Neurospora crassa*: Nucleotide-sequence and regulation of messenger-RNA synthesis and turnover. *Mol. Gen. Genet.* 227, 213–223.
- Parkos, C.A., Allen, R.A., Cochran, C.G., & Jesaitis, A.J. (1987). Purified cytochrome *b* from human granulocyte plasma membrane is comprised of two polypeptides with relative molecular weights of 91,000 and 22,000. *J. Clin. Invest.* 80, 732–742.
- Parkos, C.A., Dinanuer, M.C., Walker, L.E., Allen, R.A., Jesaitis, A.J., & Orkin, S.H. (1988). Primary structure and unique expression of the 22-kilodalton light chain of human neutrophil cytochrome *b*. *Proc. Natl. Acad. Sci. USA* 85, 3319–3323.
- Ponder, J. & Richards, F. (1987). Tertiary templates for proteins: Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193, 775–791.
- Prosser, I.M. & Lazarus, C.M. (1990). Nucleotide-sequence of a spinach nitrate reductase cDNA. *Plant Mol. Biol.* 15, 187–190.
- Purisima, E.O. & Scheraga, H.A. (1984). Conversion from a virtual-bond chain to a complete polypeptide backbone chain. *Biopolymers* 23, 1207–1224.
- Roos, D., de Boer, M., de Klein, A., Bolscher, B.G.J.M., & Weening, R.S. (1993). Chronic granulomatous disease: Mutations in cytochrome *b*558. *Immunodeficiency* 4, 289–302.
- Rotrosen, D., Kleinberg, M.E., Nunoi, H., Leto, T.L., Gallin, J.I., & Malech, H.L. (1990). Evidence for a functional cytoplasmic domain of phagocyte oxidase cytochrome *b*558. *J. Biol. Chem.* 265, 8745–8750.
- Royer-Pokora, B., Kunkel, L.M., Monaco, A.P., Goff, S.C., Newburger, P.E., Baehner, R.L., Cole, F.S., Curnutte, J.T., & Orkin, S.H. (1986). Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature* 322, 32–38.
- Schluchter, W.M. & Bryant, D.A. (1992). Molecular characterization of ferredoxin-NADP⁺ oxidoreductase in cyanobacteria: Cloning and sequence of the *petH* gene of *Synechococcus* sp. PCC 7002 and studies on the gene product. *Biochemistry* 31, 3092–3102.
- Segal, A.W. (1987). Absence of both cytochrome *b*-245 subunits from neutrophils in X-linked chronic granulomatous disease. *Nature* 326, 88–91.
- Segal, A.W. (1989). The electron transport chain of the microbicidal oxidase of phagocytic cells and its involvement in the molecular pathology of chronic granulomatous disease. *J. Clin. Invest.* 83, 1785–1793.
- Segal, A.W. & Jones, O.T.G. (1978). Novel cytochrome *b* system in phagocytic vacuoles from human granulocytes. *Nature* 276, 515–517.
- Segal, A.W., West, I., Wientjes, F., Nugent, J.H.A., Chavan, A.J., Haley, B., Garcia, R.C., Rosen, H., & Scrase, G. (1992). Cytochrome *b*245 is a flavocytochrome containing FAD and the NADPH-binding site of the microbicidal oxidase of phagocytes. *Biochem. J.* 284, 781–788.
- Shaw, J.P. & Harayama, S. (1992). Purification and characterization of the NADH-acceptor reductase component of xylene monooxygenase encoded by the *tol* plasmid pWWO of *Pseudomonas putida* MT-2. *Eur. J. Biochem.* 209, 51–61.
- Stainthorpe, A.C., Lees, V., Salmond, G.P.C., Dalton, H., & Murrell, J.C.T. (1990). The methane monooxygenase gene-cluster of *Methylococcus capsulatus* (BATH). *Gene* 91, 27–34.
- Taylor, W.R. (1988). A flexible method to align large numbers of biological sequences. *J. Mol. Evol.* 28, 161–169.
- Taylor, W.R. (1991). Towards protein tertiary fold prediction using distance and motif constraints. *Protein Eng.* 4, 853–870.
- Taylor, W.R. (1993). Protein fold refinement: Building models from idealized folds using motif constraints and multiple sequence data. *Protein Eng.* 6, in press.
- Yubisui, T., Naitoh, Y., Zenno, S., Tamura, M., Takeshita, M., & Sakaki, Y. (1987). Molecular cloning of cDNAs of human liver and placenta NADH-cytochrome *b*-5 reductase. *Proc. Natl. Acad. Sci. USA* 84, 3609–3613.