The NifL PAS domain

- Insight into its structure and function -

Marco H. Hefti

Promotor: Prof. dr. Sacco de Vries
Hoogleraar in de Biochemie
Departement Agrotechnologie en Voedingswetenschappen
Laboratorium voor Biochemie
Wageningen Universiteit

Co-promotor: dr. ir. Jacques Vervoort
Universitair hoofddocent
Departement Agrotechnologie en Voedingswetenschappen
Laboratorium voor Biochemie
Wageningen Universiteit

Promotiecommissie: Prof. Ray Dixon, John Innes Centre, Norwich
Prof. dr. Ton Bisseling, Wageningen Universiteit
dr. Cristofer Enroth, EMBL-Outstation, Hamburg
dr. Willem van Berkel, Wageningen Universiteit

PROPOSITIONS

- The statement by Hill *et al.* that NifL contains 3 FAD molecules per tetramer, and the statement by Hefti *et al.* that the NifL PAS domain contains 2 FAD molecules per tetramer, are both incorrect.

  Hill, S., S. Austin, T. Eydmann, T. Jones, & R. Dixon (**1996**) *Azotobacter vinelandii* NifL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch, *Proceedings of the National Academy of Sciences of the United States of America* **93**, 2143-2148.

  Hefti, M. H., C. J. G. Van Vugt - Van der Toorn, R. Dixon, & J. Vervoort (**2001**) A novel purification method for histidine-tagged proteins containing a thrombin cleavage site, *Analytical Biochemistry* **295**, 180-185.

- The nomenclature S1 / S2, PAS-A / PAS-B, PAS domains / PAC motifs, and LOV domains should be abandoned, and instead the name 'PAS fold' should be used.

  This thesis, chapter 7.

- NifL is a hemo-flavo-protein.

- When performing large-scale modelling of protein sequences, the time it takes to perform the calculations is never the rate limiting step.

- Crystallisation is art, crystallography is science.

- Making a model of a horse from photographs does not tell how fast it will run.

  Gutfreund, H. & J. R. Knowles (**1967**) The foundations of enzyme action, *Essays in Biochemistry* **3**, 25-72.

  This thesis, chapter 7.

- The graduate school Experimental Plant Sciences sends emails which, according to their own definition, can be classified as SPAM.

  "*If you set your e-mail programme to prevent you from SPAM, it will refuse messages to great numbers of recipients. Since EPS-NEWS is distributed to several hundreds of addresses it will also be refused*", EPS email newsletter 251, 10 February **2003**.

- The European Union divides the EC contribution to research infrastructures into arbitrary classes. By doing so, they give a false representation of the total amount of money really contributed to the field of nuclear physics.

  Enhancing Access to Research Infrastructures 1998-2002, technical review of IHP-ARI contracts (**2003**).

Propositions belonging to the thesis:
"The NifL PAS domain: Insight into its structure and function"

Marco H. Hefti
Wageningen, 29 April 2003

The NifL PAS domain

- Insight into its structure and function -

Marco H. Hefti

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. ir. L. Speelman,
in het openbaar te verdedigen
op dinsdag 29 april 2003
des namiddags te half twee in de Aula

ISBN: 90-5808-809-X

Wanderers in the land of Osten Ard are cautioned not to put blind trust in old rules and forms, and to observe all rituals with a careful eye, for they often mask being with seeming.

The Qanuc-folk of the snow-mantled Trollfells have a proverb.

> 'He who is certain he knows the ending of things when he is only beginning them is either extremely wise or extremely foolish; no matter which is true, he is certainly an unhappy man, for he has put a knife in the heart of wonder.'

More bluntly, new visitors to this land should take heed:

> 'Avoid assumptions.'

The Qanuc have another saying:

> 'Welcome stranger. The paths are treacherous today.'

The Dragonbone Chair - Tad Williams

Contents

Abbreviations and Symbols

DAM	dummy atom model
GOGAT	glutamate synthase, or glutamine:2-oxoglutarate amidotransferase
GS	glutamine synthetase
HIC	hydrophobic interaction chromatography
IMAC	immobilised metal-ion affinity chromatography
IPTG	isopropylthio-β-D-galactoside
LB-medium	Luria-Bertani medium
MALDI-TOF	matrix-assisted laser desorption/ionisation - time of flight
NifA	nitrogen fixation regulatory protein A (*nif*A product)
NifL	nitrogen fixation regulatory protein L (*nif*L product)
Ni-NTA	nickel-nitrilotriacetic acid
NtrC	nitrogen regulatory system protein C (*ntr*C product)
PABA	para amino benzoic acid
PAS-domains	PER-ARNT-SIM homology regions
PHBH	para-hydroxybenzoate hydroxylase
PYP	bacterial blue-light photosensor Photoactive Yellow Protein
SAXS	small-angle X-ray scattering
TFA	trifluoroacetic acid

CHAPTER 1

General Introduction into the Field of Nitrogen Fixation, NifL, and PAS Domains

NITROGEN FIXATION

Biological N_2 fixation involves the enzymatic reduction of N_2 to ammonia. The ammonia produced can subsequently be incorporated in the organism by enzymatic means, to be used for the growth and maintenance of the cell. N_2 fixation is unique to bacteria. Animals and plants that need to fix N_2 must do so in association with bacteria. An understanding of biological N_2 fixation is essential to elucidate the dynamics of the global nitrogen cycle (Figure 1.1). With the exception of water, nitrogen is generally considered to be the most limiting nutrient for growth of plants in their natural environment. The capability of fixing N_2 is recognised as a process of great agronomic importance, and a variety of leguminous plants and some non-leguminous plants can obtain their nitrogen from the air by symbiotic association with micro-organisms.

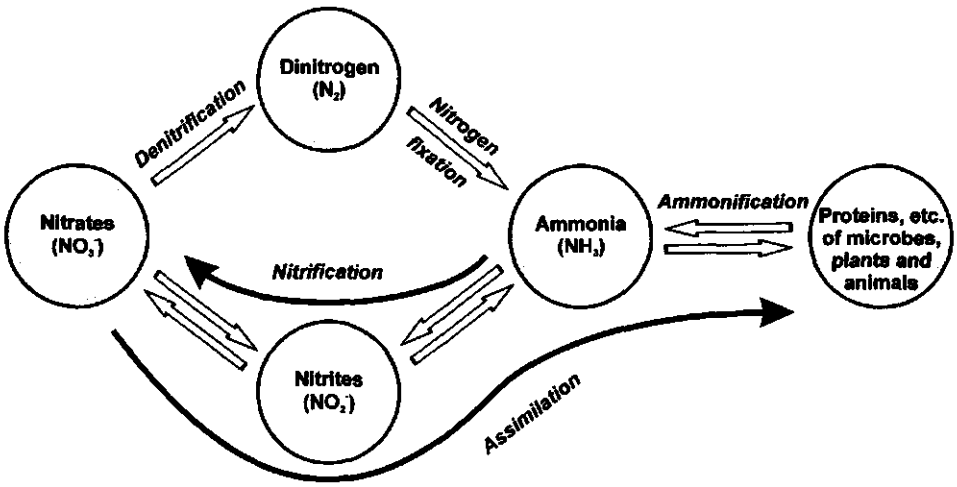


Figure 1.1 The global nitrogen cycle.

The nitrogen cycle symbolises the transformations undergone by the element nitrogen (N) on this planet (Figure 1.1). The bottom section shows the assimilation of nitrogenous living matter (mainly protein) from inorganic nitrogen compounds (nitrate, nitrite, and ammonium ions) during growth of plants and their consumption by animals. They then return to the soil as a result of decay and putrefaction of plant and animal material. The upper section shows the loss of nitrogen to the atmosphere from nitrates, and its return to the cycle by the process known as nitrogen fixation. The element nitrogen is an essential constituent of all organisms, and in nearly all agricultural areas of this planet, biological productivity is determined by the availability of inorganic nitrogen in the soil.

Nitrogen fixing bacteria, diazotrophs, are capable to convert N_2 to NH_3 by electron reduction and protonation of gaseous dinitrogen. Nitrogenase is the enzyme complex in diazotrophs responsible for this N_2 fixation. The biosynthesis of NH_3 is determined by

15 to 20 different nitrogen fixation (*nif*) gene products[54]. The *nif* genes may be carried on plasmids, as is the case in most *Rhizobium* species. More commonly, these *nif* genes are located on the major chromosome in free-living and associative nitrogen-fixing bacteria. These genes and their genomic arrangement is rather conserved in the various diazotrophs[54]. Comparison of the physical and genetical maps of *nif*-specific genes from *Klebsiella pneumoniae*, *Azotobacter vinelandii*, *Rhodobacter capsulatus*, *Bradyrhizobium japonicum*, and *Clostridium pasteurianum* reveals that spacial arrangements of the individual *nif* genes are not always conserved at the interspecies level. For example, *nif*W and *nifZ* genes are located adjacent to each other in *K. pneumoniae* and *A. vinelandii*, but not in *R. capsulatus*[54]. The primary products of the nitrogenase genes are not capable of catalysing N_2 fixation. The products of a number of other nif-specific genes are required for the maturation of the nitrogenase structural components.

Ammonia, the product of nitrogen fixation, is assimilated in diazotrophs by the enzyme glutamine synthetase (GS). In this reaction, ammonia is incorporated into glutamate to yield glutamine. Glutamate is regenerated by the activity of GOGAT (glutamate synthase).

$$NH_3 + \text{L-glutamate} + ATP \xrightarrow{GS} \text{L-glutamine} + ADP + P_i$$

$$\text{L-glutamine} + \alpha\text{-ketoglutarate} + NADPH \xrightarrow{GOGAT} 2\text{-glutamate} + NADP^+$$

By the combined activities of GS and GOGAT, cells are supplied with glutamate and glutamine. Glutamate is the source of α-amino groups of all amino acid residues, and half the nitrogen in pyrimidine, purine and imidazole rings, and the amino group of adenine. These together represent approximately 90 % of all nitrogen-containing metabolites. Glutamine, on the other hand, provides nitrogen in amino-sugars, NAD, para amino benzoic acid (PABA) and the other nitrogens in purines, pyrimidines, histidine and tryptophan. Together, these represent about 10 % of all nitrogen-containing metabolites[239].

The activity and/or synthesis of nitrogenase in diazotrophs is inhibited by an abundance of fixed nitrogen compounds in the environment. In addition to transcriptional control of *nif* genes in response to fixed nitrogen, nitrogenase enzyme activity itself may be inhibited in the presence of ammonium. This inactivation occurs for instance in several nitrogen-fixing bacteria including Cyanobacteria, *Rhodobacter*, and *Azospirillum*[130, 330]. In the latter three organisms, the nitrogenase Fe-protein is modified by attachment of an ADP-ribose moiety by the DraT enzyme in cells grown with excess ammonium, causing nitrogenase inactivation. The ADP-ribose is removed by the DraG enzyme when nitrogen becomes limiting[130]. In all free-living or associative diazotrophs examined, excess ammonium prevents *nif* gene expression.

Regulation of nitrogenase (*nif*) gene expression in *Azotobacter vinelandii* is complicated by the fact that the organism has the ability to synthesise two alternative nitrogenases. Expression of all three nitrogenase systems is repressed by high concentrations of fixed nitrogen, and expression of the individual systems is determined by the metal availability. Nitrogenases require anaerobic conditions, a source of low-potential electrons, and a large amount of ATP to reduce N_2 to NH_3[36, 118]. Consistent with the oxygen-sensitivity of these enzymes, their synthesis is repressed under aerobic regimes where nitrogen fixation is inhibited[115]. The interrelationship of these different gene systems and the co-ordination of their regulation have been investigated. Discrimination between expression of different systems must occur either by selective expression of the different *nif*A homologues or by control of the activity of their products. *nif*A is a regulatory gene required for activation of *nif* gene transcription.

In *K. pneumoniae*, *H. seropedicae* and *R. capsulatus*, expression of the *nif*A gene is prevented in cells grown in ammonium enriched medium[80, 184, 267]. In *K. pneumoniae*, *A. vinelandii*, *A. brasilense*, *H. seropedicae* and *R. capsulatus*, the activity of the NifA protein is inhibited in cells grown in ammonium. In the former two, but not the latter three organisms, the *nif*A gene lies downstream from and is cotranscribed with *nif*L[25]. In *A. vinelandii* and *K. pneumoniae*, the NifL protein is a sensor for the environmental status of both ammonium and oxygen supply and if either is plentiful or excessive, NifL inactivates NifA so that transcription of the *nif* genes does not occur.

NifL

NifL is a flavoprotein, which can inactivate NifA depending on the redox state of the cell[116, 253]. The NifL protein is a sensor for the environmental status of both ammonium and oxygen supply. If either of the latter two molecules is present, NifL inactivates NifA so that transcription of the *nif* genes does not occur. How NifL is converted from its inert form to an active NifA-inhibitory form in cells with ammonium (or other repressive fixed N sources), and conversely, how active NifL is converted to its inert form in the absence ammonium or oxygen, are key questions in understanding *nif* gene regulation in this nitrogen fixing member of the gamma group of *Proteobacteria*.

In *A. vinelandii*, the activation of *nif* gene expression by the transcriptional regulatory enhancer binding protein NifA is controlled by the sensor flavoprotein NifL in response to changes in oxygen level or the level of fixed nitrogen[116, 194]. This regulation provides a tight control of nitrogen fixation and avoids the unnecessary consumption of energy. Inhibition of NifA activity occurs via direct protein-protein interactions, and requires stoichiometric amounts of both proteins. This is shown by immunological studies and the demonstration of complex formation between NifL and NifA by co-chromatography and the 'yeast-two-hybrid'-system[74, 145]. The formation of the inhibitory complex between NifL and NifA may be regulated by the intracellular ATP/ADP ratio. Adenosine dinucleotides promote complex formation *in vitro*, allowing isolation of the

NifL-NifA complex. Removal of the dinucleotide causes dissociation of the complex. The N-terminal domain of NifA and the C-terminal region of NifL potentiate the ADP-dependent stimulation of NifL-NifA complex formation. The NifL-NifA complex is composed of a tetramer of NifL and a tetramer of NifA[193, 194].

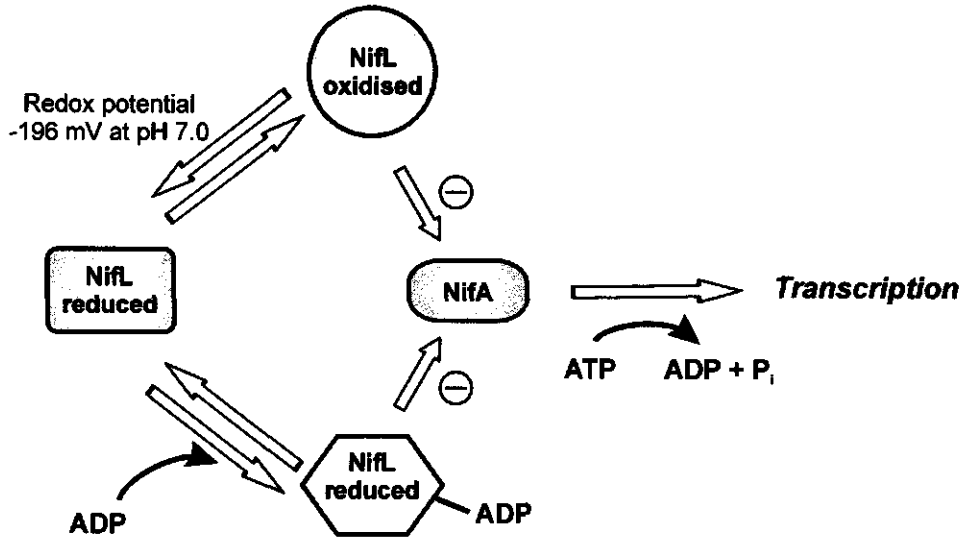


Figure 1.2 The proposed conversion of inactive NifL to active protein forms that inhibit NifA (after[57]).

NifL is a novel redox-sensing flavoprotein that also responds to adenosine nucleotides *in vitro*, particularly ADP[116, 265]. This response overrides the influence of redox status on NifL, and is also observed with refolded NifL apoprotein, which lacks the flavin moiety. The ability of NifL to inhibit NifA activity is responsive to the oxidation state of the chromophore, indicating that NifL is a redox-sensitive regulator.

Redox Sensing

The redox- and nitrogen-responsive functions of NifL are discrete[265]. The oxidised form of NifL is competent to inhibit NifA activity, but when the flavin moiety is reduced with sodium dithionite, NifA activity is not inhibited[74]. Gel filtration experiments with native NifL (residues 1-519) have shown that the protein is a tetramer in solution. The N-terminal domain (amino acid residues 1-284) of the NifL protein also forms a tetramer in solution[265]. NifL can interact with appropriate electron donors to maintain NifL in the reduced (inactive) state under anaerobic conditions. The redox potential of *A. vinelandii* NifL is -226 mV at pH 8.0, and –196 mV at pH 7.0, as determined by xanthine oxidase/xanthine in the presence of appropriate mediators. This suggests that NifL might be reduced by a number of different electron donors and NAD(P)H-dependent enzymes. Re-oxidation of NifL occurs rapidly in the presence of air[155].

Although NifL is maintained in an inactive state under reducing conditions *in vitro*, high NifL/NifA ratios inhibit activation under anaerobic nitrogen-limiting conditions *in vivo*. NifL is reduced relatively slowly, while re-oxidation is fast, as expected for a reduced flavoprotein. This provides an appropriate redox switch, allowing electron transfer to oxygen as acceptor at a faster rate than electron donation to NifL. The low reduction rate of NifL is consistent with the requirement for sustainable reducing conditions to support nitrogenase activity. The synthesis of nitrogenase is a large investment in biosynthetic capacity. A gradual switching on of *nif* transcription would be economic. As a result of the oxidation of NifL by oxygen, a fast switching off of the transcription of *nif* genes would prevent the wasteful synthesis of proteins that are otherwise destinated for inactivation[155].

Carboxyterminal domain of NifL

Sequence analysis of NifL indicates that the protein is composed of two domains, separated by a glutamine-rich flexible linker (Figure 1.3). The C-terminal domain shows strong homology with the histidine protein kinase family of two-component regulatory proteins, and binds adenosine dinucleotides[265].

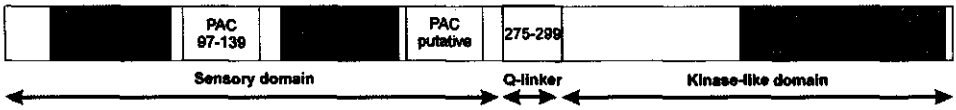


Figure 1.3 The flavoprotein NifL is comprised of 2 domains, tethered by a Q-linker.

The *A. vinelandii* NifL protein possesses conserved regions also found in other transmitter domains. Although NifL contains a conserved histidine residue at position 304, known to be phosphorylated in other members of this family, most mutations at this residue do not effect the function of NifL. This implies that sensory transduction by NifL does not involve phosphorylation of this residue[6, 144, 323]. Moreover, phosphotransfer to NifA has not been detected *in vitro*[6, 144].

Q-linker region of NifL

Q-linkers are short (mostly 20 to 30 residues), hydrophilic sequences rich in glutamine, arginine, glutamate, serine and proline[324]. They serve as a flexible linker or a 'coil' to connect the structurally-distinct but interacting domains of some regulatory proteins[9, 138, 179]. The NifL Q-linker has a length of about 25 amino acid residues, connecting the C-terminal kinase-like domain and the N-terminal sensory domain.

Aminoterminal domain of NifL

NifL contains an N-terminal sensory domain, required for the redox response[265]. This domain shows homology to the *bat* gene product from *Halobacterium halobium*, which potentially has an oxygen sensing function[25]. The N-terminal domain is

comprised of two PAS domains. Each PAS domain has a PAC like motif at its C-terminal side.

PAS DOMAINS

PAS domains are ubiquitous motifs present in archaea, bacteria and eucarya[332]. PAS domains are found in sensor proteins and are named after homology between the Drosophila period protein (PER), the aryl hydrocarbon receptor nuclear translocator protein (ARNT) and the Drosophila single-minded protein (SIM). These domains are sometimes referred to as LOV domains; light-, oxygen or voltage domains[33, 44, 45, 49, 50, 127]. PAS-domains, unless most other sensory domains, are located in the cytoplasm[287] and include serine/threonine kinases[231], histidine kinases[1], photoreceptors and chemoreceptors for taxis and tropism[269], cyclic nucleotide phosphodiesterases[266], circadian clock proteins[128, 251], voltage-activated ion channels[316], as well as regulators of responses to hypoxia[121] and embryological development of the central nervous system[210]. To date, most PAS domains known have a cofactor associated with this domain. This cofactor partly determines the specificity of a PAS domain for detecting input signals.

For understanding the different mechanisms in which PAS domains work detailed information about their cofactor binding region is needed. In the PFAM Protein Families Database[13] there are currently (version 7.8) 958 PAS domains present in 607 different proteins. According to PFAM, a PAC motif is found at the C-terminus of a subset of known PAS domains, and is proposed to contribute to the PAS domain fold[332]. For comparison of proteins it is necessary to abandon the use of the commonly used annotations S1 / S2[332], PAS-A / PAS-B[48, 117], and PAS domain / PAC motif[231] which are now in use to specify sequence similarities. The LOV domain is a PAS domain as well. It is more important to discuss the PAS functional fold rather than sequential relations describing common motifs. The basis for a redefinition of the PAS structural fold is provided in Chapter 7. When referring to this structural element, we would like to suggest the term PAS fold.

OUTLINE OF THIS THESIS

The EU-Biotechnology project 'TASTES: Transcriptional Activation and Sensory Transduction; Elucidation of Structures' started as a collaboration between several groups in the EU: The John Innes Centre in Norwich (Martin Drummond, Ray Dixon), the Imperial College in London (Martin Buck), the CNRS in Toulouse (Daniel Kahn and Jean-Pierre Samama), the Wageningen University (Jacques Vervoort), and the EMBL in Hamburg (Paul Tucker and Dmitri Svergun). Within this consortium I have focussed on the PAS domain of the nitrogen fixation regulatory protein NifL from *Azotobacter*

vinelandii. A general introduction into the field of nitrogen fixation, as well as the biochemical importance of PAS domains is given in Chapter 1. The NifL protein is a flavoprotein, with FAD as the prostethic group. For studying flavoproteins, it might be necessary to remove this prosthetic group, or replace it with an isotopically labeled variant. Chapter 2 summarises the tools currently available within the field of flavoprotein deflavination and reconstitution. In Chapter 3 a new purification method for His-tagged proteins is described. This purification method makes use of the commonly incorporated Histidine-tag. On-column cleavage of the protein with thrombin facilitates the separation of the protein of interest and the His-tag. This procedure was tested for four different proteins, and proved to be a novel tool for protein purification. In Chapter 4 the His-tag is again used as a tool, this time to deflavinate and reconstitute the NifL PAS domain protein. The method described is efficient, as shown by NMR spectroscopy. Crystals of the NifL PAS domain were obtained and diffracted to 3 Ångstrom resolution. Chapter 5 describes the current status of the structure elucidation of this domain, using X-ray crystallography. SAXS (Small-Angle X-ray Scattering) studies with this domain clearly showed the tetrameric state of the PAS domain. In an envelope, created using SAXS data, four monomeric models of the PAS domain were fitted, in order to elucidate the structural arrangement of these four monomers. Details are given in Chapter 6. In Chapter 7 the term PAS fold is introduced to denote a three-dimensional fold present in several proteins: The photoactive yellow protein (PYP) from *Ectothiorhodospira halophila*[27, 92], the heme-binding domain of the Rhizobial oxygen sensor FixL from *Bradyrhizobium japonicum*[97], and from *Rhizobium meliloti*[190], the N-terminal domain of the human ether-a-go-go-related potassium channel HERG[198], the FMN containing phototropin module of the chimeric fern Adiantum photoreceptor[49, 50] and the average NMR structure of the N-terminal PAS domain of human PAS kinase[4]. We demonstrate that sequence comparisons alone are not sufficient to correctly annotate and identify PAS domains in newly discovered protein sequences. This approach leads to a pool of hypothetical PAS domain containing proteins. Modelling studies, and subsequent independent objective analysis of the 3D models give more insight into this intriguing family of sensory proteins. In Chapter 8 some concluding remarks are given, summarising the main conclusions of the preceding chapters.

CHAPTER 2

Deflavination and Reconstitution of Flavoproteins

-Tackling Fold and Function-

A modified version of this chapter has been submitted to the European Journal of Biochemistry

ABSTRACT

Flavoproteins are ubiquitous proteins that are involved in many biological processes. In the large majority of flavoproteins, the flavin cofactor is non-covalently bound, providing the opportunity to replace the natural cofactor by flavin analogs. Studying apoflavoproteins and their reconstituted holo forms can give valuable insight in the thermodynamics of flavin binding, protein-flavin and subunit interactions, flavoprotein folding and stability, regulation processes, reaction mechanisms and redox potential regulation. In this review we summarise the advances made in the field of flavoprotein deflavination and reconstitution. Several chromatographic procedures are described that can be used to either deflavinate or reconstitute the flavoprotein on a large scale. These chromatographic procedures can combine the use of affinity tags both for the purification of the protein of interest as well as replacement of the protein-bound cofactor. Protein engineering of covalent flavoproteins provides insight into the mechanism of covalent flavinylation and the rationale for this atypical protein modification. Understanding the flavin-apoprotein interaction is also of medical relevance. Several diseases are associated with mutations in human flavoprotein genes. These mutations may result in enzyme inactivation but can also lead to impaired flavin binding.

INTRODUCTION

Flavoproteins are involved in many biological processes and are present in all kingdoms of life. Flavoproteins contain a flavin prosthetic group, most often non-covalently bound. The dissociation constants of the apoprotein-flavin complexes are usually between 10^{-7} and 10^{-10} M. Therefore, in most cases, it is impossible to remove the flavin by simple dialysis. The predominant flavin cofactors are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Both cofactors are synthesised *in vivo* from riboflavin (vitamin B2) by the action of riboflavin kinase and FAD synthetase[158, 268]. The structures of riboflavin, FMN and FAD are shown in Figure 2.1.

The isoalloxazine ring system is the functional part of the flavin prosthetic group. It may undergo one- or two-electron transitions which are modulated by the surrounding protein structure. The versatility of one- or two-electron transitions make flavoproteins to be at the crossroads of cellular redoxchemistry. The electronical and geometrical properties of protein-bound flavin in the different redox forms, i.e. the oxidised state, the one-electron reduced semiquinone state, and the two-electron reduced hydroquinone state, are of main importance for determining flavoprotein activity. There are large differences in the redox potential of flavoproteins[160, 271, 302], ranging from -367 mV for nitroalkane oxidase[90] to +55 mV for vanillyl-alcohol oxidase (VAO)[84]. In this paper we will present an overview of the field of flavoprotein deflavination and reconstitution.

First, however, some general considerations about the interactions between flavoproteins and their cofactors are made.

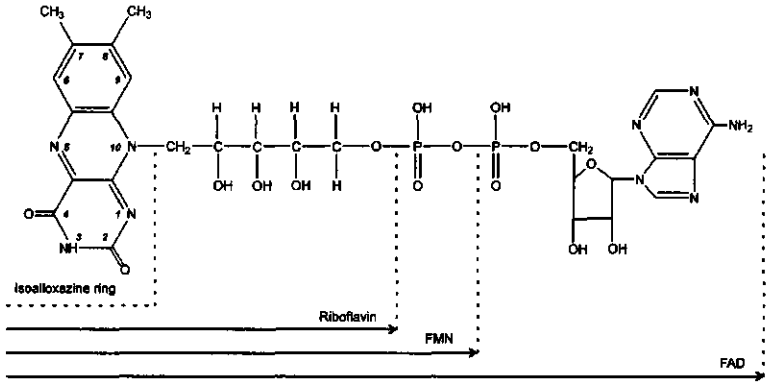


Figure 2.1 Schematic representation of the structures of riboflavin, FMN and FAD.

Thermodynamics of flavin binding

The binding interaction between apoproteins and flavin prosthetic groups have been extensively studied. The strong and specific binding of FMN or FAD to apo-flavoproteins is driven by the enthalpic contribution to the free energy change of binding[154, 171, 276]. The thermodynamics of flavin binding to *Azotobacter vinelandii* as well as *Anabaena* flavodoxin has been studied in detail[11, 37, 154, 235, 303]. In the study of Carlson and Langerman[37] it was shown that the enthalpy, entropy and free energy associated with the binding of FAD, 8-carboxyriboflavin, or the natural cofactor FMN to *A. vinelandii* apoflavodoxin were all negative. However, striking differences in the enthalpy for binding FMN, FAD and 8-carboxyriboflavin (-28.3, -16.6, and -14.0 kcal/mol respectively) were observed. It was concluded that the difference in binding enthalpy between FMN and 8-carboxyriboflavin results from the phosphate group in FMN, which binds in a predefined binding pocket in the apoflavodoxin at the N-terminus of an α-helix. The unique hydrogen bonding network surrounding the FMN-phosphate group stabilises the FMN-apoprotein complex[225]. This observation was recently corroborated by Losato et al.[154], who dissected the binding energies of the Anabaena apoflavodoxin-FMN complex in great detail, using a large variety of methods: site-directed mutagenesis, titration calorimetry, equilibrium binding constant determinations, and X-ray crystallography. It was shown that the contribution of the phosphate to the binding energy is the greatest (7 kcal/mol), that the contribution of the isoalloxazine is around 5-6 kcal/mol, and that the ribityl side chain contributes only 1 kcal/mol. Both the phosphate and the isoalloxazine contribute significantly to the enthalpy of binding. The ionic strength of the solution does not have a large effect on the stability of the apoflavodoxin-FMN complex.

The thermodynamics of FAD binding to D-amino acid oxidase (DAAO) has been studied by Matteo and Sturtevant[171]. The free energy of binding was shown to be largely independent of temperature. However, the enthalpy and the entropy of the binding interaction were strongly temperature dependent. In contrast to the binding of FMN to apoflavodoxin, where the entropy strongly opposes binding, the binding of FAD to DAAO is enforced by a large positive entropic contribution due to a decrease in the exposure of non-polar groups to the solvent and a smaller negative entropic contribution resulting from a tightening of the protein structure with losses in (vibrational) heat capacity when the coenzyme is bound. In general it is clear that (preformed) hydrogen bonding networks are important in flavin-binding, resulting in large negative enthalpy factors and that the entropic contribution can be either negative in rigid apoproteins (flavodoxins) or positive in more flexible apoproteins (DAAO).

The enthalpic and entropic contributions of flavin binding are very important, when dealing with apoflavoprotein preparation. First the enthalpic contribution should be decreased in order to release the flavin from the holoprotein and second the entropic contribution in the preparation of the apoprotein should be decreased. The enthalpic contribution in flavin binding can be decreased by the addition of solutes which interfere with the specific interactions between flavin cofactor and apoprotein. These solutes can be unfolding agents (urea, guanidinium hydrochloride) which break hydrogen bonds, or monovalent anions (chloride, bromide) which influence protein hydration. Phosphate and pyrophosphate can be used additionally[63] as these compounds can bind specifically in the phosphate or pyrophosphate binding pocket of FMN and FAD containing flavoproteins, thereby diminishing the binding interaction of the flavin prosthetic group with the apoprotein[235, 264]. Often a combined effect of urea and bromide provides the best result in the preparation of apoflavoproteins. In those cases where the apoprotein is relatively unstable it is important to positively influence the entropic contribution in the preparation of apoflavoproteins. Many FAD-containing flavoproteins are less stable when they loose their cofactor and care should be taken when removing the flavin prosthetic group. In several studies performed within our group we have shown that in those cases where the apoprotein is unstable, the best approach in preparation of apoflavoproteins is to bind the holoprotein covalently or non-covalently to a chromatographic support, thereby diminishing the entropic contributions.

Flavin binding sites

In flavoprotein biochemistry we deal with a combination of several factors influencing the different flavin binding modules. Nature facilitated the binding of flavins to proteins by the evolution of different flavin binding folds[62, 82, 172]. Within these folds, specific parts of the flavin molecule are recognised by different structural motifs, i.e. the P-P loop, the TIM barrel, the GG motif and the Rossmann fold. These structural motifs are characterised by their unique fingerprint sequences[66, 70, 81, 83, 260, 277, 284, 293]. Many

proteins that do bind FAD do not interact with FMN, and vice versa, and this is reflected in these fingerprint sequences. A known exception is methylenetetrahydrofolate reductase (MTHFR). This enzyme has a β8α8-barrel, which resembles the structure of FMN-binding proteins rather than FAD-binding proteins. The FAD cofactor of MTHFR, however, binds to a different part of the flavin binding domain compared to the barrels that bind FMN[101].

Most flavoproteins bind their cofactor in a non-covalent mode. However, in some flavoproteins, the isoalloxazine ring of the flavin is covalently linked to either a histidine, tyrosine, or cysteine residue[185]. The rationale for covalent flavinylation is not always clear, but for VAO from *Penicillium simplicissimum* it was shown that the covalent linkage between the C8α atom of the isoalloxazine ring of the flavin and the N3 atom of His422 raises the redox potential of the flavin, thereby facilitating substrate oxidation[84]. Moreover, from crystallographic analysis of non-covalent VAO variants and the VAO apoprotein, it could be established that the flavin binds in this enzyme to a preorganised binding site and becomes covalently linked to the apoprotein by a self-catalytic mechanism[85]. Members of the VAO family have a remarkable tendency for covalent anchoring of the flavin cofactor[83], whereas flavoproteins with a Rossmann fold[82, 174] clearly prefer a non-covalent flavin binding mode. Human monoamine oxidase[23, 24, 129] is one of the few Rossmann fold enzymes that contains a non-dissociable covalently bound cofactor.

A new type of flavin attachment was recently reported for the NqrB and NqrC subunits of the Na^+-translocating NADH-quinone oxidoreductase (NQR) from *Vibrio alginolyticus*[110, 209]. From MALDI-TOF MS analysis of proteolytic digests, it was concluded that in both NqrB and NqrC subunits, the FMN cofactor is attached by its 5'-phosphate moiety to a threonine of the polypeptide chain. In agreement with this, no covalent flavin was detected in the efficiently expressed Thr225Leu mutant of NqrC from *Vibrio cholerae*[12]. Furthermore, from sequence comparisons it was predicted that this novel type of phosphoester binding between FMN and the apoprotein is conserved in the NQR sodium pumping systems of a number of marine and pathogenic bacteria and that in some of these systems the target threonine is replaced by a serine residue[109].

There are also flavoproteins that have more than one flavin binding site. Among these diflavin enzymes is NADPH-cytochrome P450 reductase (CPR) which contains both a FMN- and a FAD-binding domain in order to shuttle electrons from NADPH to the P450 heme. When removing flavin from CPR, FMN is released more easily than FAD[26]. The FMN-binding domain of CPR has a structure similar to flavodoxin, whereas the FAD binding domain relates to ferredoxin-$NADP^+$ reductase (FNR)[314, 331]. CPR accepts a pair of electrons from NADPH, with FAD and FMN being the port of entry and exit, respectively. The electrons are transferred, one at a time, to cytochrome P450. The two flavins are not stacked, but rather are end-on, and communicate witch each other through the 7- and 8-methyl groups of the xylene rings. The two isoalloxazine

rings are in an angle of about 150°, with the closest distance between them being about 4 Å[314].

Nitric-oxide synthase (NOS)[169, 249] and flavocytochrome P450 BM3[206] contain FAD, FMN, and heme b in one single polypeptide chain. Comparison of the structure of the neuronal NOS FAD/NADPH domain with CPR reveals a strict conservation of the FAD-binding site. This suggests that the hydride transfer mechanisms of both enzymes are very similar. In contrast, the FMN-binding site surface of NOS is less positively charged than CPR, indicating a different nature of interaction between the two flavin domains, and a different mode of regulation in electron transfer[329]. Flavocytochrome P450 BM3 is a bacterial P450 system in which a fatty acid hydroxylase P450 is fused to a mammalian-like diflavin NADPH-P450 reductase. The enzyme is hydrophilic (unlike mammalian P450 redox systems) and its fusion arrangement results in the highest catalytic activity of any P450 monooxygenase[206].

Another member of the family of diflavin reductases is mammalian methionine synthase reductase[142], a monomeric protein of 78 kDa with both a FAD and a FMN cofactor[215]. In bacteria, the physiological reduction system involved in methionine synthesis is comprised of ferredoxin oxidoreductase and flavodoxin[88, 107, 108]. Since mammals lack flavodoxin, methionine synthase reductase has been recruited as electron donor in these systems for methionine synthase activation[215].

Other members of the emerging family of dual flavoproteins include sulfite reductase[40, 71, 72, 219, 328], and the novel reductase 1 (NR1)[221] which appears to be widely expressed in human cancer cell lines and, therefore, could play a potential role in the activation (or deactivation) of drugs used in cancer therapy.

Artificial Flavins

In order to gain insight into how the protein environment influences the reactivity of the flavin, it is desirable to remove the native prosthetic group in a, for the protein, non-destructive way. The flavin prosthetic group can then be replaced with an artificial[94, 157, 164, 208, 241, 296] or isotopically enriched analog[15, 79, 86, 113, 187, 188, 195, 196, 245, 248, 273, 309, 311]. Replacement with a flavin analog should result in the (functionally active) reconstituted holoprotein. FMN and FAD analogs can conveniently be synthesised from riboflavin, either chemically[268] or enzymatically[158], and can be isotopically enriched[202].

Artificial flavins can be used to study the relative distance between particular flavin atoms and the protein, especially when using a series of flavins which are modified at a specific position, thus varying the Vanderwaals radius of the substituent. The solvent accessibility of the dimethylbenzene- and pyrimidine part of the isoalloxazine ring system can be probed by introduction of a reactive (photoactive) group[46, 78, 162, 165, 166, 255]. Artificial flavins can also be used to covalently link the flavin to the protein[91], or to determine the nature of amino acid residues in the flavin binding site[22, 46, 164, 168]. The chemical modification of the flavin molecule can have drastic influences on its

reactivity, spectral properties or redox potentials, thus providing the researcher with a powerful tool to study the enzyme reaction mechanism and/or the flavin-apoprotein interaction[94, 217, 222].

Isotopically enriched flavins are very suitable to get a detailed view into the molecular and submolecular structure of the protein-bound flavin molecule. ^{13}C and ^{15}N NMR chemical shifts can reveal both π electron density and the presence of specific hydrogens at the carbon and nitrogen atoms investigated. ^{13}C and ^{15}N have a natural abundance of 1.1 % and 0.4 %, respectively. Therefore, the flavoprotein has to be reconstituted with ^{13}C and ^{15}N enriched flavins. This approach gives a detailed chemical view of a single atom (of the isoalloxazine ring system) of the flavin prosthetic group in the various redox states. It is possible to monitor changes during catalysis of a particular atom, i.e. to measure intermediates formed[308]. There is no need for the isotopic enrichment of phosphor atoms, since ^{31}P, the NMR active isotope, has a natural abundance of 100 %.

Medical implications

Understanding the factors that influence the interactions between flavoproteins and their cofactors is also of medical importance. In 1969, it was reported that a reduced activity of human glutathione reductase due to a mutation can lead to nonspherocytic haemolytic anaemia. The abnormal enzyme had a reduced affinity for FAD, but normal *K*m values for glutathione and NADPH. Administration of a daily flavin preparation to the patient had a therapeutic effect, and normal glutathione reductase activity was restored within a few weeks[270].

Patients suffering from chronic granulomatous disease (CDG) carry mutations in one of the protein components of NADPH-oxidase, a superoxide-producing enzyme system involved in protection against microbial infection[242]. One of the CDG patients was found to carry a mutation in the FAD-binding region of the cytosolic part of the membrane bound flavocytochrome b558[149]. Based on mutagenesis studies of structurally related flavin-dependent oxidoreductases[3], it could be inferred that this mutation, which leads to the substitution of Thr341 by Lys, does not affect the proper binding of the flavin cofactor but results in impaired hydride transfer from NADPH to FAD[149]. This contrasts the situation with another CDG patient, where it was found that a single amino acid substitution (His338Tyr) in the putative ADP-binding pocket of the flavin-binding site of flavocytochrome b558 results in severe depletion of the FAD[327].

NADH: cytochrome b5 reductase (b5R) is a membrane-bound flavoenzyme involved in the elongation and desaturation of fatty acids, cholesterol biosynthesis and cytochrome P450-mediated drug metabolism. In erythrocytes, a soluble form of b5R exists, which is functional in methemoglobin reduction. b5R deficiency manifests itself in two distinct ways. In type I methemoglobinemia, the patients only suffer from cyanosis, whereas in type II methemoglobinemia, the patients suffer in addition from

severe mental retardation and neurological impairment[147]. Biochemical data have shown that the type I mutations cause enzyme instability whereas the type II mutations result in complete enzyme deficiency or enzyme inactivation. From homology modelling with the known three-dimensional structure of porcine b5R, it could be inferred that the type I mutations in human b5R are located in non-essential regions and not involved in NADH or FAD binding. However, the other substitutions, known to cause the type II form of the disease, were found to directly affect the consensus FAD-binding site or indirectly influence NADH binding[20, 56].

MTHFR catalyses the conversion of methylenehydrofolate to methyltetrahydrofolate, the major methyl donor for the conversion of homocysteine to methionine. Mutations in human MTHFR that lead to loss of enzyme function are associated with severe hyperhomocysteinemia and aggressive development of cardiovascular disease[42]. The common C677T polymorphism in the human MTHFR gene that leads to substitution of Ala222 by Val is associated with a milder form of hypercysteinemia[87, 99]. From structural and functional analysis of recombinant MTHFR from *E. coli* it was predicted that the Ala222Val substitution decreases the affinity of the human enzyme for FAD[101]. However, *E. coli* MTHFR is a homotetramer of 33 kDa subunits, whereas the human enzyme is composed of two identical subunits of 70 kDa that contain in addition to the catalytic domain a regulatory domain for the binding of the allosteric inhibitor S-adenosylmethionine. To probe the FAD binding properties of the human enzyme, recombinant human MTHFR was expressed in a baculovirus expression system[325]. Biochemical analysis of the purified enzymes showed that the Ala222Val variant is more thermolabile and more susceptible towards monomerisation than the wild-type enzyme and confirmed that subunit dissociation is linked to the loss of FAD.

Apoptosis-inducing factor (AIF) is a flavoprotein that can stimulate a caspase-independent cell-death pathway required for early embryonic morphogenesis[52, 122]. AIF shows NADPH oxidase activity, has a glutathione reductase-like fold and is structurally most similar to the ferredoxin reductase component of *Pseudomonas* biphenyl dioxygenase[170]. To gain further insight into the redox properties of AIF, Lys176 and Glu313, located near the isoalloxazine ring of FAD, were individually changed into alanine by site-directed mutagenesis. Both AIF variants appeared to be highly active when assayed in the presence of excess FAD. However, during purification Lys176Ala and Glu313Ala easily lost the flavin cofactor, yielding the corresponding apoproteins[170]. This again demonstrates that changing a specific amino acid residue can considerably reduce the strength of flavin binding.

PREPARATION AND RECONSTITUTION OF APOFLAVOPROTEINS

A. General remarks

There are many methods for the preparation of reconstitutable apoflavoproteins. Some already exist for over 60 years, other are more recently developed. They are all based upon either weakening of flavin binding or stabilising the formed apoprotein. In theory, there are four basic principles for protein deflavination; lowering the pH, increasing the salt concentration, changing the solvent, and increasing the temperature. Because the final goal is to obtain highly reconstitutable apoprotein in high yield, the approach should be chosen with care. Prolonged exposure, for instance by dialysis, to non-native conditions, can lead to irreversible denaturation of the apoprotein formed. Therefore, the time needed for flavin removal should be relatively short.

A number of different methods for the reversible removal of flavins from flavoproteins have been described over the years. Since each flavoprotein has its own physio-chemical characteristics, several strategies have been developed to find the best non-destructive method for the protein of interest. Initial deflavination protocols were based upon precipitation, partial unfolding, or dialysis of the protein[119]. More recent techniques focus on the binding of the protein to a chromatographic support, facilitating the removal of the flavin, and reconstitution of the apoprotein[202]. When studying the properties of apo or reconstituted flavoprotein, one needs to consider the side-effects of residual flavin in the endproduct. If replacement with a flavin analog is desired, residual natural flavin might influence the catalytic properties of the reconstituted enzyme considerably. The presence of residual flavin might even be more problematic when one wants to investigate the physical and spectroscopic properties of the apoprotein itself. Below we describe several apoflavoprotein preparation procedures, starting with the conventional (and still often used) methods. Then we turn to the growing field of immobilisation-based deflavination methods in which one uses specific characteristics of the holo flavoprotein to obtain the corresponding apoprotein. Immobilisation is achieved either by non-covalent or covalent binding of the holoprotein to a chromatographic support, and is combined with one of the common methods of flavin removal, resulting in less protein deterioration.

B. Conventional methods

In theory[119], simple dialysis will never lead to pure apoflavoprotein at physiological pH, since the reassociation of the apoprotein-flavin complex competes with the physical removal of the flavin. In practice, flavin removal by dialysis is feasible, if conditions are met that lead to a strongly increased dissociation constant. In 1935, Theorell was the first who reported that flavoproteins could be reversibly resolved into their constituents apoprotein and prosthetic group. To weaken the binding of the flavin, the pH of the dialysis solution of Old Yellow Enzyme was lowered to pH 2.7, thus releasing the non-

covalently bound FMN[288]. A few years later it was reported that DAAO[315] and the yeast enzyme of Haas[103] could release their flavin in the presence of high concentrations of ammonium sulfate. Based on these initial findings a more common procedure for the preparation of apoflavoproteins was developed by combining a low pH with a high ammonium sulfate concentration[285]. Reversible precipitation methods such as acid ammonium sulfate and TCA precipitation proved to work very well at large scale, as long as the protein of interest survives the acidic conditions. TCA precipitation works good for flavodoxin[181] and the FMN-binding domain of cytochrome P450 BM3[105], which can stand these extreme conditions very well. The acid ammonium sulfate precipitation method appeared to be especially useful for 'recalcitrant' flavoenzymes such as lipoamide dehydrogenase (LipDH) from pig heart[123] and glucose oxidase from *Aspergillus niger*[285], although with both enzymes partial irreversible denaturation occurred, and variable amounts of reconstitutable apoprotein were obtained. Typical yields were in the 35 to 80 % range. Strittmatter showed that addition of a high concentration of KBr (1 M) to the acid ammonium sulfate solution weakens the electrostatic interactions between the flavoprotein and the flavin. This increased the apoprotein yield from 75 to 90 % in the case of cytochrome b5 reductase[275], and from 60 to 75 % in the case of oxynitrilase[256]. Further addition of charcoal, to adsorb the free flavin, resulted in an even more efficient removal of the prosthetic group[34, 180, 241].

Unlike flavodoxins[182], many flavoproteins are not very stable at low pH. Therefore, a procedure of apoprotein preparation was developed based on dialysis at more physiological pH. Flavoproteins can be deflavinated using a high concentration of halide anions[163, 214]. Equilibrium techniques such as dialysis and ultrafiltration[26] can be used successfully at physiological pH values for flavoenzymes, which bind the flavin prosthetic group rather weakly. D-amino acid oxidase[38, 163, 229], cyclohexanone monooxygenase[244], phenol hydroxylase[211] and 2,4-dichlorophenol hydroxylase[236] apoprotein can be obtained by simple dialysis in the presence of KBr. KCl has been used as well[292], but it is a less effective salt because it is less chaotropic.

Instead of time-consuming dialysis or ultrafiltration procedures, gel filtration methods are more efficient. Biogel P-6 and Sephadex G-25 have been applied successfully because with these resins the relatively small flavin molecules can be quantitatively separated from the apoprotein at relatively large scale[69]. Care should be taken to stabilise the thus formed apoprotein, for instance by addition of stabilising factors to the washing buffer. Glucose oxidase apoprotein, for instance, is prepared by acidification of an enzyme solution to pH 1.7, followed by separation on a Sephadex G-25 column[199, 241, 248].

Another procedure of apoprotein preparation is to partially unfold the holoprotein with guanidinium hydrochloride[32, 204] or urea[43, 306, 313], thus lowering the binding strength between protein and flavin. A disadvantage of this method is that one needs to find conditions in which the partially unfolded apoprotein is capable to refold. Circular

dichroism spectroscopy[156, 205, 303, 305] and fluorescence spectroscopy[207, 299, 305] are useful here to probe the folding behaviour of the protein of interest.

For many metalloflavoproteins, most of the conventional apoflavoprotein preparation procedures cause extensive denaturation. A likely explanation for this is that deflavination and reconstitution of flavoproteins is more difficult if the quaternary structure is more complex, i.e. when the protein contains more cofactors and/or subunits. The molybdenum-containing flavoprotein xanthine oxidase[167] can be deflavinated by dialysis at physiological pH in the presence of 2 M $CaCl_2$[136]. The FAD released is hydrolysed to FMN by the high concentration of salt present. Other methods are based upon this technique, and include $Ca(C_2H_3O_2)_2$[10] or KI[124]. An interesting metalloflavoenzyme is the molybdo iron-sulfur flavoprotein carbon monoxide dehydrogenase from *Oligotropha carboxidovorans*, which X-ray structure was elucidated recently[59, 60]. The recombinant flavoprotein was deflavinated on a gel filtration column, but the free apoflavoprotein did not bind FAD. However, when the apoflavoprotein was complexed with the iron-sulfur protein and molybdoprotein components, the flavin binding site is capable to integrate FAD[100].

Addition of a phosphodiesterase (i.e. *Naja naja atra*, snake venom) or phosphatase to dilute solutions of flavoprotein shifts the equilibrium to the apo form, because the free FAD or FMN is hydrolysed. The reaction is relatively fast, but not very useful at large scale. Moreover, extra care must be taken to remove the remaining cleavage enzyme[201].

C. Chromatographic procedures

For several flavoproteins, the conventional methods of apoprotein preparation are most satisfactory. A good example is the TCA precipitation procedure developed for flavodoxins[181]. Multidimensional NMR studies have shown that the extreme conditions for apoflavodoxin preparation do not lead to any structural perturbation in the reconstituted holoprotein, compared to the native protein[304]. However, with many flavoproteins, and especially when large amounts of apoprotein are required, (affinity-based) chromatographic procedures are the methods of choice. One of the main advantages of these methods is that on-column protein aggregation is unlikely to occur, since each molecule is at a relatively large distance from its neighbours. Particularly during partial unfolding, this helps stabilising the apoform of the protein, before flavin reconstitution.

C1. Non-covalent binding

Ion-Exchange Chromatography (IEC)

Separation in ion-exchange chromatography depends upon the reversible adsorption of charged solute molecules to immobilised ion exchange groups of opposite charge. For successful on-column flavin removal, conditions are needed where the protein interacts with the ion-exchanger (low ionic strength) and not with the flavin (low pH). This

concept was first worked out for the riboflavin-binding proteins (RfBPs) from chicken egg white[14, 75] and chicken egg yolk[173]. These proteins belong to a family of carrier proteins which are involved in vitamin delivery to the developing embryo[320]. The holo forms of the chicken RfBPs can be separated in the apo forms and free riboflavin by cation-exchange chromatography at pH 3.7. The apoprotein-riboflavin complex dissociates between pH 3.8 and 4.0[183]. At pH 3.7, the neutral highly fluorescent riboflavin is released from the column whereas the apoprotein remains tightly bound. The apoprotein can subsequently be released from the column by raising the pH and ionic strength of the elution buffer to pH 4.1 and 0.5 M NaCl.

Unlike most other flavoproteins, apo RfBP interacts strongly with a large number of flavin derivatives but not with FMN or FAD[14]. The structure of RfBP is rather unusual[192]. Besides from an N-terminal ligand-binding domain that is strongly conditioned by nine disulfide cross-links, it contains a flexible phosphorylated motif with nine phosphoserines, which is essential for vitamin uptake[186]. The isoalloxazine ring of the tightly bound riboflavin molecule is stacked in between two tryptophans, explaining the strong fluorescence quenching observed upon flavin binding.

The interaction between apo RfBP and riboflavin in the oxidised and two-electron reduced state has been addressed by reconstitution of the protein with ^{13}C- and ^{15}N-enriched riboflavin derivatives[196]. Except from giving insight into the solvent accessibility and conformational properties of the protein-bound flavin, these studies revealed that in the reduced state, the pK_a of the N1 atom of the flavin is shifted by about 0.8 pH unit to more alkaline pH values as compared to that of free reduced flavin. This shift to a pK_a value of 7.45 is quite unusual, because in many flavoproteins, the N1 atom of the reduced flavin is negatively charged[39, 200, 291].

The effective binding of riboflavin to the apoform of RfBP has also received biotechnological attention. Apo RfBP was shown to scavenge riboflavin in model beer solutions, thereby inhibiting the light-induced formation of reactive oxygen species and sunstruck off-flavour[139, 297].

Hydrophobic Interaction Chromatography (HIC)

To circumvent severe protein loss using conventional techniques, a generally applicable immobilisation procedure was developed for the large scale preparation of apo flavoproteins using hydrophobic interaction chromatography[300]. This method makes use of the fact that many flavoproteins bind to phenyl agarose at neutral pH in the presence of 1 M ammonium sulfate. After immobilisation, the flavin can then be removed by the addition of high concentrations of KBr and/or lowering the pH of the elution buffer. The HIC method has been successfully applied for a number of flavoproteins[295, 300], sometimes with slight modifications, to get optimal results[79, 143, 257].

The HIC method for preparing apoflavoproteins works very well for prokaryotic and eukaryotic disulfide reductases, including LipDH, glutathione reductase, and mercuric reductase, and is preferred over classical methods[300]. For LipDH, it was shown that the kinetics of holoenzyme reconstitution are dependent on the source of enzyme[295] and on the type of flavin[53]. Initial FAD binding to the monomeric apoprotein results in regaining of DCPIP activity and quenching of tryptophan fluorescence. Then, dimerisation occurs as reflected by the regain of lipoamide activity, strongly increased FAD fluorescence and increased hyperchroism of the visible absorption spectrum[295]. For LipDH from *A. vinelandii*, the conformational stability of the monomeric apoprotein was compared with that of the dimeric holoenzyme[299]. Unfolding of the apoenzyme in GdnHCl follows a simple two-state mechanism and is fully reversible. However, the midpoint of unfolding of the monomeric apoprotein (C_m GdnHCl = 0.75 M) is much lower than that of the dimeric holoenzyme (C_m GdnHCl = 2.4 M). GdnHCl unfolding was also used to probe the conformational stability of *A. vinelandii* lipoamide dehydrogenase in the different redox states[299]. From this and additional mutagenesis studies it was inferred that overreduction by NADH promotes subunit dissociation and that the C-terminus of the protein plays an important role in dimer stabilisation[16, 17, 176].

Sometimes the apoprotein interacts very strongly with the HIC column material and apoprotein elution or on-column reconstitution can be difficult. In such case, the apoprotein may undergo irreversible conformational damage, e.g. by local unfolding or subunit dissociation. This was observed to some extent with the apoprotein of butyryl-CoA dehydrogenase from *Megaspheara elsdenii*[300], which can not be reconstituted when bound to the column. By using high concentrations of ethylene glycol as eluent, stable apoprotein showing negligible residual activity could be isolated with 50-80 % yield. Spectral analysis of the apoprotein revealed that the coenzyme A persulfide ligand present in the native protein[322] is removed during apoprotein preparation. At pH 7.0 and 25 °C, the apoprotein is a mixture of dimers and tetramers, and reassociates to a native-like tetrameric form in the presence of FAD. The reconstitution with FAD is relatively slow, and is stimulated in the presence of CoA ligands (Figure 2.2).

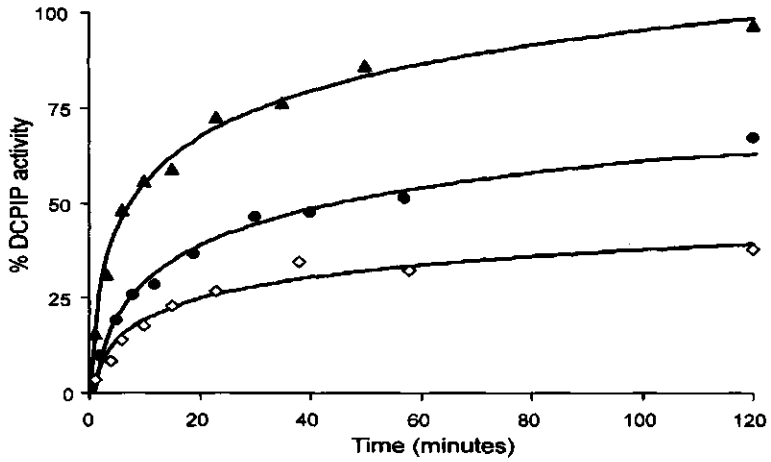


Figure 2.2 Time-dependent reactivation of apo butyryl-CoA dehydrogenase with FAD, in the absence (•) and presence (▲) of 1 mM CoASH, or in the presence of 0.5 mM acetoacetyl-CoA (◊), according to[300]. Note that the presence of 0.5 mM acetoacetyl-CoA results in about 65 % inhibition.

Binding of CoA ligands stimulates tetramerisation of the reconstituted holoenzyme and improves protein stability. This is in agreement with the crystal structure of butyryl-CoA dehydrogenase[132] which shows that the CoA ligand binds in an extended conformation near the dimer-dimer interface. Fluorescence-polarisation experiments revealed that the reconstituted protein is somewhat less stable than the native holoprotein, and that FAD dissociates more easily[300]. In Figure 2.3, a schematic representation is given of the possible mechanisms of reconstitution of apo butyryl-CoA dehydrogenase with FAD, in the absence or presence of CoA derivatives.

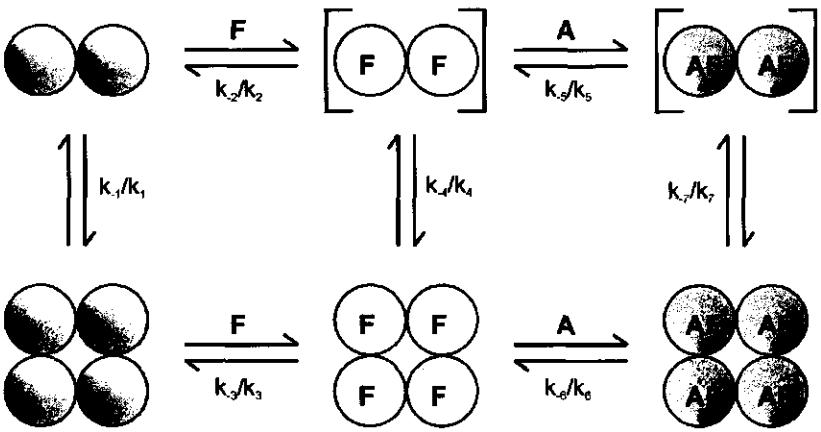


Figure 2.3 Possible mechanisms for the reconstitution of apo butyryl-CoA dehydrogenase by FAD (F), and in the presence of coenzyme A (A). For the species between brackets, no experimental evidence is available (figure according to[300]).

Another protein that was successfully deflavinated by the HIC method is L-amino acid oxidase (LAAO) from the venom of *Crotalus adamanteus*, the eastern diamondback rattlesnake[318]. This rattlesnake, which has a row of large dark diamonds with brown centres and cream borders down its back, feeds primarily on small mammals, from mice to rabbits. They locate their prey by odour, as well as by sensing the infrared waves emitted by warm-blooded prey[5, 47]. LAAO is a dimeric glycoprotein, containing one FAD per monomer, that catalyses the oxidative deamination of L-amino acids. The deflavination method for LAAO[237] is similar to the original protocol developed for LipDH. Apo-LAAO remains bound to the HIC column, while the FAD cofactor is washed away in a buffer with 1.5 M ammonium sulfate. Apo-LAAO can be reconstituted with FAD on-column. However, the FAD-reconstituted LAAO is inactive, having a perturbed conformation of the flavin binding site. In the presence of 50 % glycerol the reconstituted LAAO becomes nearly completely active. It was suggested that repulsion of glycerol from hydrophobic surfaces of the protein and the simultaneous interaction with protein polar regions initiates the restoration of the internal protein hydrophobic core. Thus, glycerol can have a restorative effect on the proposed partially unfolded equilibrium intermediates, and acts as a molecular chaperone[237]. This function of glycerol was first demonstrated for the pig kidney D-amino acid oxidase (DAAO) mutant H307L[189, 317]. The replacement of His307 with leucine perturbs the active site conformation causing a dramatic loss of enzyme activity due to the weakening of the protein-flavin interaction. The negative effect of this mutation can be eliminated in the presence of glycerol, resulting in up to 50 % activity recovery and more than 16-fold increase of flavin affinity. From this it was concluded that glycerol assists in the rearrangement of the protein towards the holoprotein conformation, as well as in reducing the solvent accessibility of the protein hydrophobic core[238].

Hydroxyapatite Chromatography (HAP)

Hydroxyapatite, $Ca(PO_4)_3OH_2$, is a crystalline form of calcium phosphate which can be used for the separation and purification of proteins. This column material often separates proteins shown to be homogeneous by other chromatographic techniques. Ceramic hydroxyapatite is a recently developed, spherical, macroporous form of hydroxyapatite, which overcomes the flow limitations of the crystalline material and allows use in large-scale columns. It has a high protein-binding capacity and similar elution characteristics as crystalline hydroxyapatite.

For preparation and reconstitution of apoflavoproteins, HAP has the advantage that high salt concentrations, which can stimulate apoprotein formation, are not necessarily a limitation. The interaction between the protein and HAP is primarily the result of non-specific electrostatic interactions between the positively charged protein amino groups and the negatively charged column material[98]. The deflavination process is influenced by the charge distribution of the protein, as well as the kind of salt that is used as eluent.

The HAP method of apoflavoprotein preparation was used for a recombinant form of the flavin-binding subunit of the flavocytochrome p-cresol methylhydroxylase (PCMH) from *Pseudomonas putida*[65] and for the His61Thr variant of vanillyl-alcohol oxidase (VAO) from *Penicillium simplicissimum*[85, 286]. In native VAO, the 8α-methyl group of the FAD isoalloxazine ring is covalently bound to the N3 atom of His422[18, 175]. Structural and functional analysis of His422 and His61 variants have shown that His61 activates the neighbouring His422 for covalent binding of FAD[84, 85] (see Figure 2.4). In the His422Ala and His61Thr variants, the covalent His422-C8α-flavin linkage is not formed but the mutant enzymes are still active and capable of binding FAD.

Figure 2.4 Flavin binding in VAO. In native VAO, His61 (red) activates the neighbouring His422 (green) for covalent binding of the FAD (yellow) cofactor. The aromatic ring of the inhibitor isoeugenol (blue) is situated parallel to the flavin isoalloxazine ring.

Binding the His61Thr variant to hydroxyapatite appeared to be a very gentle and efficient method of obtaining the VAO apoprotein. Upon washing with 200 mM phosphate buffer pH 7.2, the His61Thr protein remains tightly bound to the column, whereas the FAD is easily removed. This removal of FAD is superior to other methods. For instance, when the His61Thr holoenzyme is eluted on a Biogel P6 gel filtration column in 50 mM phosphate buffer pH 7.0 in the absence or presence of high salt,

almost no flavin is released. Another advantage of the HAP method for protein deflavination is that the His61Thr VAO apoprotein can be eluted in concentrated form by simply washing the column with 600 mM potassium phosphate pH 7.2.

While native VAO forms primarily octameric assemblies of 507 kDa, the apo His61Thr variant exists in solution as a dimer of 126 kDa (see Figure 2.5). Binding of FAD or ADP to the dimeric apoenzyme induces octamerisation. In contrast, incubation with riboflavin or FMN does not stimulate octamer assembly, suggesting that upon FAD binding, small conformational changes in the ADP-binding pocket of the dimeric VAO mutant are transmitted to the protein surface, thus promoting oligomerisation[286].

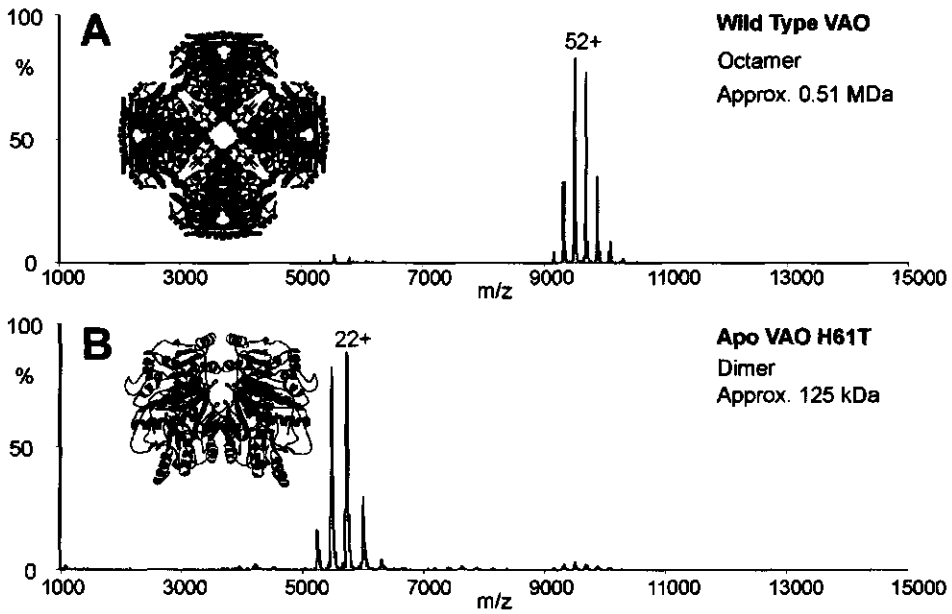


Figure 2.5 Nano-electrospray mass spectra of native VAO (**a**) and mutant apo His61Thr VAO (**b**), with a representation of the corresponding quaternary structures (figure according to[286]).

Fluorescence titration studies showed that the dissociation constant of the apoprotein-FAD complex in the His61Thr variant is about 0.2 μM. In the His422Ala variant, the FAD is much more strongly bound[84]. This might explain why the HAP method described above for the His61Thr variant does not work for the His422Ala variant of VAO. Thus, for on-column releasing the FAD cofactor in mutant His422Ala, conditions need to be found that weaken the protein-flavin interaction, while the protein remains bound to the HAP support. As noted above, an interesting approach here would be to introduce other type of salts in the washing buffer.

Dye-Affinity Chromatography (DAC)

Polyaromatic dyes can bind to proteins that use a cofactor with a nucleotide moiety[55]. Cibacron Blue has a strong affinity for the Rossmann βαβ dinucleotide binding fold[156, 243], and many flavoproteins possess such a fold, either for binding the ADP part of FAD and/or NAD(P)[321]. This feature can be used to separate apo and holo flavoproteins[289]. Such a separation was performed with Cibacron Blue Sepharose for flavocytochrome b2[230], and for DAAO[146].

Red-A agarose, another polyaromatic dye-containing material, has been used for the deflavination of *p*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens*[223]. Under low ionic strength conditions, the red dye of the column binds to the enzyme, displacing the FAD. The column is then eluted with high-ionic strength buffer containing the artificial flavin 6-azido-FAD, which binds to the protein and displaces the dye. The 6-azido-FAD cofactor can be covalently linked to the protein by irradiation. Enzyme that has not been photolabeled is separated from the covalently photolabeled enzyme by applying the reaction mixture to a Red-A column again. At low ionic strength, the nonlabeled enzyme binds to the column material, whereas the photolabeled enzyme passes directly through the column[91, 222].

C2. Covalent binding

When partial unfolding of the holoprotein is required to weaken the protein-flavin interaction, the above mentioned chromatographic methods for protein deflavination may not work properly because of the presence of high concentrations of unfolding agents. Therefore, we introduced the concept of covalent enzyme immobilisation for improving the yield and quality of the apoprotein[201].

For PHBH, the prototype of the family of single-component flavoprotein hydroxylases[68], it was shown that large amounts of highly reconstitutable apoprotein can be prepared by covalently binding the enzyme to thiol (DTNB) agarose[201]. PHBH forms a dimer in solution, and contains 5 sulfhydryl groups per monomer[203]. However, Cys116 is the only solvent exposed thiol group, accessible to N-ethylmaleimide and DTNB[301]. Using this property, it is possible to bind the enzyme covalently to the thiol-agarose column. Oxidation[298] or mutation[73] of Cys116 does not influence catalysis but prohibits binding of the protein to the DTNB column. In Figure 2.6, the PHBH dimer is shown, as well as the solvent accessibility of the Cys116 sulfur atom. After coupling to the column material, the FAD can be released from the protein with urea and KBr. The resulting column bound apoprotein is released after reaction with DTT. On-column reconstitution of the apoprotein with an artificial or isotopically labelled flavin analog is possible as well. Upon reaction with DTT, the colour of the column turns from pale yellow to dark red due to the formation of the 2-nitro-5-thiobenzoate anion. After regeneration with DTNB, which allows to determine the capacity of protein binding[151, 201], the column is slightly yellow again.

The mildness of the thiol agarose method for apoflavoprotein preparation was confirmed by the high resolution crystal structure of PHBH reconstituted with arabino-FAD[296]. Nevertheless, the thiol affinity chromatography method requires the presence of a freely available thiol group at the protein surface, making the method not generally applicable[298]. Although the introduction by mutagenesis of a surface accessible cysteine is in principle possible with every flavoprotein, knowing beforehand that this cysteine is indeed solvent exposed and will not interfere with catalysis might sometimes be a long shot.

Figure 2.6 3D-structure of PHBH. PHBH is a dimer, and contains 1 FAD cofactor (green) per monomer. In the dimer, the only solvent-exposed cysteine is Cys116. In the right panel, all amino acid residues within a distance of 10 Å to the Cys116 sulfur atom (yellow) are shown. The solvent accessibility of the sulfur atom is indicated with yellow dots, and Cys116 is drawn with Vanderwaals-radii.

Furthermore, the newly introduced cysteine should not interact intra- or intermolecularly with other cysteines. This can easily be tested with the DTNB reaction[67]. Due to its high standard oxidation-reduction potential, DTNB reacts with protein thiols by an exchange reaction to form a mixed disulfide of the protein and 1 mole 2-nitro-5-thiobenzoate per mole of protein sulfhydryl groups. The formed nitromercaptobenzoate anion has an intense yellow colour (with ε_{412}=13,600 $M^{-1}cm^{-1}$), and solvent-exposed protein thiols give complete colour development within 2 minutes[104]. For preparation of apo PHBH, the original DTNB-agarose has recently been replaced by commercially available thiopropyl sepharose[216-218].

Immobilised Metal-Affinity Chromatography (IMAC)

The introduction of recombinant DNA techniques has opened the possibility to add an affinity tag to proteins, facilitating protein purification. These affinity tags can be used to covalently bind a protein to a column. In IMAC, certain solvent exposed amino

acids (e.g. histidine, cysteine and tryptophan) form complexes with the chelated transition metals at neutral pH. Introduction of a polyhistidine tag to either the N-terminal or C-terminal end of the protein of interest facilitates a strong binding to the column bound metal ion. This strong binding is only disrupted by high concentrations of imidazole, an acidic pH or chelating agents such as EDTA. Thus, when a flavoprotein is bound to such a column, in principle rather harsh washing steps can be used to successfully remove the flavin prosthetic group. After on-column deflavination, the apoprotein or reconstituted protein can be eluted from the affinity matrix with a high concentration of imidazole.

The IMAC method of apoflavoprotein preparation was developed for the flavin-containing PAS domain of NifL, a redox-sensing protein from *A. vinelandii*[57]. The PAS domain is purified as described in chapter 3[114]. Deflavination was achieved by exploiting the available N-terminal His-tag (chapter 4,[113]). Usually, 10 mg of purified protein in 50 mM Tris-HCl, pH 7.5 containing 0.3 M NaCl, was applied onto a 1 ml nickel-nitrilotriacetic acid (Ni-NTA) superflow column, preequilibrated in the same buffer. Protein-bound FAD was efficiently removed by washing the column with 10 volumes of starting buffer, containing 2 M KBr and 2 M urea, resulting in a column-bound apo-form of the protein. The apoprotein could be removed from the column with 250 mM imidazole, but precipitated within 12 hours after release from the column. Therefore, on-column reconstitution was achieved by circulating a 10 mM solution of 2,4a-$^{13}C_2$-FAD or 2,4a-$^{13}C_2$-FMN overnight (4 °C) at 0.2 ml/min in 50 mM Tris-HCl, pH 7.5 containing 0.3 M NaCl (see Figure 4.1).

The reconstituted PAS domain containing a new flavin cofactor was eluted from the column with 2 column volumes 50 mM Tris-HCl buffer pH 7.5, containing 0.3 M NaCl and 250 mM imidazole. Reconstitution of the flavoprotein is 100 % efficient, as shown by NMR and UV-VIS absorption spectroscopy (see Figure 4.2 and Figure 4.4).

In the case of NifL, the IMAC method can best be used to prepare reconstituted protein instead of apoprotein, due to the instability of the apoprotein formed. Due to the immobilisation of the apoprotein during the process of de- and reflavination, no severe protein loss is observed. The reconstituted protein, containing isotopically labelled FAD or FMN, is used to study the PAS core, the direct environment of the flavin prosthetic group[113]. Although the NifL PAS domain normally binds FAD, the His-tag reconstitution method unambiguously showed that this protein also tightly interacts with FMN. ^{13}C and ^{31}P NMR experiments confirmed that there is no significant difference in the active site between the FAD- or the FMN- reconstituted protein in either the reduced or oxidised state (see chapter 4;[112]).

The use of IMAC for the preparation and reconstitution of His-tagged apo flavoproteins by immobilisation on a Ni-NTA column, might become a general method since His-tag incorporation in recombinant proteins is every-day work.

CONCLUSIONS AND FUTURE PERSPECTIVES

Since the pioneering work of Theorell[288], many methods have been developed for the (large scale) preparation and reconstitution of apoflavoproteins. Conventional precipitation methods, such as the acid ammonium sulfate procedure, are rapid but the yield and reconstitutability of apoprotein may vary dramatically. The more recently developed chromatographic procedures have the advantage that the apoprotein is stabilised by immobilisation, and that large amounts of apo- or reconstituted flavoproteins can be obtained.

His-tagged flavoproteins can be purified, deflavinated and reconstituted on the same column. Therefore, the use of IMAC for flavoprotein deflavination and reconstitution should be further exploited. An interesting possibility to facilitate flavin release with such a column is to deflavinate the protein in its reduced state. In the reduced form, flavin binding sometimes is less tight than in the oxidised form[160, 271]. The introduction of a solvent exposed cysteine residue is another engineering strategy that allows the covalent binding of a flavoprotein to a solid support. Although not always straightforward, this method produces high quality apoflavoproteins in very good yield[201].

Site-directed mutagenesis can also be used to change the strength of the flavin-apoprotein interaction. With such approach, it is important that the amino acid replacements (or deletions) do not result in dramatic changes in protein stability and/or enzyme catalysis. The feasibility of this approach for the preparation of apoflavoproteins is demonstrated by the above mentioned studies on AIF, LipDH, NQR, MTHFR, and VAO and is further illustrated by the following example: By rational design, a surface loop of 14 amino acid residues was removed from *Rhodotorula gracilis* DAAO[228]. This shifted the protein from the dimeric into the monomeric state. In the monomeric state, the FAD interacts more weakly with the protein, because the cofactor is now solvent accessible. The apoprotein of the DAAO mutant was then obtained by dialysis in the presence of a chaotropic agent[38].

Sometimes, site-directed mutagenesis is needed to improve the flavin-apoprotein interaction. With L-aspartate oxidase, a FAD-dependent enzyme involved in the microbial biosynthesis of NAD^+, the protein could only be crystallised in the apoform[177]. All attempts to crystallise the wild-type enzyme in the holo form were unsuccessful. However, the Arg386Leu mutant, in which an active site arginine is replaced, turned out to be amenable to crystallisation and structure elucidation in the active FAD-bound state[28].

In the near future, other recombinant-based methods such as the construction of fusion proteins will undoubtedly emerge for flavoprotein deflavination and reconstitution. These procedures will be also valuable for the preparation of apo forms of other cofactor-containing proteins, specially when the apoprotein is relatively unstable in solution.

CHAPTER 3

A Novel Purification Method for Histidine-tagged Proteins containing a Thrombin Cleavage Site

This chapter is a modified version of:

Marco H. Hefti, Caroline J. G. Van Vugt - Van der Toorn, Ray Dixon and Jacques Vervoort
Analytical Biochemistry **295**, 180–185 (2001)

ABSTRACT

A general procedure for the purification of histidine-tagged proteins has been developed using immobilised metal-ion affinity chromatography. This two-step purification method can be used for proteins containing a hexahistidine tag and a thrombin cleavage site, yielding high amounts of purified protein. The advantage of this method is that thrombin is used instead of imidazole, in the final purification step. Imidazole can influence NMR experiments, competition studies or crystallographic trials, and the presence of imidazole often results in protein aggregates. Removal of the His-tag results in a form of the protein of interest in which no additional tags are present, resembling the native form of the protein, with only 3 additional amino acids at the N-terminal side. Our method is compared with a more conventional method for the purification of the *Azotobacter vinelandii* NifL PAS domain, overexpressed in *E. coli*. It also proves to be successful for three different His-tagged proteins, the *Klebsiella pneumoniae* NtrC protein, and the *A. vinelandii* NifA and NifL proteins, and therefore it is a general method for the purification of His-tagged proteins.

INTRODUCTION

In 1975, Porath and co-workers[232] first described immobilised metal-ion affinity chromatography (IMAC). Metal-ions bound to an agarose matrix have the tendency to form complexes with proteins that have surface-exposed imidazole and thiol groups such as histidines, tryptophans or cysteines. This observation made it possible to easily purify proteins from complex mixtures by attaching a polyhistidine-tag to the N- or C- terminus of proteins (for an overview read[64, 233, 278]). In principle a single step IMAC purification procedure is used, in which the polyhistidine-tagged protein is washed from the matrix with imidazole[7, 51, 141, 150, 252] and the imidazole is subsequently removed by dialysis. In some cases, however, even extensive dialysis cannot remove all imidazole bound to the protein. For certain experiments, such as NMR measurements, crystallisation trials or interaction studies, it is necessary to fully remove imidazole. Moreover, the presence of imidazole can result in aggregation of the protein[259].

The introduction of an endopeptidase cleavage site adjacent to the histidine-tag facilitates removal of the tag. A thrombin cleavage site is frequently used[263, 272, 312] since thrombin recognises a highly specific cleavage sequence[41].

Here we present a general applicable method, in which the protein of interest is cleaved from the His-tag directly on the column and no extensive dialysis is necessary. This two-step purification method for His-tagged proteins results in pure protein in the absence of the tag, and without requirement for imidazole in the final elution step.

EXPERIMENTAL PROCEDURES

Materials

All chemicals used were commercially available and at least of reagent grade. Unless stated otherwise double distilled deionised water was used (at least 18 megaOhm/cm) using the NANOpure system from SYBRON Barnstead, Boston, USA. For chromatography, the following buffers were made and filter-sterilised (0.22 µm GVWP Millipore filters): Buffer A: 50 mM Tris - 300 mM NaCl - 20 mM imidazole (Merck), pH 7.5; buffer B: 50 mM Tris - 300 mM NaCl - 250 mM imidazole, pH 7.5; buffer C: 50 mM Tris - 300 mM NaCl, pH 7.5; buffer D: 50 mM Tris - 300 mM NaCl, 2.5 mM $CaCl_2$, pH 7.5.

Protein expression

Overproduction of the *Klebsiella pneumoniae* NtrC protein and the *Azotobacter vinelandii* NifA, NifL, and the NifL PAS domain proteins was achieved using the T7 expression system (for an overview on the different proteins and their function read[57]). DNA fragments derived either by PCR amplification or restriction digestion were cloned into the expression vector pET28b (Novagen) and expressed in the BL21 (DE3) pLysS *E. coli* strain. Plasmid pMB737 contains the full-length *nif*A gene derived from pDB737[6], plasmid pDW78 the full-length *ntr*C gene from pNH31[58], and plasmid pTJ45 the full-length *nifL* gene derived from pPW53[6]. Plasmid pKH416, which overexpresses the amino-terminal PAS domain of NifL from amino acid M1 to H140, was derived by PCR amplification of pPW53. All plasmids contained an N-terminal His-tag with an adjacent thrombin cleavage site and in all cases coding sequences were flanked by an *Nde*I site at the start codon and a *Bam*HI site following the stop codon.

E. coli cells were first grown on LB plates supplemented with 50 µg/ml kanamycin (Sigma). Single colonies were transferred to 5 ml prewarmed LB medium containing 50 µg/ml kanamycin. After overnight growth this culture was transferred to a 2-litre flask containing 500 ml prewarmed medium. Cells were grown with vigorous shaking (300 rpm) to a density of A_{600} = 0.4 (corresponding to approx. $0.5*10^9$ *E. coli* cells per ml) before protein expression was induced with 1 mM IPTG (Gibco). Cells were harvested by centrifugation at 4 °C for 20 minutes at 10,000g after reaching a cell density of A_{600} = 0.6-0.8. The cells were resuspended in 40 ml cold buffer A, and passed 3 times through a French Press. This suspension was centrifuged for 20 minutes at 20,000g. The cell pellet was resuspended with buffer A and centrifuged again, reducing the protein loss in this step. The supernatant was passed through a 0.45 µm non-sterile FP 30 filter (Schleicher & Schuell) to avoid clogging of protein aggregates or cell debris on the Ni-NTA column.

Purification Procedure

All experiments were performed in a cold room (4 °C). The Ni-NTA Superflow (Qiagen, Valencia, USA) column material is comprised of Ni^{2+}-nitrilotriacetic acid

coupled to Superflow resin (Sterogene Bioseparations, Inc.). The TALON (Clontech, CA, USA) column material consists of a special polydentate metal chelator, which tightly retains cobalt, and uses Sepharose CL-6B (Pharmacia-LKB Biotechnology, Uppsala, Sweden) as a durable matrix. The Ni-NTA column can be regenerated, while the TALON column cannot, according to the manufacturers protocols. For this reason the Ni-NTA column was always used as the first column, because after several purifications a brownish smear was observed on top of the column. Though both IMAC columns use the same principle, there may be differences in their specific affinity. IMAC columns were prepared by pouring 25 ml resin into a XK-16/20 (Ni-NTA) or a XK-26/20 (TALON) glass tube column (Pharmacia) containing water, and settling of the gel bed was accelerated by connecting the column outlet to a pump running with a flow speed of 0.5 ml/min. The column was connected to a Gradifrac system with an in-line A_{280}-detector (Pharmacia) connected to a recorder, and equilibrated with at least 10 column volumes of buffer A. At all times the salt concentration was kept at 300 mM to avoid non-specific hydrophobic or ionic interactions.

A) Imidazole-based purification method

This method was used to purify the *Azotobacter vinelandii* NifL PAS domain protein. After French Press disruption the protein solution was loaded onto the Ni-NTA column at a flow-speed of 1 ml/min. Weakly bound proteins were washed from the resin with buffer A. The recombinant histidine-tagged protein was eluted by increasing the imidazole concentration from 20 mM (buffer A) to 250 mM (buffer B) in 1 column volume. Eluted fractions were analysed by SDS-PAGE and fractions containing the tagged protein were pooled, diluted to 20 mM imidazole with buffer C and further purified using the TALON column. Then the protein was stripped from the TALON column using buffer B, resulting in a pure protein solution, containing a high concentration of imidazole. To fully remove imidazole it is necessary to extensively dialyse this protein solution. However, precipitation was observed upon dialysis (10 kDa membranes), independent of buffer conditions which resulted in loss of protein (Table 3.1). Even if the protein solution was first diluted to a final imidazole concentration of about 50 mM precipitation still occurred. Although changing buffer conditions using the Amicon system (with YM-10 filter) is faster than dialysis, the protein also aggregated under these conditions (see Table 3.1). Removal of imidazole by using a 60 cm length desalting column (Bio-Gel P-6 Desalting Gel from Biorad) resulted in loss of protein (Table 3.1). Protein recovered after the second Amicon step (2.7 mg) was freeze dried and subsequently dissolved in 500 µl 20 % D_2O solution with a final buffer concentration of 20 mM potassium phosphate pH 7.5; 50 mM KCl. NMR-spectroscopy showed that imidazole was still present in this protein solution, with an estimated concentration of about 50 µM resulting in two large proton peaks at 7.6 and 7.0 ppm.

Table 3.1 Overview of yields per purification step for the NifL PAS domain.

Purification step	Imidazole based purification [1]		Thrombin based purification [2]
	Volume [3] (ml)	Protein amount [3] (mg)	Protein amount [3] (mg)
Ni-NTA	750	16.1	15.2
TALON	32	15.0	13.1
Dialysis	61	11.0	
Amicon	9.8	8.1	
Biogel	20.3	3.8	
Amicon	5.1	2.7	
Pure protein		2.7	13.1

Purification started with approximately one litre *E. coli* culture; cells were broken using a French Press.

[1] The imidazole based purification method resulted in histidine-tagged protein.

[2] The thrombin based purification method resulted in protein with no histidine-tag present.

[3] Volumes and protein amounts were determined after each purification step.

B) Thrombin-based purification method

This method was developed because the imidazole-based purification method resulted in loss of protein due to the efforts necessary to fully remove imidazole. It uses on-column cleavage of the His-tag with thrombin. After French Press disruption (see Table 3.2) the protein solution was loaded on the Ni-NTA column. Following elution with imidazole all fractions containing tagged protein were pooled, diluted with buffer C to a final concentration of 20 mM imidazole and loaded on the TALON column (at a flow speed of 0.5 ml/min this took approximately 8 hours). After stabilisation of the A_{280} baseline the column was washed with at least 10 column volumes buffer D to remove imidazole. A catalytic amount of 100 NIH units thrombin (Sigma) was dissolved in 30 ml buffer D, and this solution was circulated over the column at low speed (0.1 ml/min with a Pharmacia P-50 pump) by connecting the column outlet to the inlet. A Pharmacia UV-1 detector was used to monitor the A_{280} baseline and thus the cleavage process. After two to four days, dependent on the total amount of protein loaded on the column, this baseline stabilised and the cleaved protein was collected by washing the column with buffer D. After dialysing three times against 5 litre 5 mM ammonium bicarbonate, pH 7.5 the amount of pure protein was determined (Table 3.2). To fully remove the His-tag still bound to the column at least 10 column volumes buffer B was used to clean the column before further use.

SDS-PAGE

Electrophoresis was performed using tricine as a trailing ion, according to the modified method[250] of Laemmli[140]. The discontinuous gels consisted of a 4 % T, 3 % C stacking gel and a 16.5 % T, 3 % C separating gel for the NifL PAS domain or a 10 % T, 3 % C separating gel for the NifL, NifA, and NtrC proteins which were run on a vertical electrophoresis unit (SE 250 Mighty Small II, Pharmacia). T denotes the total percentage concentration of both monomers (acrylamide and bisacrylamide), and C the percentage concentration of the crosslinker relative to the total concentration T. Protein

bands were detected by staining with Serva Blue G. The low molecular weight markers were from Pharmacia.

Table 3.2 Total protein yields in the different purification steps.

Protein	Cell wet weight (grams)	French Press (grams)	Ni-NTA (mg)	Talon (mg)
NifA	9.04	1.28	54.2	47.3
NifL PAS	10.06	1.37	60.7	52.5
NifL	12.72	1.63	39.2	28.5
NtrC	13.03	1.57	48.4	39.6

MALDI mass-spectrometric analysis

Matrix-assisted laser desorption/ionisation time of flight (MALDI-TOF) mass analysis was carried out on a PerSeptive Biosystems Voyager DE RP Biospectrometry Workstation. Matrix solutions were prepared by dissolving in a 1 ml mixture of acetonitrile:water 20 mg of either sinapinic acid (Aldrich) (1:2) or α-cyano-4-hydroxycinnamic acid (Aldrich) (1:1) containing 1 % TFA. For analysis 1 µl matrix solution was mixed with 1 µl protein solution and applied to a 100-well gold-coated plate. Dilution series (1:10 to 1:1000) of protein in matrix solution were pipetted on the plate as well. Samples were allowed to dry in a gentle air stream for at least 1 hour. For measurement of protein masses above 5 kDa, sinapinic acid was used as matrix, and measurements were performed in linear mode. α-Cyano-4-hydroxycinnamic acid was used as a matrix for masses ranging from 500 Da to approximately 5 kDa and the detector was set in reflective mode. The mass spectrometer was operated in the positive-ion mode with a delayed extraction time of 500 ns. Ions were accelerated to an energy of 25 kV before entering the TOF mass spectrometer. To obtain optimised spectra the minimum laser power possible was used, and 30-100 spectra were accumulated for each run. The mass spectrometer was calibrated externally with a mixture of 8 proteins ranging from 1 kDa to 66 kDa, prepared as described above.

Protein Quantification

Both the NifL protein and its PAS domain contain a non-covalently bound FAD cofactor that remains tightly bound during purification. NifL is a tetramer in solution with an average FAD:protein tetramer ratio of 3:1[116]. The NifL PAS domain also is a tetramer in solution (according to analytical gel filtration runs with a Superdex 200 HR-10/30 column connected to an ÄKTAexplorer; data not shown) with an average FAD:protein tetramer ratio of 2:1. This was measured by comparing protein concentrations determined at 450 nm using an ε_{FAD}=11300 $M^{-1}cm^{-1}$[319] and determined as described by Bradford[31] using Bradford Reagent (Sigma) with bovine serum Albumin (Boehringer) as a standard. NtrC and NifA do not contain FAD and therefore the protein concentration was determined using the Bradford method.

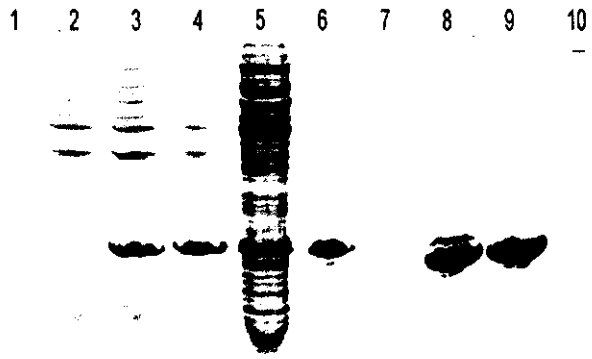
RESULTS AND DISCUSSION

Large-scale Production and Purification of the Polyhistidine-tagged proteins

Small-scale experiments indicated a higher protein-yield if the cells were first grown to a density of A_{600} = 0.4 before IPTG induction was started. After French Press disruption approximately 80 % of the His-tagged proteins were in the soluble fraction, as estimated by SDS-PAGE. Although protein solutions were carefully filtered before loading on the Ni-NTA IMAC column, a brownish band was observed on top of this column after several runs. This band could be removed by washing and regenerating the column material according to the manufacturer's protocol. Together with the His-tagged protein, typically 3 to 4 *E. coli* proteins were co-purified in small amounts (Figure 3.1, lane 6). Although binding of His-tagged protein is based on the same principle, the use of two different IMAC columns eliminates copurification, which occurs if one uses just one column[259]. Because dialysis of fractions containing the His-tagged protein often resulted in precipitation of the protein due to the high imidazole concentration, this step was omitted. We decided therefore to reduce the imidazole concentration of all protein-containing fractions by adding buffer C to a final concentration of 20 mM. This protein solution, containing 4 to 5 different proteins, was then loaded overnight on the TALON column. At a loading speed of 0.5 ml/min all His-tagged proteins bound to the IMAC columns, as no His-tagged protein was detected in the flow-through (Figure 3.1, lane 7). The column was then extensively washed with buffer D to remove imidazole. The masses of the thrombin-cleaved proteins were checked using the mass spectrometer (Table 3.3). After thrombin was added to start cleavage, the A_{280} base-line increased indicating that protein was cleaved from the histidine-tag. This resulted in a 100 % pure protein with no imidazole present (Table 3.2 and Figure 3.1, lane 9). The masses of the proteins were compared with the theoretical thrombin-cleaved masses, and cleavage always occurred at the thrombin cleavage site (Table 3.3).

Protein Quantification

Imidazole has a dramatic influence on the accuracy of protein concentration determination using the Bradford method. We observed a two-fold increase in absorbance when comparing dilution series of bovine serum albumin in the absence and presence of 500 mM imidazole. However, in the absence of imidazole (as determined by NMR spectroscopy), the amount of purified protein can be measured using the Bradford method. This method gave equivalent values to those found when using the ε_{FAD} at 450 nm or when using the estimated ε_{280}[95].



Lane 1- molecular weight marker.
2- E. coli cells prior to IPTG induction.
3- cells 2 hours after IPTG addition.
4- cells 3 hours after IPTG addition.
5- supernatant from French Press disruption.
6- after Ni-NTA column.
7- TALON column flow-through during loading.
8- TALON column eluant after thrombin digestion and cleaning the column with buffer B, about 10 times concentrated.
9- NifL PAS domain released from the TALON column by on-column cleavage.
10- molecular weight marker.

Figure 3.1 SDS-PAGE of the different steps in the on-column cleavage purification method using the NifL PAS domain.

In Table 3.2 a purification scheme is shown for all 4 proteins. Following French Press disruption protein was determined using the Bradford method, and after the Ni-NTA IMAC column purification using the ε_{FAD} (NifL, NifL PAS) or the Bradford method (NifA, NtrC). The A_{280} absorbance was measured continuously during the on-column enzymatic cleavage of these proteins using the TALON IMAC column. After 2 to 4 days no further increase in absorbance was observed, resulting in 100 % pure protein (Figure 3.1, lane 9). Addition of another 100 NIH units thrombin did not result in an increase of the absorbance, indicating that all proteins with a solvent exposed cleavage site were cleaved from the column.

Table 3.3 Comparison of the expected and measured protein masses.

Protein	Calculated mass [1]	Measured mass [2]
NifA	58404.8	58441
NifL PAS	15898.0	15897.5
NifL	58109.5	58102
NtrC	52626.1	52573

[1] The displayed values are the average $[M+H]^+$ masses calculated with the EXPASY Proteomics tools (http://www.expasy.ch/tools/peptide-mass.html) after cleavage at the predicted thrombin cleavage site.

[2] The masses determined with MALDI-TOF MS are of the thrombin-cleaved protein domains, and represent the average $[M+H]^+$ mass.

CONCLUSION

The high specificity of IMAC towards His-tagged proteins and the high specificity of thrombin for its cleavage site can be combined in a purification method for His-tagged proteins. Previous methods for histidine-tag removal have required purification and extensive dialysis[224, 261]. The method described here works very efficiently with the 4 His-tagged proteins tested in this study. The high selectivity of IMAC column material towards histidine-tags, combined with on-column removal of the tag with thrombin results in a two-step purification with high overall yield. If the TALON column is not washed extensively with buffer D before starting on-column thrombin cleavage, high concentrations of imidazole remain bound to the protein after repeated dialysis, as detected by NMR spectrometry. The presence of imidazole can be problematic for NMR measurements, or for long time storage[259]. In the final purification step, about 15 % protein is lost during the cleavage procedure. This is partially due to the removal of copurified *E. coli* proteins, but, after cleaning the TALON column with buffer B both the His-tag and non-cleaved protein are visible on SDS-PAGE (Figure 3.1, lane 8). This suggests that thrombin cleavage is not 100 % complete and not all protein is released from the column. The protein that is released from the column, however, is 100 % His-tag free (lane 9).

A problem that could occur is clogging of the TALON column due to circulating a highly concentrated protein solution. This can be avoided by removing the cleaved protein after 2 days, and starting on-column cleavage again by addition of a new catalytic amount of thrombin. Because the on-column thrombin cleavage takes approximately 2 to 4 days at 4 °C, this method might not be applicable for less stable proteins, although higher concentrations of thrombin can be used to accelerate the cleavage process. Thrombin is expected to be several fold more active at 20 °C than at 4 °C, so we believe that the whole cleavage experiment could be performed in 12 - 24 hours. In our case 4 °C was used, because the NifL protein is known to be less stable at room temperatures. In our protocol, thrombin is not removed from the purified protein. In our experience, no further protein cleavage is observed on SDS-PAGE even after 3 months at 4 °C (0.1 % NaN_3). For NMR measurements this very low concentration of thrombin is not a problem, as long as no further cleavage occurs, but also because it is not visible in the NMR spectra. If necessary, further purification can be achieved by adding agarose beads coated with anti-thrombin (Antithrombin III from Sigma) to the final protein solution. Alternatively, the thrombin cleavage capture kit (Novagen, Calbiochem) can be used. The thrombin can then be removed by centrifugation.

CHAPTER 4

A Histidine-tag based Immobilisation method for the Preparation and Reconstitution of Apoflavoproteins

Part of this chapter has been published:

Marco H. Hefti, Fin J. Milder, Sjef Boeren, Jacques Vervoort, and Willem J. H. van Berkel
Biochimica et Biophysica Acta **1619**, 139-143 (2003)
and
Flavins and Flavoproteins **10**, 969-973 (2002)

ABSTRACT

The NifL PAS domain from *Azotobacter vinelandii* is a flavoprotein with FAD as the prosthetic group. Here we describe a novel immobilisation procedure for the large-scale preparation of apo NifL PAS domain and its efficient reconstitution with either 2,4a-^{13}C-FAD or 2,4a-^{13}C-FMN. In this procedure, the His-tagged holoprotein is bound to an immobilised metal affinity column and the flavin is released by washing the column with buffer containing 2 M KBr and 2 M urea. The apoprotein is reconstituted on-column with the (artificial) flavin cofactor, and then eluted with buffer containing 250 mM imidazole. Alternatively, the immobilised apoprotein can be released from the column matrix before reconstitution. The His-tag based immobilisation method of preparing reconstituted (or apo) NifL PAS domain protein has the advantage that it combines a protein affinity chromatography technique with limited protein loss, resulting in a high protein yield with extremely efficient flavin reconstitution. This on-column reconstitution method can also be used in cases where the apoprotein is unstable. Therefore, it may develop as a universal method for replacement of flavin or other cofactors.

INTRODUCTION

The reversible dissociation of flavoproteins into its constituents, apoprotein and flavin prosthetic group, has received much attention[119, 143, 202]. Generation of a stable soluble form of the apoprotein is of great interest as it allows the reconstitution of the holoprotein with either artificial[94, 161, 326], enzymatically modified[296] or isotopically enriched flavin analogues[79, 201, 245]. Conventional methods of apoprotein preparation include acid ammonium sulfate precipitation[275], dialysis in the presence of chaotropic salts[163] or treatment with unfolding agents[32, 197]. These methods can lead to significant irreversible protein denaturation and are often not suited for large-scale applications. Therefore, we introduced the concept of reversible protein immobilisation for improving the yield and reconstitutability of the apoprotein[201].

For *p*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens* it was shown that large amounts of highly reconstitutable apoprotein can be prepared by covalently binding the enzyme to thiol agarose[201]. The suitability of this mild procedure was confirmed by the high resolution crystal structure of PHBH reconstituted with arabino-FAD[296]. However, the thiol affinity chromatography method requires the presence of a freely available thiol group at the protein surface, making the method not generally applicable[298]. Therefore, another immobilisation procedure was developed for the large-scale preparation of apo flavoproteins using hydrophobic interaction chromatography (HIC)[300].

This method makes use of the fact that many flavoproteins bind to phenyl agarose at neutral pH in the presence of 1 M ammonium sulfate. After immobilisation, the flavin

can be released by the addition of a high concentration of KBr and/or lowering the pH of the elution buffer. The HIC method has been successfully applied for a number of flavoproteins[237, 257, 295, 300], sometimes with slight modifications to get optimal results[79, 143]. A disadvantage of the HIC method is that with certain flavoproteins, the apoprotein is more hydrophobic than the holoprotein and is not easily removed from the column. In such case, the apoprotein may undergo conformational damage (e.g. subunit dissociation), leading to loss of enzyme activity[237, 300]. Because not all flavoproteins do bind to a hydrophobic matrix, we sought for an alternative method for the large-scale preparation of apo flavoproteins. Here we describe a method, using immobilised metal affinity chromatography (IMAC). This IMAC method was developed for the FAD-containing PAS domain of NifL.

PAS domains are important signaling modules that monitor changes in light, redox potential, oxygen, small ligands, and overall energy level of a cell[332]. NifL, a redox-sensing protein from *A. vinelandii*[57], is a member of this family. The N-terminal PAS domain serves as the flavin redox-sensing domain distinct from the C-terminal nitrogen-responsive and nucleotide-binding domain. The NifL PAS domain (140 amino acids), to which an N-terminal 6xHis-tag and a thrombin cleavage site have been added[265], has been overexpressed in *E. coli*. In contrast to previous work[114], and due to an improved purification method, the PAS domain purifies as a tetramer with 4 FAD molecules per tetramer. In the rhombohedral (R32) crystal form this domain contains 1 monomer per asymmetric unit[111]. Recently, a new structure with a PAS domain fold has been released[49]. This domain from the phototropin module of *Adiantum* phy3 contains FMN as the cofactor. Because phy3 and the NifL PAS domain share an identical fold, it was of interest to study by ^{13}C NMR the interaction of the NifL PAS domain with FMN, as well as with its native cofactor FAD.

MATERIALS AND METHODS

Protein purification

The NifL PAS domain, including a 6xHis-tag and thrombin cleavage site at the N-terminus, was obtained by overexpression of plasmid pKH-416 in the *E. coli* host strain BL21 (DE3) pLysS. The expression vector was pET28b (Novagen). The PAS domain was purified as described in[114].

Synthesis of 2,4a-^{13}C-FMN and 2,4a-^{13}C-FAD

Barbituric acid was synthesised[106] from ^{13}C-labeled urea and diethyl malonate-1,3-$^{13}C_2$ resulting in a yield of 89 %. This barbituric acid was refluxed for 5 hours with 3,4-dimethyl-(6-phenylazo)-ribitylaniline, resulting in riboflavin[290, 310].

For synthesis of 5'-FMN, the riboflavin was phosphorylated with phosphorous oxychloride[258]. The resulting crude FMN was purified on a 310 x 25 mm Lichroprep

RP18 column with a flow rate of 6 ml/min in 10 % methanol, 90 % 0.1 M NH_4HCO_3 in water[213]. After freeze-drying, salt was removed by redissolving the FMN in water and reinjecting it on the same column. For this step, the eluent used is 50 % methanol in water. The pure 5'-FMN obtained was freeze-dried and stored at -20 °C before further use.

FAD was prepared enzymatically from riboflavin with FAD synthetase[158]. The resulting crude FAD was purified on a 20 x 2 cm Lichroprep C18 column at a flow rate of 4 ml/min with a gradient from 0 to 25 % methanol in water. The resulting 2,4a-^{13}C-FAD was freeze-dried and stored at -20 °C before further use.

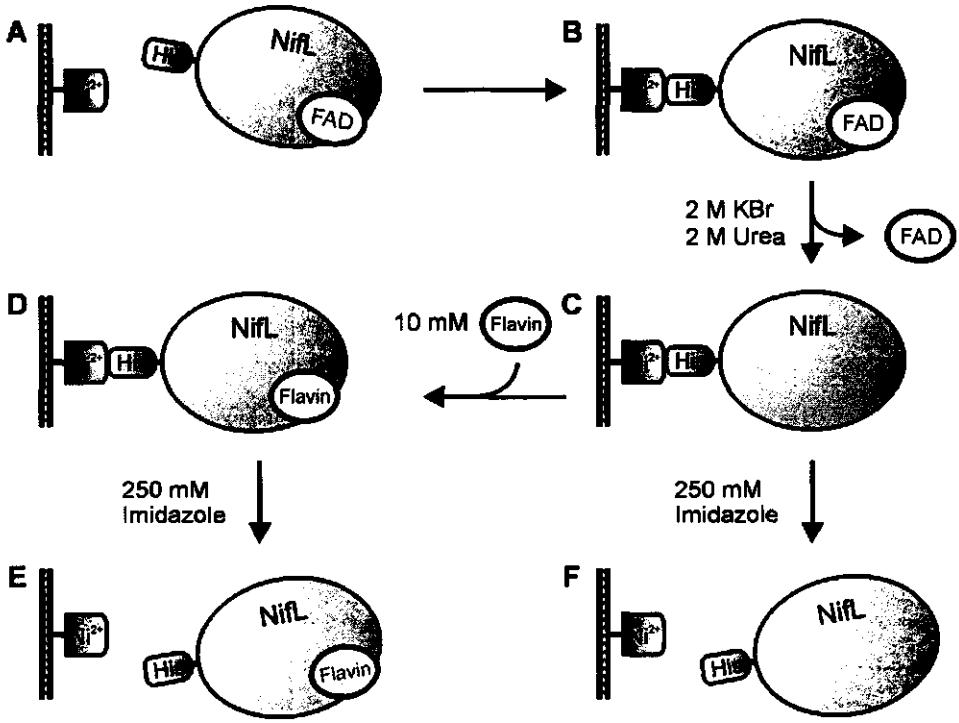


Figure 4.1 Schematic representation of the preparation and reconstitution of His-tagged apo flavoproteins by immobilisation on a Ni-NTA column.

Immobilized metal affinity chromatography

A schematic representation of the on-column reconstitution of flavoproteins is presented in Figure 4.1. 15 mg of purified PAS domain in 50 mM Tris-HCl buffer pH 7.5 containing 0.3 M NaCl, was applied onto a 1 ml preequilibrated Ni-NTA superflow (Qiagen) column (Figure 4.1A). After loading the protein solution at a flow rate of 0.2 ml/min, the column outlet was connected with the pump inlet, creating a loop to ensure maximal binding of protein to the column (Figure 4.1B). After overnight circulation (0.2 ml/min) at 4 °C in the dark, the loop was disconnected. Protein-bound FAD was then

removed by washing the column with 10 volumes of starting buffer, containing 2M KBr and 2 M urea, resulting in a column-bound apo-form of the protein (Figure 4.1C). The column was washed with 10 column volumes of a 50 mM Tris-HCl buffer pH 7.5 containing 0.3 M NaCl. On-column reconstitution of the apo form was done by circulating a 10 mM solution of 2,4a-^{13}C-FAD or 2,4a-^{13}C-FMN overnight (4 °C) at 0.2 ml/min (Figure 4.1D) in 50 mM Tris-HCl buffer pH 7.5 containing 0.3 M NaCl. Excess cofactor was then removed by washing with 50 mM Tris-HCl pH 7.5, containing 0.3 M NaCl. The reconstituted apo PAS domain containing a new flavin cofactor was eluted from the column with 2 column volumes 50 mM Tris-HCl buffer pH 7.5, containing 0.3 M NaCl and 0.25 M imidazole (Figure 4.1E) and concentrated to 10 mg/ml by ultrafiltration (Millipore Ultrafree 5K centrifugal filter).

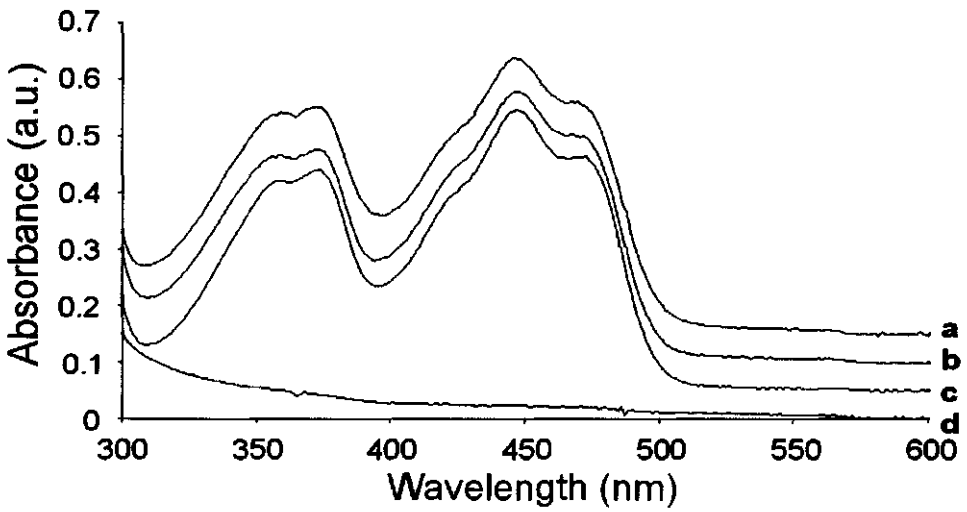


Figure 4.2 Absorption spectra of the NifL PAS domain. The untreated holo protein (line **a**) shows similar flavin absorption characteristics as protein on-column reconstituted with ^{13}C-FAD (**b**) and ^{13}C-FMN (**c**). Part of the apoprotein (**d**) formed aggregates immediately after removal from the IMAC column, but the remaining protein-containing solution clearly does not contain FAD or FMN. For clarity, the absorption value at 600 nm was set to zero for (**d**) and shifted upwards 0.05 a.u. for (**c**), 0.1 a.u. for (**b**), and 0.15 a.u. for (**a**), respectively.

For preparing the apo form of the PAS domain the apoprotein was eluted from the column after flavin removal, with 2 volumes of 50 mM Tris-HCl buffer pH 7.5, containing 0.3 M NaCl and 0.25 M imidazole (Figure 4.1F), and concentrated to approximately 3 mg/ml by ultrafiltration (Millipore Ultrafree 5K centrifugal filter). The FAD content of the apoprotein was < 0.1 % as evidenced by absorption spectral analysis (Figure 4.2).

Analytical methods

Protein concentrations were determined using the Bradford assay (Sigma). Calibration was done with known BSA concentrations (Pierce) using disposable cuvettes. Absorbance spectra were recorded at a Hewlett Packard Chemstation 8433 UV-visible diode-array spectrometer using Hellma 1cm light path quarts cuvettes.

Circular dichroism spectra (195-260 nm) were recorded with a J-715 Jasco Spectropolarimeter with Starna 1 mm light path quarts cuvettes. All CD measurements were performed at 20 °C with a scanning speed of 20 nm/min and 8 times averaging. Protein concentrations (in 50 mM KCl and 50 mM potassium phosphate buffer pH 7.4) were adjusted to an optical density of 0.2 at 450 nm. The ellipticity at 250 nm was recorded for each sample and used as a baseline value[131].

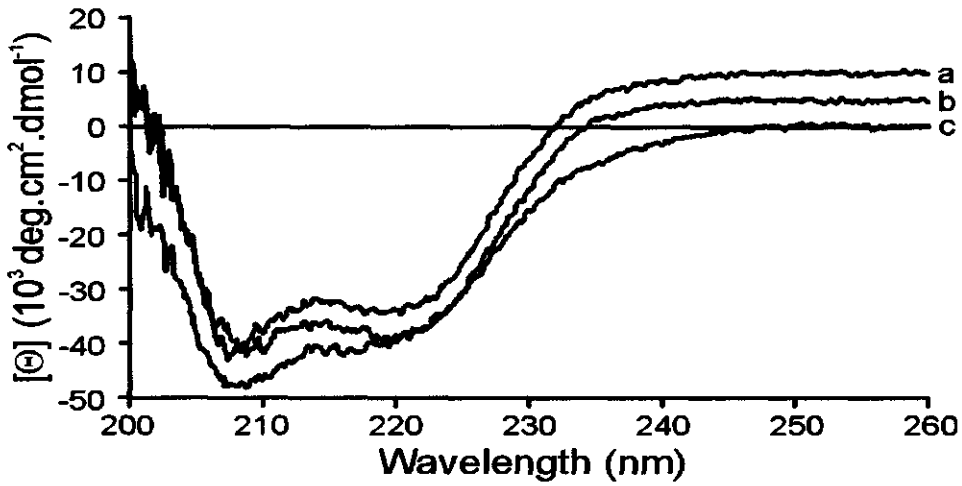


Figure 4.3 Circular dichroism spectra of the NifL PAS domain. The untreated holo protein (line **a**) and the ^{13}C-FAD (**b**) and ^{13}C-FMN (**c**) reconstituted proteins show similar characteristics, indicating similar secondary structure elements. All proteins were diluted in 50 mM potassium phosphate pH 7.4, 50 mM KCl to an A_{450} = 0.2 (lines (**a**) and (**b**) were shifted upwards for clarity).

For ^{13}C-NMR experiments, 0.7 mM protein was dissolved in 50 mM Tris-HCl, 50 mM KCl, pH 7.4, containing 10 % D_20. Measurements were performed at 15 °C using a 9.4 T Avance 400 NMR spectrometer. Proton decoupling was achieved with a GARP sequence and spectra were recorded as described before[307, 310]. For measuring protein samples in the reduced state, the Wilmad 5 mm NMR tubes were filled with 0.50 ml sample, and de-aerated by flushing with argon for 10 minutes. Subsequently, a 10-fold excess of buffered dithionite over flavin concentration was added.

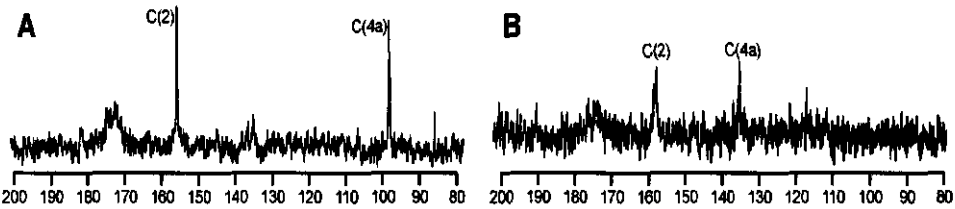


Figure 4.4 ^{13}C-NMR spectra of (**a**) reduced 2,4a-$^{13}C_2$-FAD containing NifL PAS domain, and (**b**) oxidised 2,4a-$^{13}C_2$-FMN containing NifL PAS domain.

RESULTS AND DISCUSSION

Apoprotein preparation

FAD release from the NifL PAS domain with 2 M KBr and 2 M urea (Figure 4.1) is very efficient and can easily be monitored visually. Instead of on-column reconstitution, it is possible to elute the apo PAS domain from the IMAC column with two column volumes of a buffer containing 250 mM imidazole (Figure 4.1F). Unfortunately, the free apo-protein of the PAS domain formed aggregates immediately after removal from the IMAC column and is not stable in solution.

Reconstitution of apoflavoprotein

Because the apo-PAS domain is not stable in solution, we explored the possibilities of on-column replacement of the FAD cofactor with an alternative cofactor. The strong interaction between the N-terminal His-tag of the protein and the Ni-NTA matrix turned out to facilitate this replacement. Reconstitution of the column-bound apoprotein with ^{13}C-FAD or ^{13}C-FMN is highly effective with a 100 % cofactor replacement yield. However, due to the extensive washing steps about 15 % of the loaded protein is lost.

Properties of the reconstituted NifL PAS domain

Figure 4.2 shows that replacement of the native FAD cofactor with ^{13}C-FMN or ^{13}C-FAD does not change the characteristic flavin absorption spectrum of the PAS domain. The raise in baseline for the apoprotein at lower wavelengths is caused by small aggregates formed. From the spectrum can be concluded that the sample indeed is flavin-free, and thus represents the apo-form of the protein. Figure 4.3 shows the far-UV CD spectra of the native, and the reconstituted PAS domain protein. Unfortunately, the aggregation of the apo-protein prevented reliable CD measurements. However, after on-column reconstitution the PAS domain does fold nicely, as shown by the identical CD-spectra of the holo- and FAD-reconstituted form, and with marginally different characteristics of the FMN-reconstituted form. Analytical gel filtration experiments using a Superdex 200 30/10 column revealed that the tetrameric state of the NifL PAS domain does not change after exchange of the FAD for the FMN cofactor. Moreover,

comparison of the protein concentration with the flavin concentration showed that the cofactor to protein ratio is not changed.

To further demonstrate the powerfulness of the His-tag based flavin exchange technique, the ^{13}C-NMR spectrum of the 2,4a-^{13}C-FAD reconstituted NifL PAS domain protein in the reduced state is shown in Figure 4.4. Two peaks arising from the ^{13}C labeled FAD are visible, at 156.2 (C_2) ppm and 99.8 (C_{4a}) ppm, respectively. For the reduced 2,4a-^{13}C-FMN reconstituted protein the C_2 and C_{4a} atoms have the same ^{13}C chemical shift (156.2 ppm and 99.8 ppm), indicating that there is no difference in the chemical environment of the isoalloxazine moiety of reduced FAD- or FMN-reconstituted PAS domain. Moreover, the ^{13}C-NMR linewidths of the bound FAD and of the bound FMN are nicely Lorentzian shaped, indicating that the cofactor is tightly bound with absence of conformational averaging. The ^{13}C chemical shift of the C_2 and C_{4a} values are in close agreement with the corresponding values[79] of *p*-hydroxybenzoate hydroxylase[311]. This indicates that in the reduced state the NifL PAS domain protein has a negatively charged FAD cofactor. In the oxidised state the C_2 and C_{4a} ^{13}C resonances of FAD- and FMN- labeled PAS domain are 157.4 ppm, 134.7 ppm (FAD) and 157.2 ppm, 134.7 ppm (FMN), indicating again that there is no difference in the chemical environment of the two flavin cofactors. The ^{13}C linewidths indicate that in the oxidised state conformational averaging occurs.

CONCLUSIONS

This paper clearly shows that the FAD cofactor of the His-tagged PAS domain of NifL can be replaced on-column by IMAC. The polyhistidine tag allows a strong interaction between the protein and the IMAC column material. The efficient removal of FAD, and the subsequent on-column reconstitution, is fast and can be used on a preparative scale.

Since the apo-PAS domain is not very stable in solution, the on-column reconstitution is the method of choice to prevent precipitation of this apo-protein. On-column reconstitution of the flavin cofactor with ^{13}C-FAD or ^{13}C-FMN does not change the oligomerisation state of the NifL PAS domain, and neither does it change the chemical environment of the flavin. Thus, this method is especially useful when the apoform of the protein is not stable in solution due to aggregation and precipitation, since it stabilises the apoform of the protein during the reconstitution procedure.

In principle, His-tags can be introduced in every recombinant protein. Any His-tagged protein with a non-covalently bound cofactor can be immobilised on an IMAC column, opening possibilities to exchange the cofactor in high yield. This cofactor does not necessarily need to be a flavin. Therefore, on-column replacement of protein-bound cofactors by IMAC may develop as a universal method for replacement of flavin or other cofactors.

CHAPTER 5

Crystallisation and Partial three-dimensional Structure of the NifL PAS domain protein from Azotobacter vinelandii

Part of this chapter has been published:

Marco H. Hefti, Jörg Hendle, Cristofer Enroth, Jacques Vervoort and Paul A. Tucker
Acta Crystallographica Section D -*Biological Crystallography*- **D57**, 1895–1896 (2001)

ABSTRACT

The *A. vinelandii* NifL protein is a redox-sensing flavoprotein, which inhibits the activity of the nitrogen-specific transcriptional activator NifA. The N-terminal PAS-domain has been overexpressed in *E. coli* and crystallised by the hanging-drop vapour-diffusion method. The crystal belongs to the rhombohedral spacegroup *R*32, with unit cell parameters $a = b = 64.9$ Å, $c = 157.4$ Å and has one molecule in the asymmetric unit. Native data were collected to 3.0 Å on the BW7B synchrotron beamline at the EMBL Hamburg Outstation. Initial attempts to crystallise Seleno-Methionine labeled protein failed, but structure elucidation using molecular replacement strategies seems to be successful. In this chapter preliminary data of the partially elucidated structure are reported (Jason Key *et al.* unpublished results).

INTRODUCTION

PAS-domains are ubiquitous motifs present in bacteria, archaea and eucarya[332]. Unlike most other sensory domains, PAS-domains are located in the cytosol[287] and include histidine[1] and serine/threonine kinases[231], chemoreceptors and photoreceptors for taxis and tropism[269], circadian clock proteins[128], voltage-activated ion channels[316], cyclic nucleotide phosphodiesterases[266], as well as regulators of responses to hypoxia[121] and embryological development of the central nervous system[210]. The specificity of a PAS-domain for detecting input signals is partly determined by the cofactor associated with the PAS-domain. Five PAS-domain protein structures are known; the bacterial blue-light photoreceptor PYP[27, 92], the heme-binding domain of the rhizobial oxygen sensor FixL[97, 191], the N-terminal domain of the human ether-a-go-go-related potassium channel HERG[198], the FMN containing phototropin module of *Adiantum* PHY3[49], and the human PAS kinase[4]. To date, no crystal structures of PAS-domains are known which contain FAD as cofactor. The *A. vinelandii* NifL protein is an FAD-containing redox-sensing protein which inhibits the activity of the nitrogen-specific transcriptional activator NifA, in response to molecular oxygen and fixed nitrogen[57]. The N-terminal PAS-domain serves as the flavin redox-sensing domain[265]. In order to understand the different mechanisms by which PAS-domains work, detailed information about their cofactor binding regions is required.

PROTEIN PRODUCTION AND CRYSTALLISATION

The amino-terminal PAS-domain of *A. vinelandii* NifL, containing an N-terminal histidine tag with an adjacent thrombin cleavage site, was overexpressed in the *E. coli* strain BL21 (DE3) pLysS and purified as described previously[114]. For the production of selenomethionyl-labeled protein *E. coli* strain BL21 (DE3) pLysS was first grown on LB plates supplemented with 50 μg/ml kanamycin (Sigma). Single colonies were transferred

to 5 ml prewarmed LB medium (37 °C) containing 50 μg/ml kanamycin. After overnight growth 500 μl medium was transferred to a 2-litre flask containing 500 ml prewarmed non-sterile M9 medium[246]. This M9 minimal medium was supplemented with 50 mg/l of all 20 aminoacids (in which methionine was replaced with selenomethionine), 100 mg/l kanamycin and 0.6 % w/v glucose. Cells were grown with vigorous shaking (300 rpm) for 15 hours, before protein expression was induced with 1 mM IPTG. After 6 hours the cells were harvested and the protein purified as described before[114], resulting in 9 mg pure protein from a 4 litre culture. Prior to crystallisation the freeze-dried protein was dissolved in double-distilled deionised water and the protein concentration was determined using Bradford's assay. Initial crystallisation trials were carried out using the hanging-drop vapour-diffusion method on Linbro plates. A hanging drop (2 μl) prepared by mixing 1:1 each of the protein solution and the reservoir solution was placed on a siliconised cover slip over 0.6 ml of the reservoir solution. Initial screening for crystallisation conditions was performed with Screen I[120] and II (Hampton Research), and with Wizard I and II (Emerald Biostructures) screening solutions using protein concentrations of 6 to 8 mg/ml, at 293 K. Initial crystal growth was observed with 2.0 M ammonium sulfate as salt, and 5 % v/v isopropanol as precipitant. Because crystals were observed in conjunction with amorphous precipitate, subsequent crystallisation experiments were carried out to try to optimise the conditions, including pH, salt and precipitant concentration, as well as protein concentration. Protein crystals always formed after initial light precipitate (Figure 5.1), and seems to be insensitive to the pH used.

Figure 5.1 Yellow crystals of NifL PAS-domain protein, containing the FAD cofactor, grown by the hanging drop method. The average dimensions of these crystals were 0.1x0.1x0.1 mm.

The best conditions were found to be 1.8 M ammonium sulfate and 4.5 % isopropanol. Small crystals of the selenomethionine containing protein formed after 12 days under the same conditions, except for the addition of 10 mM β-mercaptoethanol to the well-solution, but they dissolved again after a further five days. No new crystal growth was observed subsequently.

X-RAY ANALYSIS

Diffraction data to 3 Å resolution were collected at 100K with synchrotron radiation (λ = 0.8424 Å) at the BW7B beamline at the EMBL-Hamburg outstation using a MAR345 image plate detector. A total of 110 images were recorded with an oscillation angle of 0.5° and an exposure time of 4 minutes per image. The crystal-to-detector distance was set to 300 mm. Prior to data collection, a crystal was equilibrated for a few seconds in a cryoprotectant solution prepared by mixing 1:1 well-solution and 50 % v/v glycerol, then directly mounted in a CryoLoop (Hampton Research) and flash-cooled in the cold nitrogen stream. The intensity data were processed and scaled with the HKL package[220]. A high rejection probability was used in SCALEPACK (0.099) during scaling because of the high completeness and redundancy of the data. The crystals are rhombohedral with a=64.9 Å, c=157.4 Å. With one molecule in the asymmetric unit the Matthews coefficient is 2.01[178]. Data collection details are given in Table I, and a typical crystal is shown in Figure 5.1.

Table 5.1 Data collection statistics for the NifL PAS domain crystal.

X-ray source	EMBL Hamburg, BW7B
Wavelength (Å)	0.8424
Temperature (K)	100
Resolution (Å)	40 - 3.0 [3.11 - 3.0]
Space group	*R*32
Unit-cell parameters (Å)	*a* = 64.89, *c* = 157.39
Total observations	28038
Independent reflections	2667
Completeness (%)	97.5 [98.5]
$I/\sigma(I)$	17.7 [3.3]
R_{sym} (%)	5.1 [38.8]
Mosaicity (°)	0.82

Values in square brackets are for the highest resolution shell

MASS-SPECTROMETRY STUDIES

Prior to crystallisation the mass of the purified PAS domain protein was checked with MALDI-TOF to verify the correctness of the thrombin cleavage. The determined mass (15897.5 Da) was in excellent agreement with the expected theoretical mass (15898 Da). Protein crystals were washed in well-solution and dissolved in the MALDI-TOF matrix solution. Mass spectra were recorded, with a low signal to noise ratio. The estimated mass was approximately 16000 ± 150 Da, indicating that the initial crystals

contained thrombin-cleaved protein. Selenomethionyl labelled protein contains two selenium atoms (15991 Da), as expected (theoretical mass 15991.3 Da) from the number of methionines in the protein sequence. The selenium incorporation was over 95 % efficient, estimated from the peak-heights.

MOLECULAR REPLACEMENT

A number of search models for the NifL PAS domain were constructed based upon the known structures of PHY3, PYP, FixL, and HERG. Of these, the most successful models by far were constructed using the FMN containing structure PHY3. The search model contained the side chains of 15 residues that are strictly conserved between NifL and PHY3. All other amino acid residues, except glycine, were changed to alanines. Thus, a search for the 140 NifL PAS residues was performed with a search model containing 100 residues, of which 85 were glycine or alanine.

The program EPMR[134] is a program that finds crystallographic molecular replacement solutions using stochastic search models. The program simultaneously optimises the three rotational and three positional parameters for a search model. EPMR was used to search for the molecular replacement solution using several resolution ranges between 25 and 3 Å. The range between 10 to 4 Å yielded the best results. Due to inadequacy of the used search models, all solutions produced by EPMR had low correlation coefficients and high R-factors. In order to find the correct solution, several thousand runs were performed with each search model, while writing out each solution for each run. The resulting solutions with the most promising correlation coefficients and R-factors were examined for realistic packing interactions in the cell unit. Most candidate solutions violated packing considerations, produced overlapping symmetry mates, and were therefore incorrect. One solution consistently appeared with both the highest correlation coefficient and lowest R-factor, and showed a unit cell with reasonable packing interactions (Figure 5.2a). The correlation coefficient of this solution, containing only 15 residues with side chain atoms, was 37 % with an initial R-factor of 55.5 %. This solution occurred in less than 0.8 % of the EPMR runs. After adding the co-ordinates of FMN from PHY3 to the search model, the correlation coefficient increased to 41.5 % with an R-factor of 54.8 %. With this addition of FMN, the solution occurred in approximately 2 % of the EPMR runs.

OMIT MAPS

In order to determine if the molecular replacement solution was correct, $2F_o$-F_c maps were calculated from the original search model which lacked most side chain residues and the FAD cofactor. The FAD omit maps revealed density for the flavin isoalloxazine ring as well as the phosphate tail of FAD (Figure 5.2b). In addition, density is visible for

the majority of the peptide backbone of the model, and a number of side chains. Although the initial R-factor of this solution was 55 %, it was possible to add a number of side chains to the model, and build part of the N- and C-terminal residues. Upon doing so, the free R-factor dropped to 48 %. Density modification by solvent flipping was useful in building some portions of the model, and currently the R-factor is 42 %. To date, of the 143 amino acid residues, 117 backbone atoms and about 75 side chain residues are added to the search model (in total 755 atoms). This means that about half of the residues in the search model lack their side chain.

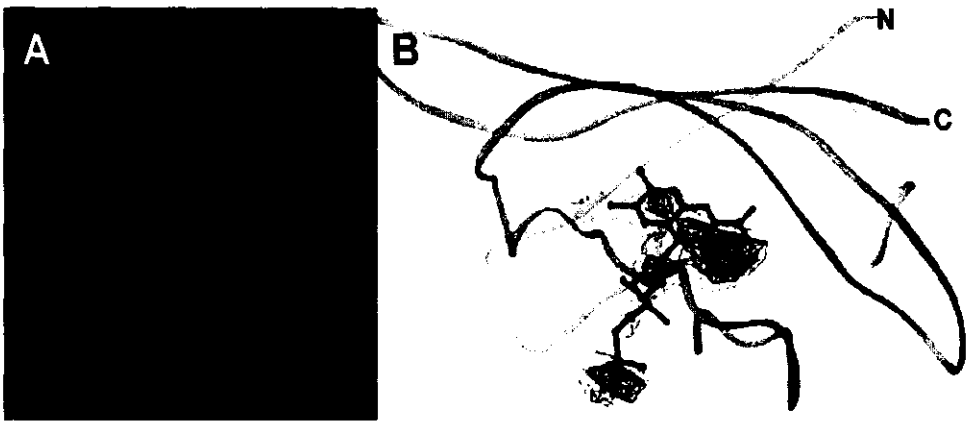


Figure 5.2 **(A)** Crystal packing interactions of the EPMR solution for the NifL PAS domain in the *R*32 unit cell. **(B)** The $2F_0$-F_c map reveals electron density for the FAD flavin ring. The map is contoured at 1σ (cyan) and 2σ (blue).

DISCUSSION

It is questionable as whether this model can be refined to a satisfactory R-factor. Due to the low resolution and the incompleteness of the model, there is some difficulty using software which relies on molecular dynamics (simulated annealing, conjugate-gradient minimisation) to refine the model. As the model is continuously improved, this should become less of a problem. That means that at this stage considerable manual refinement has to be done. Significant sections of the model are incorrect, most obviously the first turn of the N-terminus.

Given the high completeness and $I/\sigma(I)$ in the last resolution shell (Table 5,1), it seems likely that higher resolution data could be obtained from these crystals. To date, no crystals with diffraction beyond 3 Å have been obtained. Growing high quality crystals may be a challenging task, being a key to a successful structure determination and to sound biological interpretation.

ACKNOWLEDGEMENTS

This chapter presents ongoing work about the structure determination of the NifL PAS domain. I would like to thank several colleagues who have contributed to this work. Ray Dixon and Richard Little (Department of Molecular Microbiology, John Innes Centre, Norwich) have supplied the plasmid necessary for this study, and Sjef Boeren advised about mass spectrometry of the protein crystals. Further, I would like to thank Paul Tucker, Jörg Hendle, and Cristofer Enroth from the EMBL outstation in Hamburg, as well as Jason Key (University of Chicago) for running thousands of EPMR calculations.

CHAPTER 6

Low Resolution Structure of the Azotobacter vinelandii NifL PAS Domain revealed by X-ray Solution Scattering

ABSTRACT

The *A. vinelandii* NifL protein is a redox-sensing flavoprotein which inhibits the activity of the nitrogen-specific transcriptional activator NifA. The N-terminal PAS domain of NifL is a sensory domain, with 4 FAD molecules per tetramer. This domain was examined by synchrotron small-angle X-ray solution scattering. The experimental scattering data leads to a global fold of the tetrameric domain at low-resolution. Fitting a model of the PAS domain to this dummy atom model leads to a better understanding of the dimerisation / tetramerization interface. The resulting low resolution structure of the NifL PAS domain shows a well developed, anisometric shape of four monomers attached to each other. This tetrameric form is in agreement with analytical gel filtration experiments which also conform a homo-tetrameric state of this domain.

INTRODUCTION

In *A. vinelandii* the redox-sensitive, flavin-containing sensor protein NifL regulates NifA, an enhancer binding protein that controls the expression of nitrogen fixation genes in response to the concentration of fixed nitrogen and extracellular oxygen[57, 116]. Signal transduction is mediated via protein - protein interactions[193, 194, 254]. NifA is active in the absence of NifL, both *in vivo* and *in vitro*. The NifA protein activates transcription at the σ^N-dependent promoters for nitrogen fixation genes. NifL negatively controls the transcriptional activation by NifA, by acting as a specific anti-activator in response to the environmental signals. The signals that lead to this inhibitory effect of NifL are diverse. It is activated in response to the level of fixed nitrogen, as well as the redox state[152, 153, 240].

The NifL protein is comprised of two domains; the N-terminal PAS domain and the C-terminal ADP-binding domain, which are linked by a Q-linker[25]. Q-linkers are relatively short, hydrophilic sequences, rich in Arg, Gln, Glu, Pro and Ser. The Q-linker serves a simple but essential role in tethering the structurally distinct, but interacting domains of regulatory proteins[324]. The C-terminal domain of NifL shows homology to the histidine protein kinase transmitter domains[25]. The presence of ADP increases the inhibitory activity of NifL, binding to the C-terminal domain of NifL, independent of the redox state[74, 265]. The NifL redox sensing function is well understood[116, 155, 254], and is located in the N-terminal PAS domain. This domain of the *A. vinelandii* NifL protein senses the redox status, and is a member of the PAS domain superfamily[332]. These PAS domains are important protein domains that can monitor changes in oxygen (FixL), light (e.g. phytochromes), and redox state (Aer, NifL, ArcB) of a cell. These motifs have been found in bacteria, eucarya and archaea. They are, unless most of the other sensory domains, located in the cytoplasm[287].

NifL is a homotetramer in solution in oxidised form, requiring the PAS domain region to retain this tetrameric state[116, 265]. Also NifA is a homotetramer in solution. In

the presence of MgADP, NifL-NifA complexes, which were either isolated from cell extracts or formed *in vitro*, sieved with an apparent molecular mass of approximately 500 kDa on a Superose 12 column. This would be consistent with stoichiometry of a tetramer of NifL bound to a tetramer of NifA[194]. The FAD:protein tetramer ratio is believed to be 4:1 both for NifL as for the PAS domain. This is observed when using a fast purification protocol, or after reconstitution of apo-NifL or apo-PAS domain with FAD[113], although in the past this number used to vary between 3:1 and 2:1[114, 116].

EXPERIMENTAL PROCEDURES

Sample preparation

The amino-terminal PAS domain of *A. vinelandii* NifL, containing an N-terminal histidine tag with an adjacent thrombin cleavage site, was overexpressed in the *E. coli* strain BL21 (DE3) pLysS and purified as described previously[114]. This PAS domain (amino acid residue M1-H140, monomeric mass 17.8 kDa) with N-terminal extension (MGSSHHHHHHSSGLVPRGSH) was concentrated to 15 mg/ml and buffer-exchanged into 50 mM Tris-HCl, 300 mM NaCl. pH 7.5, at 4 °C. The protein was centrifuged at high speed for 20 minutes, to avoid large protein aggregates in the sample, before scattering experiments. Protein concentration was determined as described by Bradford[31] using Bradford reagent (Sigma) with bovine serum albumin (Boehringer) as a standard.

Analytical gel filtration

The NifL protein from *A. vinelandii* is a tetramer in solution[116, 265]. To find out if the same applies for the PAS domain analytical gel filtration experiments were run on an Äkta Explorer (Pharmacia Biotech). 50 µl of the His-tagged PAS domain protein (2 mg/ml), in 100 mM potassium phosphate buffer, 150 mM KCl (pH 7.0) was loaded on a Superdex 200 10/30 column at a flow speed of 0.5 ml/min. Calibration of this column was performed with blue dextran, alcohol dehydrogenase, ovalbumin, myoglobin, and cytochrome C. The retention time of the PAS domain sample corresponds to a tetrameric form of the protein.

MALDI mass-spectrometric analysis

Matrix-assisted laser desorption/ionisation - time of flight (MALDI-TOF) mass analysis was carried out on a PerSeptive Biosystems Voyager DE RP Biospectrometry Workstation. The matrix solutions used to prepare the protein samples were prepared, and the MALDI-TOF operated, as described in[114]. The mass spectrometer was calibrated externally, with a mixture of 8 proteins in the range from 1 to 66 kDa. In Figure 6.1 the MALDI-TOF MS spectrum is shown of this protein. The NifL PAS domain protein, including the N-terminal His-tag, has a experimental average mass of

17652 Da ± 25 Da (peak A). This is in close agreement with the theoretical mass of the NifL PAS domain, without the N-terminal methionine (17648 Da). It is often observed that this N-terminal Methionine is missing[254]. An additional peak at 18433 Da is observed (peak B). The difference between both peaks is 780 Da, caused by the addition of one FAD cofactor (783.5 Da).

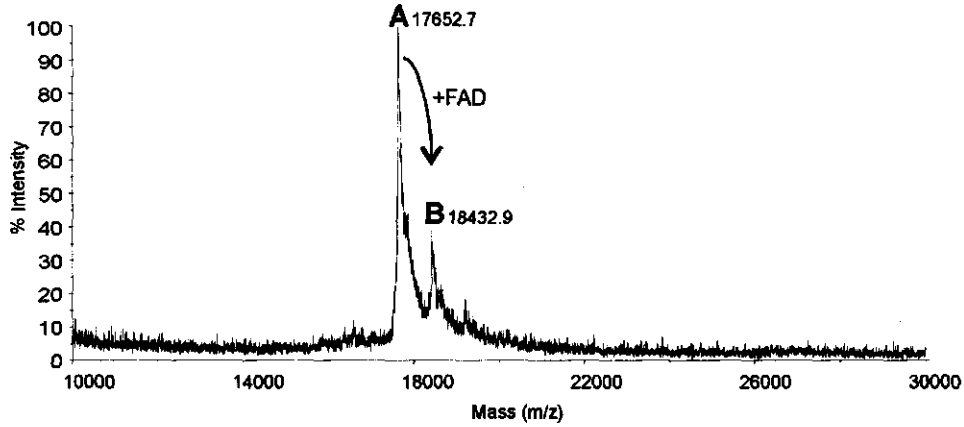


Figure 6.1 MALDI-TOF MS spectrum of the His-tagged NifL PAS domain. Peak A (17652 Da) corresponds to the average mass of one monomer, while the mass-increase of 780 Da is caused by FAD (peak B).

Modelling of the NifL PAS domain

The partially elucidated X-ray structure listed in Chapter 5 is used to create a model of the monomeric PAS domain. In this X-ray structure, all amino acid side-chain residues that have not been elucidated (yet) are substituted for alanines. This protein sequence is shown in Figure 6.2.

```
                                10        20        30        40        50        60        70        80
                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NifL 3D structure    ---------- ---------- ---------- ---------- -ALPAIARQA VEHAPAAIAA TDLAAAILAA NRAFATAAGY
NifL model           MGSSHHHHHH SSGLVPRGSH MTPANPTLSN EPQAPHAESD ELLPEIFRQT VEHAPIAISI TDLKANILYA NRAFRTITGY
                                    Thrombin
                                90        100       110       120       130       140       150       160
                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NifL 3D structure    GSEEALGAAA SILANGAAPA AVAQALAGAA AAAKPWSGAL ANAAAAATLY LAELTVAPVL AAAGATAYYA GAHRAA----
NifL model           GSEEVLGKNE SILSNGTTPR LVYQALWGRL AQKKPWSGVL VNRRKDKTLY LAELTVAPVL NEAGETIYYL GMHRDTSELH
```

Figure 6.2 Sequence alignment of the NifL PAS domain model with the partially elucidated X-ray structure. All amino acid residues that are identical in both sequences are bold and underlined, and the thrombin cleavage site is indicated.

There are two reasons why there is no template available for the N-terminal 41 amino acid residues of the model sequence. First, the protein used for SAXS experiments had an N-terminal His-tag and thrombin cleavage site. This site is removed in the protein crystallography experiments, resulting in a difference in sequence length of 17 amino

acid residues between the protein used for x-ray crystallography and the protein used for the SAXS measurements. Second, the N-terminal region of the protein sequence of the protein used in the x-ray studies (24 residues in total) has still not been elucidated in the partially elucidated X-ray structure, probably due to large mobility of this area. The protein sequence was modelled to the partially elucidated NifL PAS structure (Figure 6.2) using MODELLER version 6.2[2]. From 100 models created, the 9 best models were used to create low resolution SAXS solution models. These 9 models were fitted against the experimental SAXS data. This resulted in one tetrameric model with the best fit. This model was energy-minimised using the Amber forcefield in Insight II. The number of iterations was set to 1000, and the pH was 7.0. After energy minimisation, a monomeric model was fitted again against the experimental SAXS data, resulting in a new tetrameric model with a lower χ^2. After another optimisation-round, the final χ^2 was 0.90. The resulting best fit is shown in Figure 6.3.

Scattering experiments and data treatment

The synchrotron radiation X-ray scattering data were collected using standard procedures on the double focusing monochromator-mirror camera X33[29, 30, 135] at the European Molecular Biology Laboratory (EMBL) Hamburg Outstation on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) using multiwire proportional chambers with delay line readout[89]. Samples with protein solutions at concentrations between 5 and 18 mg/ml were analysed. The scattering curves were recorded at a wavelength λ = 0.15 nm for a detector-sample distance of 1.7 m covering the momentum transfer range $0.15 < s < 2.3$ nm^{-1} ($s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle). Fifteen successive frames of 60 seconds each were recorded for each sample. The data were normalised to the intensity of the incident beam, and corrected for the detector response. The scattering of the buffer was subtracted, and the difference curves were scaled for protein concentration and extrapolated to infinite dilution following standard procedures[76] using the data reduction and processing program SAPOKO[282]. The molecular masses (MM) of the solutes were calculated by comparison with the forward scattering from reference solutions of bovine serum albumin (MM = 66 kDa). The maximum dimensions of the particles, D_{max}, were estimated from the experimental data using the orthogonal expansion program ORTOGNOM[280]. The forward scattering I(0), and the radii of gyration R_g were evaluated using the Guinier approximation[102] assuming that at very small angles ($s<1.3/R_g$) the intensity is represented by

$$I(s) = I(0)^{-s^2 R_g^2/3}$$

These parameters were also computed from the entire scattering patterns using the indirect Fourier transform program GNOM[279, 283]. This program also provides the distance distribution function, p(r), of the particles.

Shape determination / Structure modelling

By comparison of the extrapolated forward scattering I(0) with that of a reference solution of bovine serum albumin, the molecular masses of the solutes were estimated. Prior to the shape analysis, undesirable contributions from the scattering due to the internal particle structure at higher scattering angles that become significant above approximately s = 2.0 nm^{-1} were removed by subtracting a constant from the experimental data. The Porod volume V_p of the PAS domain was calculated from the processed scattering curves $I_p(s)$ as

$$V_p = \frac{2\pi^2 I(0)}{\int_0^\infty I_p(s) s^2 ds}$$

where the computation of the integral in the denominator was performed as described[76]. This procedure ensures that the intensity decays as s^{-4} following Porod's law[234] for homogeneous particles and yields an approximation of the "shape scattering" curve (i.e. scattering due to the excluded volume of the homogeneous particle with a constant density). The shape scattering curves in the ranges up to s_{max} = 2.0 nm^{-1} were used to compute the excluded volumes V of the hydrated particles and for the *ab initio* shape determination. The outer portions of the curves, dominated by the scattering from the internal structure, were discarded in the further analysis.

The low resolution model of the NifL PAS domain was built from the experimental data using the DAMMIN program. This program uses a single phase dummy atom model for the *ab initio* shape determination by simulated annealing[281]. A sphere of diameter D_{max} is filled by densely packed small spheres (or dummy atoms) of radius r_0 << D_{max}. The scattering intensity I(s) of a given dummy atom model (or DAM) is calculated by a global summation over all dummy atoms (M >> 1) using spherical harmonics expansion, to represent partial amplitudes. The scattering intensity from the DAM is computed using

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| A_{lm}(s) \right|^2$$

with the following partial amplitudes,

$$A_{lm}(s) = i^{-l} \sqrt{2/\pi} \sum_j j_l(sr_j) Y_{lm}^{+}(\omega_j)$$

where the sum runs over the dummy atoms. The spherical Bessel function is denoted by $j_l(x)$, and r_j and ω_j are the particle atoms polar co-ordinates.

Finding a DAM configuration corresponding by minimising the discrepancy between experimental and DAM calculated scattering curves is an *ab initio* reconstitution. Each bead belongs either to the protein (index = 1) or to the solvent (index = 0), and the shape is thus described by a binary string of length M. Starting from a random string, simulated

annealing[133] is employed to find a compact configuration of the beads, minimising the discrepancy χ between the experimental $I_{exp}(s)$ and the calculated $I_{calc}(s)$ curves:
where N is the number of experimental points, c is a scaling factor and $\sigma(s_j)$ is the

$$\chi^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left[\frac{\langle c(s)\rangle I_{DR}(s_j) - cI_{exp}(s_j)}{\sigma(s_j)}\right]^2$$

experimental error at the momentum transfer s_j. The X-ray scattering curves at higher angles contain significant contribution from the internal particle structure which must be removed prior to the shape analysis. For this, a constant is given by the slope on an $s^4I(s)$ *vs.* s^4 plot, is subtracted from the experimental data to ensure that the intensity would decay as s^{-4} following Porod's law for homogeneous particles. The resulting 'shape scattering' curve (*i.e.* scattering due to the excluded volume of the particle with unit density) in the range up to s = 2.0 nm^{-1} was used in the *ab initio* shape restoration. The particle was restored assuming a *P*222 point symmetry group. The excluded volume of the hydrated particle (Porod volume) was computed from the shape scattering curve using the equation

$$V_p = \frac{2\pi^2 I(0)}{\int_0^\infty I_p(s)s^2 ds}$$

The orientation of the output DAM structure has an arbitrary orientation. Three-dimensional models of the DAM structure with a model of the NifL PAS domain were displayed (Figure 6.4) using the program ASSA[137]. The results of the analysis of the scattering curves, as well as the experimental and modelling data, are listed in Table 6.1.

Table 6.1 Experimental and modelling parameters of the NifL PAS domain.

	Experimental	Modelling	SAXS DAM [a]
Mass (kDa)	73-75 [b]	74 [c]	80 [d]

[a] Parameters of the Dummy Atom Model

[b] The analytical gel filtration retention time corresponds to a tetrameric protein

[c] Theoretical value of a tetrameric form of the NifL PAS domain protein, including N-terminal His-tag

[d] The DAM mass of 1 monomer is 80 / 4 = 20 kDa

The experimental scattering curves of the *A. vinelandii* NifL PAS domain are presented in Figure 6.3. The initial decay of these curves is linear, indicating a monodisperse protein solution.

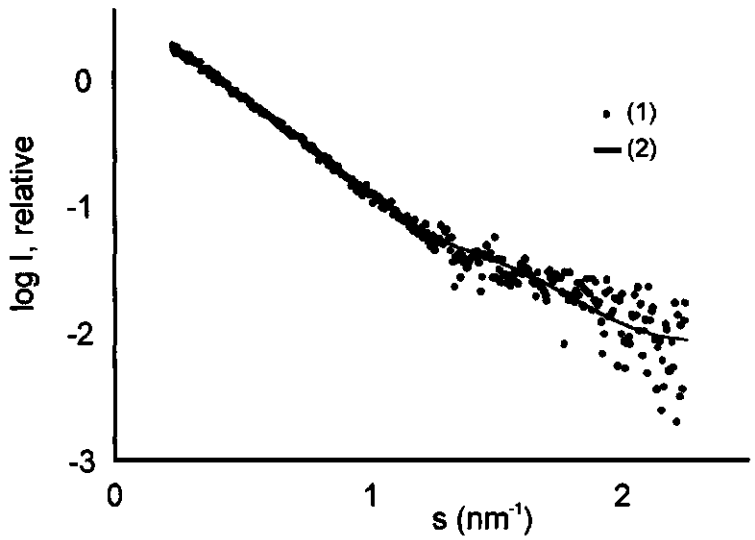


Figure 6.3 Experimental scattering curves of the NifL PAS domain from *Azotobacter vinelandii* (**1**), and calculated scattering patterns from the DAM (**2**).

DISCUSSION

The low resolution solution structure of the *A. vinelandii* NifL PAS domain protein shows a well developed, anisometric shape of four monomers attached to each other. This tetrameric form corresponds to the analytical gel filtration experiments that also conform a homo-tetrameric state. Fitting of the modelled PAS structure to the global solution structure is not straightforward, and several different fits are possible. Optimisation of the fit data could be obtained when the three dimensional X-ray structure of the NifL PAS domain is elucidated. As can be observed in Figure 6.4, the FAD (or actually FMN) cofactor) is buried inside the NifL PAS domain monomers, and the FAD phosphate groups have the space to stick out in solution. Small angle X-ray scattering measurements of full-length NifL protein might reveal that the tetramerization domain is in the PAS domain region.

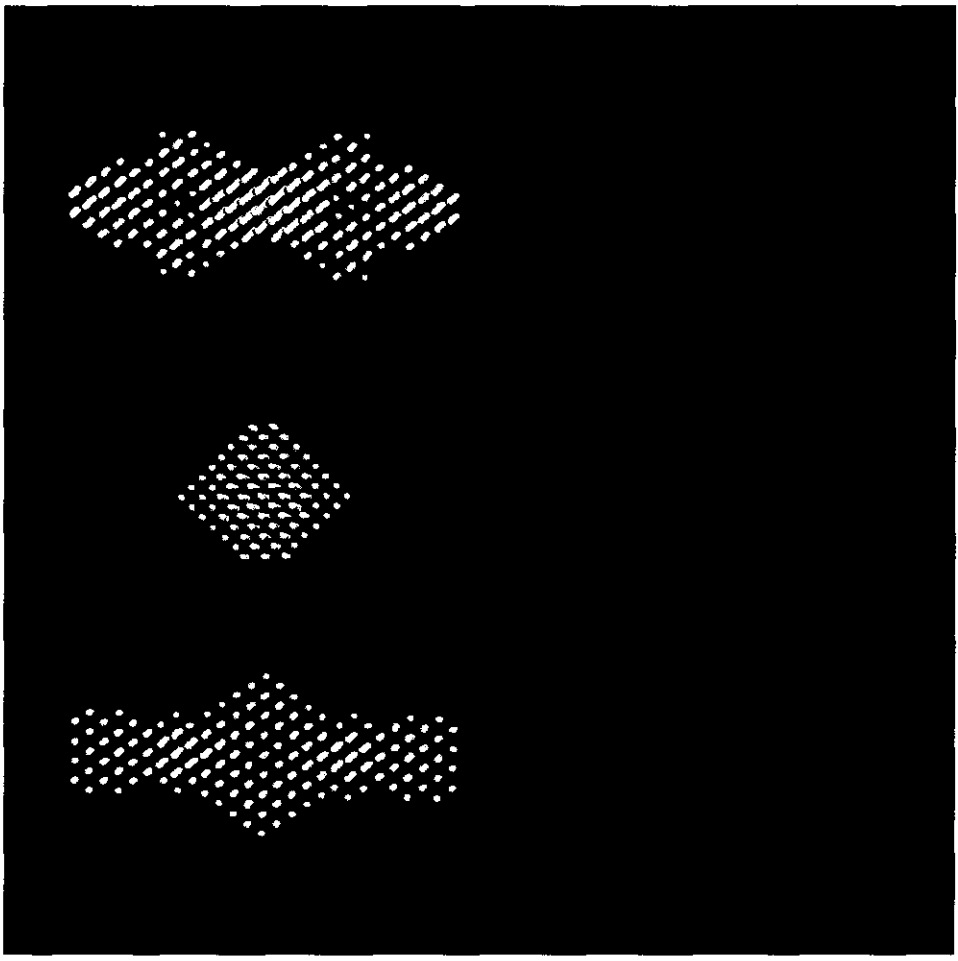


Figure 6.4 Dummy Atom Model (left) of the NifL PAS domain, assuming *P*222 symmetry, displayed in 3 different directions. On the basis of the structural homology of this domain and the FMN containing phototropin module of *Adiantum* PHY3 (PDB accession number 1G28), four FMN molecules (yellow) were modelled into the tetrameric NifL PAS domain model (right).

ACKNOWLEDGEMENTS

This chapter presents ongoing work about the low resolution structure determination of the tetrameric NifL PAS domain by X-ray solution scattering. I would like to thank several colleagues who have contributed to this work. Ray Dixon and Richard Little (Department of Molecular Microbiology, John Innes Centre, Norwich) have supplied the plasmid necessary for this study. Further, I would like to thank Marc Malfois for data collection, and Dmitri Svergun (both from the EMBL outstation in Hamburg) for analysis of the data and the numerous fits of potential models.

CHAPTER 7

The PAS fold: Redefinition of the PAS domain based upon Structural Prediction

-a Large-Scale Homology Modelling Study-

"Making a model of a horse from photographs does not tell how fast it will run"

Gutfreund, H. & Knowles, J. R. The foundations of enzyme action
Essays in Biochemistry **3**, 25-72 (1967)

ABSTRACT

In the post-genomic era it is essential that protein sequences are annotated correctly in order to assign putative functions. Over 1300 proteins in current protein sequence databases are predicted to contain a PAS domain based upon amino acid sequence alignments. One of the problems with the current annotation of the PAS domain is that this domain exhibits limited homology at the amino acid sequence level. From three dimensional X-ray and NMR structures available to date, however, it is clear that these domains have a conserved three-dimensional fold as shown here by structural alignment of six representative 3D structures from the PDB database. Large-scale modelling of the PAS sequences present in the PFAM database against the 3D structures of these six structural prototypes was performed. All 3D models generated were evaluated using ProsaII. From our large-scale modelling studies it is clear that the PAS and PAC motifs are inseparable and that these two motifs form the PAS fold. The existing subdivision in PAS and PAC motifs, as in use by the PFAM and SMART databases, appears to be linked to major differences in sequences in the loop region linking these two motifs. This loop region, as has been shown by Gardner and co-workers for human PAS kinase[4], is very flexible and adopts different conformations depending on the ligand bound. Here we propose that this loop region and the adjacent secondary elements are the main structural determinants in the biological response of PAS fold proteins. Finally, some PAS sequences present in the PFAM database did not produce a good structural model, even after realignment using Align-2D, suggesting that these representatives are unlikely to have a fold resembling any of the structural prototypes of the PAS domain superfamily.

INTRODUCTION

In 1997 Zhulin, Taylor and Dixon[332], and Ponting and Aravind[231] observed that conserved motifs representative of PAS domains were ubiquitous in archaea, bacteria and eucarya, and that many PAS containing proteins were involved in the sensing of oxygen, redox or light. **PAS** domains were first found in eukaryotes, and were named after homology to the *Drosophila* period protein (**PER**), the aryl hydrocarbon receptor nuclear translocator protein (**ARNT**) and the *Drosophila* single-minded protein (**SIM**). These domains are sometimes referred to as **LOV** domains; **l**ight-, **o**xygen or **v**oltage domains[33, 45, 49, 50, 127]. Unlike many other sensory domains, PAS domains are located in the cytoplasm[287] and are found in serine/threonine kinases[231], histidine kinases[1], photoreceptors and chemoreceptors for taxis and tropism[269], cyclic nucleotide phosphodiesterases[266], circadian clock proteins[128, 251], voltage-activated ion channels[316], as well as regulators of responses to hypoxia[121] and embryological development of the central nervous system[210]. Many PAS domains bind cofactors or ligands, which are required for the detection of sensory input signals.

The first 3D structure determined of a PAS domain containing protein was the structure of the *Ectothiorhodospira halophila* blue-light photoreceptor PYP (photoactive yellow protein[27, 92]). Pellequer and coworkers suggested that PYP is a prototype for the three-dimensional fold of the PAS domain superfamily[226]. PYP undergoes a self-contained light cycle. Light-induced trans-to-cis isomerisation of the 4-hydroxycinnamyl chromophore and coupled protein rearrangements produce a new set of active-site hydrogen bonds. Resulting changes in shape, hydrogen bonding, and electrostatic potential at the protein surface form a likely basis for signal transduction[92]. In recent years more PAS-like protein structures have been determined. This includes the 3D structure of the heme-binding domain of the rhizobial oxygen sensor FixL from *Bradyrhizobium japonicum*[97], and from *Rhizobium meliloti*[190]. FixL is an oxygen sensing histidine protein kinase, forming part of a two-component system that regulates symbiotic nitrogen fixation in root nodules of host plants[190]. The PAS domain in FixL is a heme-based oxygen sensor that controls the activity of the associated histidine protein kinase domain. FixL is regulated by binding of oxygen and other strong-field ligands. The heme domain permits kinase activity in the absence of bound ligand, but when the appropriate exogenous ligand is bound, this domain turns off kinase activity[97]. The structural resemblance of the FixL heme domain to PYP indicates the existence of a PAS structural motif, although both proteins are functionally different. In addition to the PYP and FixL protein structures, other 3D structures recently determined are the N-terminal domain of the human ether-a-go-go-related potassium channel HERG (first 3D model of a eukaryotic PAS domain[198]), the FMN containing phototropin module of the chimeric fern *Adiantum* photoreceptor[49], and the average NMR structure of the N-terminal PAS domain of human PAS kinase[4]. The above mentioned protein structures were used as structural templates for the homology modelling work described in this paper.

In order to understand the different mechanisms by which PAS domains mediate signal transduction, detailed information about their sequences and structures is needed. In the PFAM Protein Families Database[13] there are currently (version 7.8) 958 PAS domains present in 607 different proteins. According to PFAM, a PAC motif is found at the C-terminus of a subset of the PAS domains. PAS domains are defined differently by different authors. The definition used by Zhulin and coworkers[332] comprises a large sequence dataset, including S1 and S2 boxes. These sensory boxes were initially detected in bacterial sensors, and these conserved regions are present in PAS domains in all kingdoms of life. The S1 and S2 boxes are separated by a sequence of variable length.

Ponting and Aravind[231], on the other hand, split this PAS sequence into two separate regions; the PAS domain and PAC motif. These two regions roughly correspond to the S1 and S2 boxes[332], with varying lengths between the PAS domain and PAC motif. The SMART[148] and PFAM databases use the definition provided by Ponting and Aravind, thereby giving rise to an annotation system based upon two domains, PAS and PAC. Although the PAC motif is proposed to contribute to the PAS domain structure[231],

many PAS sequences in the SMART and PFAM databases are not linked to a PAC motif, raising the question about possible differences within the PAS domain superfamily. The PFAM annotation system is based upon multiple sequence alignments and profile Hidden Markov Models (HMM). Although HMM is more sensitive in detecting sequence similarities than p. e. BLAST, still HMM based profiles are dependent on sequence homology. Problems with HMM based searches may arise when proteins have virtually identical 3D structures, but limited sequence similarity. Since many protein sequences are emerging from the databases, annotation of these sequences should preferably be correct. The availability of the three dimensional structures of several PAS domain containing proteins, provide the opportunity to use three dimensional information in addition to sequence comparison. Modelling all PAS sequences annotated in the PFAM database onto the known PAS structures, results in a redefinition of this intriguing family of sensory proteins.

EXPERIMENTAL PROCEDURES

Description of the Modelling Templates

Seven crystal structures[27, 35, 92, 93, 227, 294] and 1 NMR structure[61] are known for the photoactive yellow protein (PYP) and PYP mutants from *Ectothiorhodospira halophila* in the Protein Data Bank (PDB)[19].

Table 7.1 The six representative structures selected, their Protein Data Bank accession number, and their PFAM annotated domains.

PDB name	Name	Accession number #	PFAM PAS	PFAM PAC
3PYP	PYP	P16113	PAS	-
1EW0	FixL	P10955	PAS	-
1DRM	FixL	P23222	PAS	PAC
1G28	PHY3	n/a	- *1	PAC *1
1BYW	HERG	n/a	- *1	PAC *1
1LL8	PAS kinase	n/a	PAS *1	- *1

Some proteins are not annotated in the SWISS-PROT protein sequence database or its supplement TrEMBL[8]. Therefore, they are not annotated in the PFAM database. However, PFAM has the possibility to BLAST a sequence against their HMM search profile *1.

The structure with accession number 3PYP was chosen as template structure since it has the highest resolution (0.85 Å)[93]. The oxygen sensor FixL has been crystallised from two different organisms. We selected from the two *Rhizobium meliloti* FixL structures deposited in the PDB 1EW0[190], as this has the most recent release date, and also because the resolution of the two FixL structures is identical. The five different PDB files of *Bradyrhizobium japonicum* FixL[96, 97]) have similar three dimensional folds; they are only different with respect to the bound ligand. 1DRM[97] was selected, being an apo-protein with the highest resolution (2.4 Å). The FMN binding domain (1G28[49]) of the fern photoreceptor protein from *Adiantum capillus-veneris* has a

resolution of 2.7 Å, and the N-terminal domain of the human-Erg Potassium Channel (1BYW[198]) has a resolution of 2.6 Å. The last structure used for modelling is the average NMR structure of the human PAS kinase N-terminal PAS domain (1LL8)[4]. These six representatives are listed in Table 7.1.

Structural Alignment of the Representative PAS Structures

The six representative PAS domain structures were aligned structurally using the homology module of Insight II (MSI/Biosys, San Diego, California, 1997; version 2000), running on a Silicon Graphics O_2 Workstation. The six proteins are compared automatically by calculating the root mean square difference between their alpha carbon distance matrices. Peptide segments are classified as being conserved when they have similar local conformations and similar orientations with respect to the rest of the protein. In regions of structural conservation among the proteins, the amino acid sequences are aligned, and atom co-ordinates are assigned based upon these alignments.

Table 7.2 Backbone root mean square deviation values (in Ångstrom) of the structural alignment of the six representative structures present in the Protein Data Bank.

1.25	3PYP	1EW0	1DRM	1G28	1BYW	1LL8
3PYP	0	1.0	0.9	1.4	1.3	1.5
1EW0	1.0	0	0.7	1.2	1.5	1.3
1DRM	0.9	0.7	0	1.2	1.5	1.3
1G28	1.4	1.2	1.2	0	1.0	1.7
1BYW	1.3	1.5	1.5	1.0	0	1.5
1LL8	1.5	1.3	1.3	1.7	1.5	0

Alignment Strategy

All PFAM annotated PAS sequences, including those from proteins containing multiple PAS domains, creating a list of 958 PAS sequences. The PFAM alignment of the PAS domains was used as initial alignment. All amino-acid residues extending from the N-terminal end of the PAS domain were deleted manually, and all sequences were extended C-terminally of the PFAM PAS domain in order to incorporate the PAC motif. If a sequence had a PFAM annotated PAC motif C-terminal to the PAS domain, the corresponding alignment was used. If no PAC motif was present, the sequence was elongated to a length similar to the other sequences based upon the genomic information available in public databases. This is the best possible option available, as a Hidden Markov Model search in PFAM did not result in the assignment of a PAC motif at the C-terminal end of many PAS domains, most likely due to the limited sequence homology to the PFAM HMM defined PAC motif. In this way, an alignment of 958 protein sequences was created, with an average length of 105 amino acid residues per sequence. Each of the sequences was modelled against all six template structures representative for the PAS fold.

The PAS and PAC annotated sequences of four organisms were studied in greater detail. All PAS annotated sequences from *Arabidopsis thaliana*, *Escherichia coli*,

Azotobacter vinelandii and *Caenorhabditis elegans* were re-aligned using the Align-2D command within MODELLER version 6.2 (Table 7.3). This enables the alignment of a sequence with a structure in comparative modelling, as amino acid sequence gaps are placed in a better structural context, and could improve the alignments provided by PFAM[2]. There are eight PFAM PAC annotated sequences in these four organisms, which lack a PAS domain N-terminally of the PAC motif. These sequences were elongated N-terminally, to incorporate a possible PAS region present. The PAC alignment as present in the PFAM database, was not changed, and the N-terminal region was aligned manually. Also these sequences were re-alignment using Align-2D. These sequences and the modelling results are listed in Table 7.4.

Homology Modelling

Models of all 958 PAS containing sequences were generated using MODELLER version 6.2[2, 77, 159] running on a dual processor Xeon 1.7 GHz *Pentium* computer with 1 Gb RAM, with REDHAT Linux release 7.3. The average calculation time for one model was about 90 seconds, resulting in six days of computer calculations. To optimise CPU usage, not more than 3 MODELLER jobs were running at the same time. For the resulting 6*958 protein models, the Prosa z-score was calculated using ProsaII version 3.0[262]. MODELLER is an implementation of an automated approach to comparative structure modelling by satisfaction of spatial restraints. As input it requires an alignment file and a pdb-file of the template structure. As output it generates a pdb-file of the model. Default settings were used, and the molecular dynamics refinement level was set to 2. The Align-2D command in MODELLER aligns a block of sequences with a block of structures, using a variable gap opening penalty. This gap penalty can favour gaps in exposed regions, and avoid gaps within secondary structure elements. The Align-2D command can be used to try to improve the existing alignment, but does not always result in a better quality of the 3D model generated.

RESULTS

Alignment of Existing Structures

Six structures were chosen (listed in Table 7.1) as representatives of the 19 PAS domain structures in the PDB database for comparative analysis. The other 13 structures (mutants or structures containing a different cofactor) have very similar 3D structures as the six representatives. Of these six structures, all N- and C-terminal amino acid residues that did not align after superimposition (Figure 7.1A) were removed from the corresponding alignment file manually (Figure 7.1D). The alignment obtained incorporates the two previously identified regions, the PFAM PAS and PAC motifs (The areas on which the structural alignment is based, is indicated with a black bar below the sequence alignment in Figure 7.1D). In this way, the sequences were trimmed back to a

sequence length in which the common fold observed was equivalent for all six proteins. The root mean-square deviation for this alignment is 1.25 Å, indicating high structural overlap. As some structures are more closely related than others, Table 7.2 shows the partial root mean-square deviations for all six structures, as well as the global root mean-square deviation.

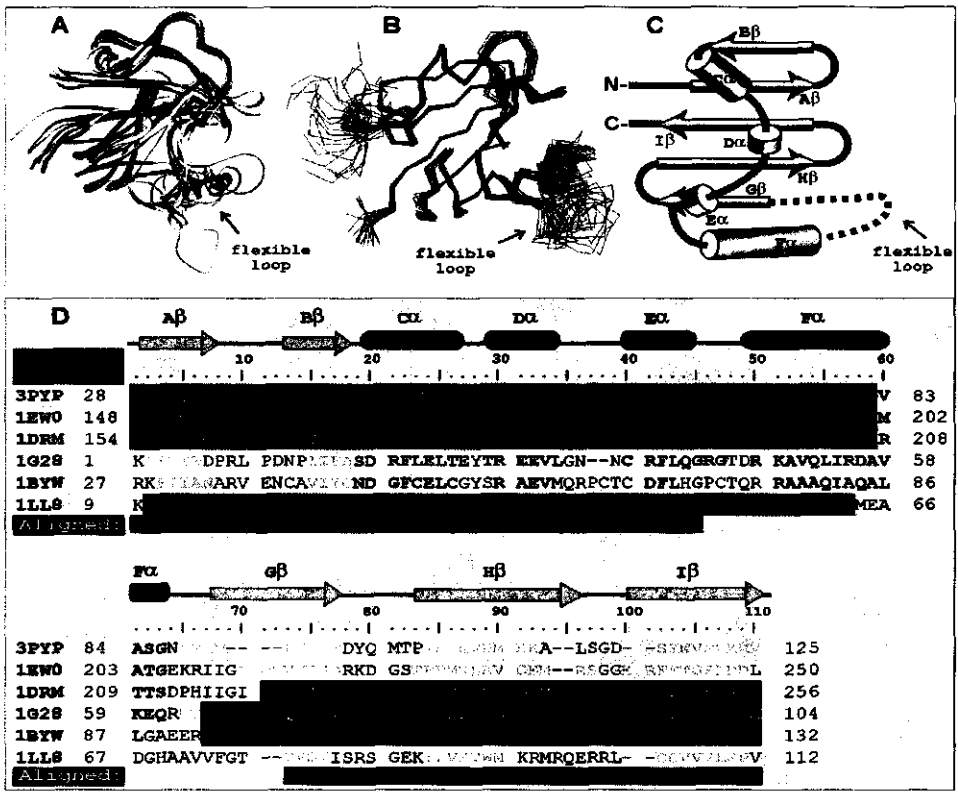


Figure 7.1 An overlay of the structural alignment of the six representative PAS structures selected is presented in panel **A**. The PFAM PAS annotated regions are coloured in blue, the PAC motif regions in orange/red. Structures and part of structures currently not assigned as either PAS or PAC are coloured in grey. Panel **B** shows the 20 lowest-energy solution structures of the human PAS kinase. In panel **C** a schematic representation of the human PAS kinase (according to[4]) is given. The flexible region between Fα and Gβ is clearly visible in **B**. This loop is located between the PAS domain and PAC motif. Panel **D** shows the structural alignment of the six structures selected. The PAS domains are indicated with blue bars, the PAC motifs with orange bars. The boxes on which the structural alignment is based are indicated in black. Helical and sheet region residues are coloured in red and yellow, respectively.

The 20 lowest-energy NMR solution structures of the human PAS kinase are shown in Figure 7.1B. The majority of the human PAS kinase structure was solved with high

precision, but portions of the Fα helix and the subsequent FG loop were poorly defined in this structural ensemble[4]. The Fα helix and the FG loop correspond to that region of the PAS fold, which tethers the PAS domain and PAC motif. A schematic representation of the human PAS kinase is depicted in Figure 7.1C.

The other PAS domain containing structures resemble a similar fold, in which the area corresponding to the Fα helix and the subsequent FG loop of human PAS kinase is believed to form specific interactions in the hydrophobic core or with bound cofactors. The FixL structures have elevated temperature factors in the FG loop region, indicating increased flexibility[97, 191]. The FG loop might be the key flexible region necessary for signal transduction[4].

According to the PFAM Protein Families Database[13], not all six structures contain both a PAS (PF00989) and a PAC motif (PF00785) (see Table 7.1). (in Figure 7.1D the PAS annotated domains are coloured with blue bars, and the PAC annotated domains with orange bars). It is obvious from the structural overlay in Figure 7.1A, that all six proteins share a common domain with a characteristic 5-stranded β-pleated α-helical structure. In comparing the structural and sequence alignments, it is clear that the subdivision of the domain into PAS and PAC motifs is incorrect, since their existence would imply that the conserved 5-stranded β-sheet is split into two sections. Therefore we propose to use the name PAS fold[226, 287] for the complete β-pleated α-helical structure, which defines PAS domains and C-terminal PAC motifs in terms of structure rather than sequence.

Large-Scale Modelling

The first, and most critical, step in protein homology modelling is the appropriate alignment of template and experimental sequences. The alignment of the six representative 3D structures (Figure 7.1A and 7.1D) provides the possibility to use all six structures as template for large-scale homology modelling. Note that not all six structures contain a PAS as well as a PAC motif, according to the PFAM database (Figure 7.1D and Table 7.1). Each of the 958 PAS domains was modelled against each of the 6 template structures presented in Figure 7.1. ProsaII Z-scores were sorted by template structure, resulting in good- and bad models. With an average sequence length of 105 amino acid residues, all models with a z-score higher than –3.57 (that is, closer to zero) were considered to be poor models[247], and were rejected. Thus, 30 % of the sequences used did not produce a good quality model. Of the resulting 672 best models, 188 were constructed using 1EW0 as template, and 177 were constructed using 1DRM. Only 2.2 % of the best models used 1LL8 as template. A pie-diagram of these results is depicted in Figure 7.2. Notably, 1EW0 and 1DRM were the best template structures, each in about 27 % of the cases. This might indicate that most PAS domain proteins would resemble a fold similar to FixL. A list of all PAS sequences modelled, as well as their best template structure, will be distributed on our website in near future.

Table 7.3 All sequences of the model organisms annotated in the PFAM PAS domain alignment. The presence of any adjacent PFAM PAC annotated domain is listed. For each sequence the template sequence with the best E-value is given, as well as the z-score of the best model before, and after realignment using Align-2D.

Name	PAS accession number	PFAM PAC*	PROSA z-score; best model	z-score after Align-2D; best model
Arabidopsis thaliana				
phytochrome A	P14712 632-737	n/a	-6.04 3PYP	-6.19 1DRM
phytochrome A	P14712 765-872	n/a	-2.02 3PYP	-3.17 1DRM
phytochrome B	P14713 676-772	n/a	-5.72 1G28	-6.04 3PYP
phytochrome B	P14713 800-904	n/a	-2.49 1DRM	-4.09 3PYP
phytochrome C	P14714 618-723	n/a	-5.96 3PYP	-5.32 3PYP
phytochrome C	P14714 751-859	n/a	-2.20 3PYP	-4.16 3PYP
phytochrome D	P42497 670-776	n/a	-5.94 1EW0	-5.29 3PYP
phytochrome D	P42497 804-908	n/a	-2.58 1G28	-3.57 3PYP
phytochrome E	P42498 609-718	n/a	-3.96 3PYP	-4.36 1DRM
phytochrome E	P42498 746-851	n/a	-1.28 3PYP	-4.57 3PYP
nonphototropic hypocotyl protein 1	O48963 201-300	PAC	-4.22 1G28	-6.10 1G28
nonphototropic hypocotyl protein 1	O48963 476-578	PAC	-5.03 1G28	-7.77 1G28
putative ser/thr kinase	O64511 38-141	PAC	-5.75 1BYW	-6.51 1G28
putative ser/thr kinase	O64511 260-364	PAC *	-4.08 1BYW	-6.23 1G28
nonphototropic hypocotyl protein 2	O81204 137-236	PAC	-4.29 1G28	-6.08 1G28
nonphototropic hypocotyl protein 2	O81204 390-492	PAC	-3.62 1DRM	-7.40 1G28
putative ser/thr kinase	O82754 102-198	PAC	-4.79 1EW0	-6.84 1EW0
putative protein kinase	Q9C547 76-172	PAC	-4.53 1EW0	-6.94 1EW0
putative protein kinase	Q9C833 76-172	PAC	-5.42 1EW0	-6.25 3PYP
putative protein kinase	Q9C902 115-211	PAC	-5.71 1EW0	-6.32 1BYW
putative protein kinase	Q9C903 76-172	PAC	-5.42 1EW0	-6.25 3PYP
hypothetical 82.2 kDa protein	Q9C9V5 113-209	PAC	-5.34 1EW0	-7.08 3PYP
protein kinase	Q9FGZ6 112-208	PAC	-4.35 1DRM	-7.49 1DRM
Escherichia coli				
hypothetical transcriptional regulator ygeV	Q46802 171-276	n/a	-4.20 1BYW	-2.86 3PYP
sensor protein atoS	Q06067 273-379	n/a	-2.95 1G28	-3.50 1EW0
sensor protein dcuS	P39272 233-339	B_19516	-4.33 1BYW	-1.72 1G28
hypothetical protein yegE	P38097 313-420	PAC	-4.14 1BYW	-6.73 1EW0
hypothetical protein yegE	P38097 566-671	PAC	-5.95 1EW0	-6.84 1BYW
hypothetical protein yciR	P77334 121-227	n/a	-4.67 1DRM	-3.25 1EW0
sensor kinase dpiB	P77510 233-341	B_39296	-3.78 1EW0	-4.00 1DRM
TraJ protein	P05837 52-158	B_39648	-4.21 1BYW	-3.17 1EW0
TraJ protein	P13949 32-138	B_39648	-4.55 1BYW	-3.58 3PYP
phosphate regulon sensor phoR	P08400 107-209	n/a	-3.91 1LL8	-2.71 1EW0
aerobic respiration control sensor arcB	P22763 164-270	n/a	-3.39 1EW0	-2.38 3PYP
hypothetical protein yddU	P76129 24-129	PAC	-7.58 1EW0	-7.69 1EW0
hypothetical protein yddU	P76129 146-254	PAC	-4.13 3PYP	-5.73 1BYW
glycerol metabolism operon regulator	P76016 214-318	n/a	-3.03 1EW0	-2.85 1DRM

Caenorhabditis elegans				
aryl hydrocarbon receptor nuclear translocator ortholog 1	O44711 128-235	n/a	-4.87 1G28	-4.35 3PYP
aryl hydrocarbon receptor nuclear translocator ortholog 1	O44711 288-394	B_66903	-4.13 3PYP	-4.83 1EW0
aryl hydrocarbon receptor ortholog 1	O44712 139-245	n/a	-6.19 1BYW	-4.47 1EW0
aryl hydrocarbon receptor ortholog 1	O44712 284-391	n/a	-2.83 1LL8	-3.09 1G28
F38A6.3B protein	Q9TVM0 200-306	n/a	-6.43 1EW0	-4.70 1LL8
F38A6.3B protein	Q9TVM0 349-445	PAC *	-4.10 3PYP	-3.88 3PYP
C25A1.11 protein	O02219 128-235	n/a	-4.87 1G28	-4.35 3PYP
C25A1.11 protein	O02219 290-396	B_66903	-4.13 3PYP	-4.83 1EW0
F38A6.3A protein	O45486 200-306	n/a	-6.43 1EW0	-4.70 1LL8
F38A6.3A protein	O45486 339-445	n/a	-5.26 3PYP	-3.88 3PYP
putative transcription factor C15C8.2	Q18018 163-271	n/a	-4.86 1G28	-3.46 1EW0
putative transcription factor C15C8.2	Q18018 304-410	PAC *	-3.52 3PYP	-1.87 3PYP
single-minded homolog T01D3.2	P90953 95-201	n/a	-3.70 1EW0	-4.79 1DRM
Azotobacter vinelandii				
nitrogen fixation regulator NifL	P30663 36-144	PAC	-2.96 1G28	-5.69 1G28
nitrogen fixation regulator NifL	P30663 162-268	n/a	-3.86 1EW0	-4.34 1DRM

* PFAM has the possibility to BLAST a sequence against their HMM search profile. Some sequences (*) are then annotated as PAC motif.

Some sequences are annotated as having a PFAM-B region (B_66903 or B_39648 or B_19516). PFAM-B regions contains a large number of small families that do not overlap with PFAM-A. Although of lower quality PFAM-B families can be useful when no PFAM-A families are found.

Arabidopsis, Escherichia, Caenorhabditis, and Azotobacter; a case study

Some of the PAS domains have been analysed in detail. We chose four representative organisms from the animal, bacterial and plant kingdoms, *Arabidopsis thaliana*, *Escherichia coli*, *Azotobacter vinelandii* and *Caenorhabditis elegans*, to analyse their complement of PAS domains. These species have been studied extensively and many details of their gene expression and function are known.

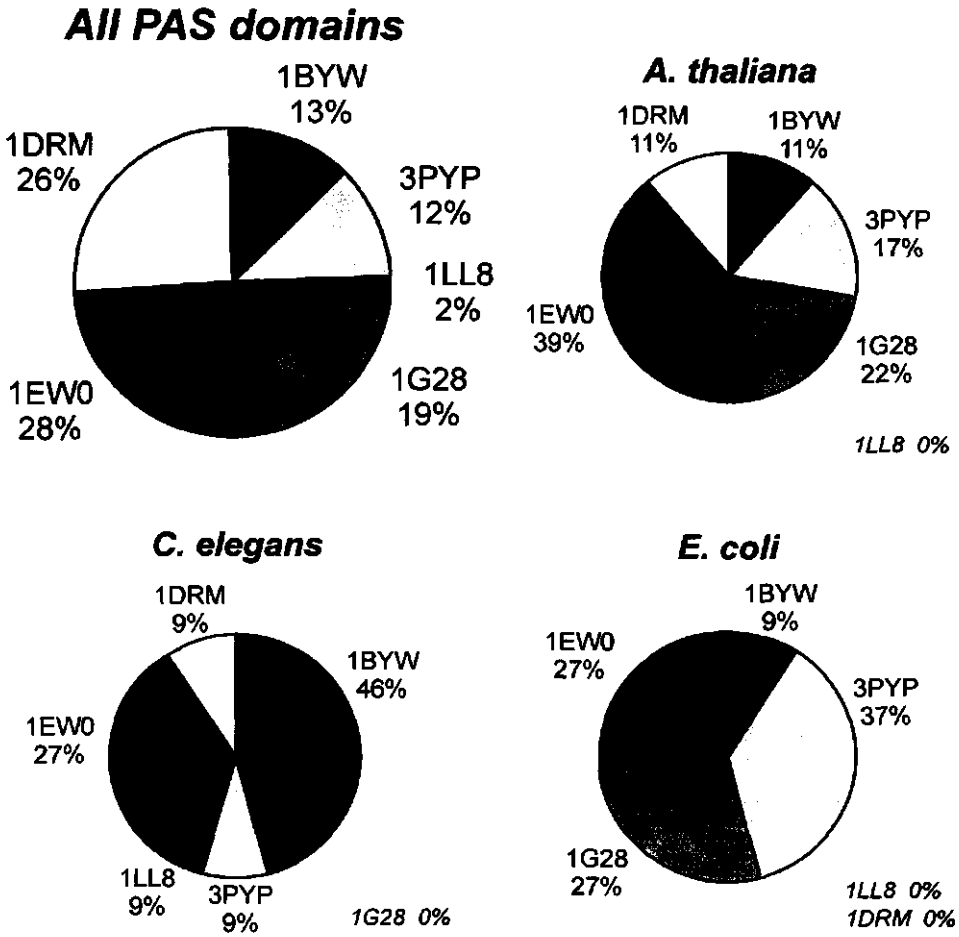


Figure 7.2 On the top-left a pie-diagram of the percentage best model, for each of the 672 best models, is presented. Of the 6 template structures used, 54 % of the sequences gives a best model with the FixL (1DRM and 1EW0) structures as template structure, while only a small percentage of the best models is created by using 1LL8 as template structure. The subsequent pie-diagrams of the percentage best model, for all PFAM PAS annotated *A. thaliana*, *C. elegans*, and *E. coli* sequences are shown. On average, for these three model organisms, 32 % of the sequences gives a best model with the 1EW0 as template structure, while only 3 % of the best models is created by using 1LL8 as template structure. Note that for the latter three, only a limited number of sequences is modelled.

The existing PFAM PAC annotation of sequences from these organisms is listed in Table 7.3. However, some sequences with a PAC motif are not annotated as having a PAS domain (Table 7.4). The full-length sequences of these proteins were aligned manually, and subsequently trimmed back to the region which we denote as representing the PAS fold. Alignment of this region from the *A. thaliana* sequences listed in Table 7.3

and Table 7.4, based upon the structural alignment (Figure 7.1D) of the six representative PAS proteins, is depicted in Figure 7.3. We conclude from this alignment that all PAS annotated *A. thaliana* proteins also contain a PAC motif, and conversely that all PAC annotated *A. thaliana* proteins contain a PAS domain. Therefore, in the case of *A. thaliana*, the PAS and PAC motifs are not separable, indicating that annotation of these proteins as containing only PAS or PAC motifs is questionable. A similar realignment was performed with the other three organisms, resulting in the same conclusion: PAS and PAC motifs do not occur independently of each other, but are parts of the same functional fold, separated by a linker region which is flexible in length. As all sequences of the four organisms studied showed inseparable PAC and PAS regions, the coexistence of PAS and PAC motifs might also apply to most other PAS and PAC protein sequences present in the PFAM database.

The sequences of these proteins were also re-aligned using the Align-2D command[2], in order to try to improve the manual alignment. Modelling based upon these alignments sometimes resulted in higher z-scores, and thus better models, as listed in Table 7.3. Indeed, some of the low-scoring models had a better z-score after realignment, resulting in more reliable models. This was specially the case for the *A. thaliana* phytochromes. The PFAM PAC motif annotated sequences, that do not have a PFAM PAS annotation, also gave reasonable z-scores after realignment (Table 7.4).

It is interesting to consider whether the best template for modelling a particular PAS domain is related to the co-factor which it contains. Unfortunately there are insufficient PAS domains characterised at the biochemical level to make any definitive correlation. The NifL PAS fold (amino acid residues 36-144) from *A. vinelandii* binds FAD as cofactor[116]. The best template was 1G28 (Table 7.3), a FMN binding PAS fold protein. The second PAS fold in this protein (amino acid residues 162-268) gives the best model when using the heme containing FixL X-ray structure 1DRM (Table 7.3). There is some indication that this domain indeed binds heme (R. Dixon, unpublished results).

PAC annotated sequences

Eight protein sequences from *A. thaliana*, *E. coli*, and *C. elegans* do not contain a PAS domain but only a PAC motif according to PFAM . All 8 sequences yielded reliable models, judged by their ProsaII z-scores (Table 7.4). For example, the *E. coli* Aerotaxis receptor (P50466) is described as containing a PAS domain by Ponting and coworkers[231, 332], although it is not annotated as such in the PFAM database. This protein has FAD as cofactor[21].

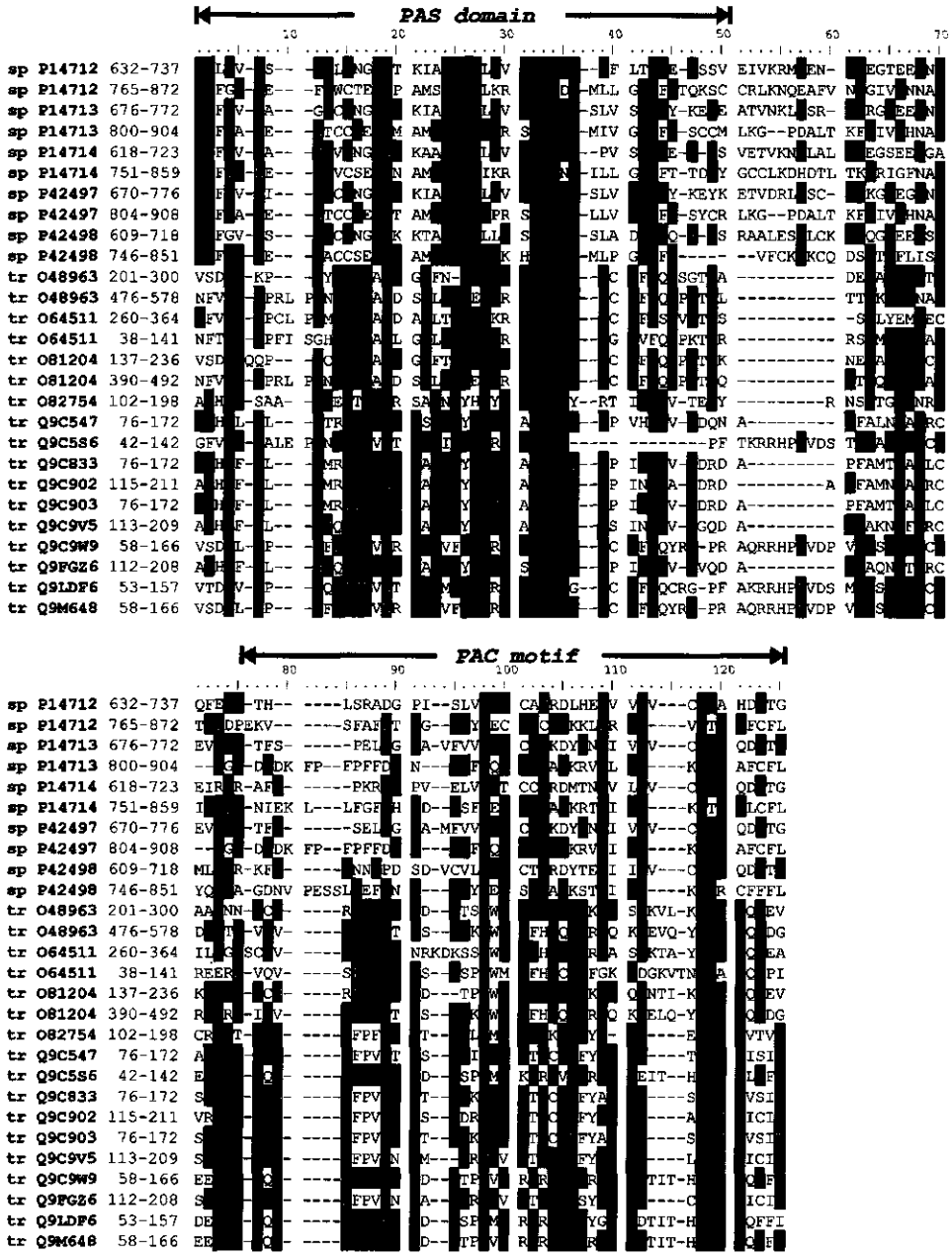


Figure 7.3 Alignment of all *A. thaliana* sequences that are either annotated as PFAM PAS domain or PFAM PAC motif. Amino acid residues having a sequence similarity > 35 % are depicted in black shading. In the left column the SWISS-PROT or TrEMBL accession numbers are listed, in the adjacent column the first and the last amino acid residue numbers. Above the sequence alignment the PAS and PAC annotated regions are indicated.

Table 7.4 Sequences that have a PFAM PAC annotation, but not a PFAM PAS annotation, were extended N-terminally to incorporate any available PAS domain. The N-terminal region of these sequences were aligned manually, and the sequences were subsequently modelled against the six template structures. Realignment with Align-2D of the *A. thaliana*, *E. coli*, and *C. elegans* sometimes resulted in better models.

Name	PAC accession number	PFAM PAS	PROSA z-score best model; after manual alignment	PROSA z-score best model; after Align-2D
Arabidopsis thaliana				
adagio 2	tr Q9C5S6 42-142	B_462	-5.36 3PYP	-6.30 1BYW
hypothetical 69.1 kDa protein	tr Q9C9W9 58-166	B_462	-5.44 1G28	-4.54 1G28
clock-associated PAS protein ztl	tr Q9LDF6 53-157	B_462	-4.96 1G28	-6.01 1G28
fkf1 (adagio 3)	tr Q9M648 58-166	B_462	-5.44 1G28	-4.54 1G28
Escherichia coli				
hypothetical protein yegE	P38097	B_45327	-3.82 1BYW	-4.30 3PYP
aerotaxis receptor	P50466	n/a	-5.72 1DRM	-6.65 1BYW
Caenorhabditis elegans				
hypothetical protein F16B3.1	O44164	B_462	-6.45 1BYW	-6.79 1BYW
EAG K channel EGL2	Q9XYX7	B_462	-6.45 1BYW	-6.79 1BYW

The two *C. elegans* sequences listed in Table 7.4 were derived from different strains, and differ only in one amino acid residue. This mutation is not in the PAS fold region, and therefore both protein sequences give the same result. Both 3D models were very reliable over the complete PAS fold sequence length. More examples of sequences that are (almost) identical are present in the PFAM PAS database (for instance the *C. elegans* sequences O02219 and O44711).

DISCUSSION

In the PFAM database there are amino acid sequences of almost 1000 PAS domains representative of all kingdoms of life. Of all PAS domain containing proteins present in the PDB database it can be concluded that the PAS and PAC motifs split the 5-stranded β-sheet into two sections. The PAS and PAC motifs are connected through a loop region, which was recently suggested to be important for the intrinsic function of PAS domain containing proteins. It is evident from the 3D structures present in the database, and from our large scale modelling studies presented here, that the PAS and PAC motif are inseparable and together give rise to a structural fold. In order to avoid confusion in protein annotation, it is important to define the sequence requirements for a given protein fold. We propose to define the complete β-pleated α-helical structure observed in the prototype structures of the PYP, FixL, human PAS kinase, HERG, and PHY3 proteins as the PAS fold. For comparison of proteins it is necessary to abandon the use of the commonly used annotations S1 / S2[332], PAS-A / PAS-B[48, 117], LOV domain[33, 44], and PAS domain / PAC motif[231] which are now in use to specify sequence similarities.

Unfortunately in recent years the meaning of the term "PAS domain" has evolved. We favour the use of the term "PAS fold" for referring to proteins sharing the PAS structural element rather than a sequence alignment.

For the large-scale homology studies, the existing PFAM PAS domain alignment was extended C-terminally by 50 amino acids in order to include the neighbouring PAC motif. Our studies show that sequence comparison alone is no longer sufficient to annotate newly discovered protein sequences as having a PAS domain, and we demonstrate that modelling studies give more insight into this intriguing family of sensory proteins, since 30 % of the PAS domains annotated in the PFAM database are unlikely to share the "PAS fold" as defined in this article. After re-alignment of PAS annotated protein sequences from four model organisms, some 3D models improved in quality, while others did not. The Align-2D command used for this realignment, could be of help in improving sequence alignments, but is not always successful. For the four organisms studied extensively, the drop-out percentage for bad models decreased significantly, from 21 % to 12 % (Figure 7.2). To date, three dimensional structures of six different PAS proteins have been elucidated. When more structures of PAS fold containing proteins will become available, it will be possible to redefine the PAS fold containing proteins into several subclasses, depending upon template structure or cofactor.

The PAS fold represents an important sensory domain present in all kingdoms of life[332], and in the PFAM database some proteins appear to have more than one PAS domain. Some proteins are known to contain two flavin cofactors[215, 314], or both a flavin and a heme[206, 249], though they do not contain a PAS fold. In the PFAM database there are currently four proteins each having five (partial) PAS annotated domains[125, 126, 212, 274]. The 20 PFAM PAS annotated regions have been modelled against the six representative PAS structures. This reveals that 50 % of the models have a poor z-score, and they are rejected as a reliable model. This percentage is significantly higher that the percentage rejected models for all PFAM PAS annotated sequences (30 %). A possible explanation is that of the 20 regions annotated as PAS, 14 are annotated as partial PAS domain. Furthermore, not all PAS regions are followed C-terminally by a PAC motif. Purification and subsequent structure analysis of these proteins might reveal the real number of PAS folds per protein.

All models of sequences of the four organisms used in the case study, which had a PFAM PAS domain annotation, had reliable z-scores, even if no PFAM PAC motif was present. Even more striking, all models of sequences with only a PFAM PAC motif annotation had good z-scores as well. This stresses the importance of better annotation of the PAS fold, based upon structural information rather than sequence information.

Annotation of protein sequences by PFAM (and SMART) is based upon sequence homology and HMM profiles. This annotation system is of great use for researchers, as for many protein sequences, a potential function based upon this annotation system can

be assigned. However, when proteins have only a limited sequence homology (as is the case for the PFAM PAC motifs), annotation of these motifs is difficult even using Hidden Markov Model. We show here that large scale homology modelling can be very useful in addition to HMM-based sequence annotation on the basis of a detailed analysis of all PAS and PAC sequences present in the PFAM database for *A. thaliana*, *E. coli*, *C. elegans* and *A. vinelandii*.

CHAPTER 8

Concluding Remarks

"All models are wrong, but some are useful"

Zeger, S. L. Statistical reasoning in epidemiology
American Journal of Epidemiology **134**, 1062-1066 (1991)

The main goal of this Ph.D. project was to obtain insight into the structure and function of the nitrogen fixation regulatory protein NifL from *Azotobacter vinelandii*. Nitrogen fixing bacteria, such as *A. vinelandii*, are capable to convert N_2 to NH_3. This conversion is essential for live, since animals and plants need NH_3 for their existence, and are unable to perform this conversion. The biosynthesis of NH_3 is controlled by 15 to 20 different nitrogen fixation (*nif*) gene products. The activation of *nif* gene expression by the transcriptional regulatory enhancer binding protein NifA is controlled by the sensor flavoprotein NifL in response to changes in oxygen level or the level of fixed nitrogen. The N-terminal domain of NifL contains two PAS domains. PAS domains are ubiquitous motifs present in archaea, bacteria and eucarya, and are named after the first letters of the Drosophila Period protein (PER), the Aryl hydrocarbon receptor nuclear translocator protein (ARNT) and the Drosophila Single-minded protein (SIM). For understanding the different mechanisms in which PAS domains work detailed information about their cofactor binding region is needed. The cofactor of the N-terminally located PAS domain from NifL is shown to be FAD. To tackle the fold and function of this region, it is of interest to deflavinate and reconstitute this domain.

A review article has been written (Chapter 2), summarising the advances made in the field of flavoprotein deflavination and reconstitution. Conventional precipitation methods, such as the acid ammonium sulfate procedure, are rapid but the yield and reconstitutability of apoprotein may vary dramatically. The more recently developed chromatographic procedures have the advantage that the apoprotein is stabilised by immobilisation, and that large amounts of apo- or reconstituted flavoproteins can be obtained. These chromatographic procedures can combine the use of affinity tags both for the purification of the protein of interest as well as replacement of the protein-bound cofactor. If the cofactor is covalently bound, protein engineering provides insight into the mechanism of covalent flavinylation. Site-directed mutagenesis can also be used to change the strength of the flavin-apoprotein interaction. As several diseases are associated with mutations in human flavoprotein genes, understanding the flavin-apoprotein interaction is also of medical relevance. An example of the use of affinity tags for the deflavination and reconstitution of the NifL PAS domain protein is given in Chapter 4.

Because imidazole can influence crystallographic trials, a general procedure for the purification of His-tagged proteins has been developed (Chapter 3). The high selectivity of IMAC column material towards histidine-tags, combined with on-column removal of the tag with thrombin results in a two-step purification with high overall yield. Moreover, this method has the advantage that instead of imidazole, thrombin is used in the final purification step. Removal of the His-tag results in a protein with no additional tags, except three additional amino acids at the N-terminal side. This structure resembles therefore the native form of the protein. It proved to be very successful for the NifL PAS domain, as well as for three other His-tagged proteins. The proof that imidazole can

prevent crystallisation of proteins, even when it is present in small amounts, is that in the case of the NifL PAS domain, the His-tagged protein never crystallised, though the non-tagged version did. This, of course, could also be caused by the removal of this very hydrophilic tag.

Although removal of the His-tag leads to protein free of imidazole traces, the tag itself can be used for the preparation and reconstitution of apoflavoproteins. A method is described in Chapter 4, in which the His-tag is used to prepare apoprotein or protein reconstituted with either 2,4a-^{13}C-FAD or 2,4a-^{13}C-FMN. The His-tagged NifL PAS domain holoprotein is bound to an immobilised metal affinity column, and the flavin is released by washing the column with buffer containing 2 M KBr and 2 M urea. The apoprotein is reconstituted on-column with the (artificial) flavin cofactor, and then eluted with buffer containing 250 mM imidazole. Alternatively, the immobilised apoprotein can be released from the column matrix before reconstitution. The advantage of this method is that it combines a protein affinity chromatography technique with limited protein loss. This results in a high protein yield with extremely efficient flavin reconstitution, and may develop as a universal method for replacement of flavin or other cofactors.

The non-tagged NifL PAS domain protein, purified as described in Chapter 3, is crystallised by the hanging-drop vapour-diffusion method. The crystals belong to the rhombohedral spacegroup *R*32, and native data were collected to 3.0 Å on the BW7B synchrotron beamline at the EMBL Hamburg Outstation. Initial attempts to crystallise Se-Met incorporated protein crystals were unsuccessful. Structure elucidation using molecular replacement strategies seems to be successful, and preliminary data of the partially elucidated structure is presented in Chapter 5. It is questionable as whether this model can be refined to a satisfactory R-factor, as this is an ongoing project.

The tagged version of the NifL PAS domain protein was examined by synchrotron small-angle X-ray solution scattering (Chapter 6). The experimental scattering data leads to a global fold of the tetrameric domain at low-resolution. Fitting a model of the PAS domain to this dummy atom model leads to a better understanding of the dimerisation / tetramerisation interface. The resulting low resolution structure of the NifL PAS domain shows a well developed, anisometric shape of four monomers attached to each other. This tetrameric form is in agreement with analytical gel filtration experiments which also conform a homo-tetrameric state of this domain. Small angle X-ray scattering measurements of full-length NifL protein might reveal that the tetramerization domain is in the PAS domain region.

One of the problems with the current annotation of the PAS domain is that this domain exhibits limited homology at the sequence level. From 3D X-ray and NMR structures available to date, however, it is clear that these domains have a conserved three-dimensional fold as shown by structural alignment of six representative 3D structures from the PDB database (Chapter 7). Large-scale modelling of the PAS sequences present in the PFAM database against the 3D structures of these six structural

prototypes was performed. The existing subdivision into PAS and PAC motifs, as in use by the PFAM and SMART databases, appears to be linked to major differences in sequences in the loop region linking these two motifs. This loop region is very flexible and adopts different conformations depending on the ligand bound. We propose that this loop region and the adjacent secondary elements are the main structural determinants in the biological response of PAS fold proteins. Finally, some sequences did not produce a good structural model, even after realignment using Align-2D, suggesting that these representatives are unlikely to have a fold resembling that of any of the structural prototypes of the PAS domain superfamily.

Curriculum vitae

Marco Hendrikus Hefti was born in Rotterdam, The Netherlands, on September 26th 1972, as the youngest son of Herman Hefti and Ceciel Hefti-Ten Broeke. In 1992, he graduated from grammar school at the Mencia de Mendoza Lyceum in Breda. Subsequently, he started to study Molecular Sciences at the Wageningen Agricultural University (WAU, nowadays Wageningen University). For his Master's degree he did 6 months research at the department of Organic Chemistry (dr. Sies van der Kerk[†] and Prof. Ernst Sudhölter) and 6 months at the department of Biochemistry (dr. Dorus Gadella). Thereafter he spent 5 months at the department of Molecular Biology at the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany (dr. Roland Brock and Prof. Thomas Jovin). In September 1997 he graduated, and started to work as a practical assistant for a first-year course at the department of Physical and Colloid chemistry (Frans Geurts).

In October 1997, he started his Ph.D. research at the department of Biochemistry at the Wageningen University. His supervisors were dr. Jacques Vervoort and Prof. Colja Laane; later dr. Jacques Vervoort and Prof. Sacco de Vries. This research with the title *'TASTES: Transcriptional Activation and Sensory Transduction, Elucidation of Structures'* was performed within the EU-Biotechnology framework. Part of his work were performed during his six months stay as a Marie Curie Fellow at the EMBL Outstation in Hamburg (Germany), where he had an on-the-spot training in protein crystallography by dr. Paul Tucker. Other parts of his work was performed at the Nitrogen Fixation Laboratory of the John Innes Centre in Norwich (UK), under supervision of Prof. Ray Dixon and dr. David Lawson. The research that was carried out during this Ph.D. period is described in this thesis. From March 2003 he works for the NVON, the Dutch organisation for scientific teachers and technical assistants working at secondary schools.

List of Publications

Vervoort, J. and M. H. Hefti
"NMR of flavoproteins." Flavoprotein Protocols. (S. K. Chapman and G. A. Reid, eds.) Humana Press Inc. **131:** 139-147 (1999).

Hefti, M. H., C. J. G. Van Vugt - Van der Toorn, R. Dixon and J. Vervoort
"A novel purification method for histidine-tagged proteins containing a thrombin cleavage site." Analytical Biochemistry **295**(2): 180-185 (2001).

Hefti, M. H., J. Hendle, C. Enroth, J. Vervoort and P. A. Tucker
"Crystallization and preliminary crystallographic data of the PAS domain of the NifL protein from *Azotobacter vinelandii*." Acta Crystallographica Section D Biological Crystallography **57**: 1895-1896 (2001).

Hefti, M. H., F. J. Milder, S. Boeren, J. Vervoort and W. J. H. van Berkel
"Preparation and reconstitution of apoflavoproteins using a His-tag based immobilization method." in: Flavins and Flavoproteins (S. K. Chapman, R. Perham and N. S. Scrutton, eds.) Berlin : 969-973 (2002).

Hefti, M. H., F. J. Milder, S. Boeren, J. Vervoort and W. J. H. van Berkel
"A His-tag based immobilization method for the preparation and reconstitution of apoflavoproteins." Biochimica et Biophysica Acta **1619**(2): 139-143 (2003).

Samenvatting

Dit proefschrift is het resultaat van een onderzoeksproject uitgevoerd in Europees verband. Een van de belangrijkste doelen van dit project was om inzicht te krijgen in de structuur en de functie van het eiwit genaamd NifL. Dit eiwit, dat stikstof-fixatie reguleert, komt van nature voor in *Azotobacter vinelandii*. *A. vinelandii* is een bacterie die in de bodem leeft, en deze bacterie heeft het vermogen om het gas stikstof (N_2) om te zetten in ammoniak (NH_3). Deze omzetting is van essentieel belang voor het leven op aarde, omdat dieren en planten ammoniak nodig hebben voor hun bestaan, maar hogere organismen zijn niet in staat zijn om stikstof naar ammoniak om te zetten. De biosynthese van ammoniak, uitgevoerd door *A. vinelandii*, wordt gereguleerd door 15 tot 20 verschillende stikstof fixatie (*nif*) gen-producten. De activering van deze *nif* gen expressie wordt verzorgd door het eiwit NifA. NifL, op zijn beurt, reguleert NifA. NifL is een sensor eiwit dat een flavine cofactor bevat.

Het N-terminale deel van NifL bevat twee PAS domeinen. PAS domeinen zijn veelvoorkomende eiwit motieven, en komen voor in bacteria, eukaryoten en archea. De naam PAS is afgeleid van de eerste letters van het *Period* eiwit, het *Aryl hydrocarbon receptor nuclear translocator* eiwit, en het *Single minded* eiwit. Om meer te leren over de verschillende werkingsmechanismen van PAS domeinen, is het nodig om meer inzicht te krijgen in dat deel van deze eiwitten dat betrokken is bij de interactie met cofactoren. De cofactor van het N-terminale PAS domein in NifL is FAD. Om de vouwing en de functie van dit deel van het eiwit te achterhalen, is het interessant om deze flavine cofactor te verwijderen en te vervangen door analoge cofactoren.

In Hoofdstuk 2 wordt een overzicht gegeven op het gebied van het verwijderen en vervangen van de cofactoren in flavine bevattende eiwitten. Conventionele precipitatie methoden, zoals bijvoorbeeld de zure ammonium sulfaat methode, zijn erg snel. Helaas valt de opbrengst vaak tegen, de reconstitutie van het apo-eiwit met een andere cofactor lukt niet altijd. Verder wordt in dit hoofdstuk ingegaan op modernere technieken. Deze technieken hebben het voordeel dat het apo-eiwit wordt gestabiliseerd doordat het geïmmobiliseerd is, zodat grote hoeveelheden apo- of gereconstitueerd eiwit verkregen wordt. Deze chromatografische methoden combineren het gebruik van een affiniteits staart zowel voor de zuivering van het gewenste eiwit alsook voor de vervanging van de eiwit-gebonden cofactor. Wanneer deze cofactor covalent gebonden is aan het eiwit, dan kan met behulp van *protein engineering* (het veranderen van de aminozuur sequentie van een eiwit door middel van mutaties) inzicht worden verkregen in dit bindingsmechanisme. *Site-directed mutagenesis* (het aanbrengen van mutaties in de eiwitsequentie) wordt ook wel gebruikt om de bindingssterkte tussen de cofactor en het eiwit te variëren. Sommige humane ziektes worden veroorzaakt door veranderingen in de aminozuur sequentie van flavine eiwitten. Meer inzicht in eiwit - cofactor interacties

is dus ook medisch gezien interessant. Een voorbeeld van het gebruik van deze affiniteits staartjes voor het verwijderen en vervangen van flavine cofactoren in eiwitten wordt gegeven in Hoofdstuk 4, hier wordt het NifL PAS domein eiwit gebruikt.

Imidazool kan een effect hebben op kristallisatie pogingen. Om dit te voorkomen is een algemene methode ontwikkeld voor de zuivering van eiwitten met een histidine staartje (Hoofdstuk 3). IMAC kolommen hebben een hoge specificiteit ten aanzien van histidine staartjes, waardoor deze kolommen uitermate geschikt zijn voor de zuivering van eiwitten met zo'n staart. Een efficiënte tweestaps zuiveringsmethode kan worden verkregen door met behulp van het enzym thrombine deze staart te verwijderen van het eiwit, terwijl deze nog vast zit aan de kolom. Behalve een grotere efficiëntie heeft deze methode nog een groot voordeel: in de laatste stap wordt geen imidazool meer gebruikt, maar thrombine. Hierdoor wordt bovendien ook nog de toegevoegde histidine staart verwijderd, zodat het eiwit meer lijkt op het oorspronkelijke, natuurlijke, eiwit. Deze methode is succesvol getest op 4 verschillende eiwitten, waaronder het NifL PAS domein eiwit. Zoals aangegeven kan imidazool, zelfs in kleine hoeveelheden, een verstorend effect hebben op de kristallisatie van eiwitten. In het geval van het NifL PAS domein bijvoorbeeld, kristalliseerde de vorm waarin de histidine staart aanwezig was niet, maar de staartloze versie wel. Dit kan natuurlijk ook komen doordat deze hydrofiele staart niet meer aanwezig was.

Alhoewel het verwijderen van de histidine staart leidt tot eiwit dat vrij is van imidazool sporen, kan de staart zelf ook voordelen hebben voor het verwijderen en vervangen van de flavine cofactor. In Hoofdstuk 4 is een methode beschreven waarbij de histidine staart wordt gebruikt voor het verwijderen van de FAD cofactor, en voor het verkrijgen van FAD of FMN wat met een koolstof-isotoop is gelabeld (2,4a-^{13}C-FAD en 2,4a-^{13}C-FMN). Nadat het NifL PAS domein met daaraan de histidine staart is gebonden aan een geïmmobiliseerde metaal affiniteits kolom, kan de flavine cofactor worden verwijderd door de kolom te wassen met een buffer waarin 2 molair kaliumbromide en 2 molair urea is opgelost. Het zo ontstane apo-eiwit kan, terwijl het aan de kolom gebonden is, worden gereconstitueerd met een artificiële flavine cofactor. Hierna kan dit gereconstitueerde eiwit van de kolom kan worden gehaald door te wassen met een buffer waarin 250 millimolair imidazool is opgelost. In plaats van reconstitutie, kan het apo-eiwit ook van de kolom verwijderd worden. Het voordeel van deze eiwit affiniteits chromatografie techniek is dat er weinig eiwit verloren raakt. Het resultaat is een erg hoge eiwit opbrengst, gecombineerd met een erg efficiënte flavine reconstitutie. Deze techniek zou zich dan ook kunnen ontwikkelen tot een universele methode voor het vervangen van flavine of andere cofactoren.

Het staartloze NifL PAS domein, zoals gezuiverd in Hoofdstuk 3, kan kristalliseren met behulp van de *hanging-drop vapour diffusion* methode. Bij deze methode hangt het eiwit als een druppeltje aan een glasplaatje boven een vloeistofoplossing. Door middel van verdamping treedt een evenwichtssituatie op, waarbij de concentratie eiwit zover

toeneemt dat deze uiteindelijk neer zal slaan in de vorm van een kristal. De zo ontstane kristallen behoren tot de rhombohedrische kristalgroep *R*32, en natieve data tot een resolutie van 3 Ångstrom werden gemeten met behulp van de BW7B synchrotron deeltjesversneller die staat bij het EMBL in Hamburg. Pogingen om het eiwit te labelen met selenomethionine en vervolgens te kristalliseren mislukte, maar de structuuropheldering door middel van de *molecular replacement* aanpak lijkt succesvol. Preliminaire resultaten van de gedeeltelijke structuuropheldering worden getoond in Hoofdstuk 5. Dit is een nog lopend project, en op dit moment is nog niet duidelijk of deze gedeeltelijke structuur ook zal leiden tot een uiteindelijke kristalstructuur.

De vorm van het eiwit waaraan de histidine staart nog gebonden was is bestudeerd met behulp van synchrotron *small-angle X-ray solution scattering* (Hoofdstuk 6). Synchrotron röntgen straling wordt verstrooid door eiwit in oplossing, en hieruit kan een globale vouwing van het tetramere eiwit worden berekend. Door een eiwit model te fitten tegen dit lage-resolutie model, kan inzicht worden verkregen over hoe de dimerisatie en tetramerisatie vlakken eruit zien. Het resultaat is een lage-resolutie structuur van het NifL PAS domein als een tetrameer, en dit tetramere model bevestigt analytische gel filtratie experimenten waaruit ook een tetramere opbouw van het domein bleek. Als men *Small-angle X-ray* verstrooiings metingen zou verrichten met het hele NifL eiwit, dan zou kunnen blijken dat het PAS domein de oorzaak is van tetramerisatie van het hele eiwit.

Een van de problemen die ontstaan door de huidige annotatie van PAS domeinen is dat dit domein een beperkte sequentie homologie heeft. Door driedimensionale kristalstructuren en NMR structuren te bestuderen van 6 eiwitten, blijkt dat deze eiwitten een geconserveerde driedimensionale vouwing hebben (Hoofdstuk 7). Alle PAS geannoteerde eiwit sequenties uit de PFAM database zijn gemodelleerd tegen deze 6 structurele PAS prototypes. De huidige onderverdeling in PAS en PAC motieven, zoals die ook in gebruik is in de PFAM en SMART databases, lijkt gecorreleerd te zijn met belangrijke verschillen in de lus-regio die deze twee motieven verbindt. Dit flexibele stuk eiwit kan zich op verschillende manieren vouwen, afhankelijk van het ligand dat gebonden is. We stellen dan ook voor zowel dit flexibele deel, alsook de twee aanliggende motieven, de naam 'PAS fold' te geven, aangezien ze gezamenlijk een belangrijk structureel element vormt die betrokken is bij de biologische interacties zoals PAS eiwitten die kunnen aangaan. Tot slot wordt in dit hoofdstuk nog ingegaan op een aantal eiwitten die niet goed gemodelleerd kunnen worden, en dus mogelijk geen structureel element vormen, en dus ook geen onderdeel uitmaken van de PAS superfamilie.

References

1. Alex, L. A. & Simon, M. I. **(1994)** Protein histidine kinases and signal transduction in prokaryotes and eukaryotes, *Trends in Genetics.* **10**, 133-138.
2. Šali, A. & Blundell, T. L. **(1993)** Comparative protein modelling by satisfaction of spatial restraints, *Journal of Molecular Biology.* **234**, 779-815.
3. Aliverti, A., Bruns, C. M., Pandini, V. E., Karplus, P. A., Vanoni, M. A., Curti, B. & Zanetti, G. **(1995)** Involvement of serine 96 in the catalytic mechanism of ferredoxin-NADP$^+$ reductase: structure-function relationship as studied by site-directed mutagenesis and X-ray crystallography, *Biochemistry.* **34**, 8371-8379.
4. Amezcua, C. A., Harper, S. M., Rutter, J. & Gardner, K. H. **(2002)** Structure and Interactions of PAS Kinase N-Terminal PAS Domain. Model for Intramolecular Kinase Regulation, *Structure.* **10**, 1349-1361.
5. Ashton, R. & Ashton, P. **(1985)** *Handbook of Reptiles and Amphibians of Florida*, Windward Publishing, inc., Miami, Florida.
6. Austin, S., Buck, M., Cannon, W., Eydmann, T. & Dixon, R. **(1994)** Purification and in vitro activities of the native nitrogen fixation control proteins NIFA and NIFL, *Journal of Bacteriology.* **176**, 3460-3465.
7. Bader, B., Knecht, W., Fries, M. & Löffler, M. **(1998)** Expression, purification, and characterization of histidine-tagged rat and human flavoenzyme dihydroorotate dehydrogenase, *Protein Expression and Purification.* **13**, 414-422.
8. Bairoch, A. & Apweiler, R. **(2000)** The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, *Nucleic Acids Research.* **28**, 45-48.
9. Bantscheff, M., Weiss, V. & Glocker, M. O. **(1999)** Identification of linker regions and domain borders of the transcription activator protein NtrC from *Escherichia coli* by limited proteolysis, in-gel digestion, and mass spectrometry, *Biochemistry.* **38**, 11012-11020.
10. Banzoli, U. & Massey, V. **(1974)** Preparation of aldehyde oxidase in its native and deflavo forms. Comparison of spectroscopic and catalytic properties, *Journal of Biological Chemistry.* **249**, 4339-4345.
11. Barman, B. G. & Tollin, G. **(1972)** Flavin-protein interactions in flavoenzymes. Thermodynamics and kinetics of reduction of *Azotobacter* flavodoxin, *Biochemistry.* **11**, 4755-4759.
12. Barquera, B., Hase, C. C. & Gennis, R. B. **(2001)** Expression and mutagenesis of the NqrC subunit of the NQR respiratory Na(+) pump from *Vibrio cholerae* with covalently attached FMN, *FEBS Letters.* **492**, 45-49.
13. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. L. **(2002)** The Pfam protein families database, *Nucleic Acids Research.* **30**, 276-280.
14. Becvar, J. & Palmer, G. **(1982)** The binding of flavin derivatives to the riboflavin-binding protein of egg white. A kinetic and thermodynamic study, *Journal of Biological Chemistry.* **257**, 5607-5617.
15. Beinert, W. D., Rüterjans, H. & Muller, F. **(1985)** Nuclear magnetic resonance studies of the old yellow enzyme. 1. ^{15}N NMR of the enzyme recombined with ^{15}N-labeled flavin mononucleotides, *European Journal of Biochemistry.* **152**, 573-579.
16. Benen, J., van Berkel, W. J. H., Dieteren, N., Arscott, D., Williams, C., Jr., Veeger, C. & de Kok, A. **(1992)** Lipoamide dehydrogenase from *Azotobacter vinelandii*: site-directed mutagenesis of the His450-Glu455 diad. Kinetics of wild-type and mutated enzymes, *European Journal of Biochemistry.* **207**, 487-497.
17. Benen, J., van Berkel, W. J. H., Veeger, C. & de Kok, A. **(1992)** Lipoamide dehydrogenase from *Azotobacter vinelandii*. The role of the C-terminus in catalysis and dimer stabilization, *European Journal of Biochemistry.* **207**, 499-505.
18. Benen, J. A., Sanchez-Torres, P., Wagemaker, M. J., Fraaije, M. W., van Berkel, W. J. H. & Visser, J. **(1998)** Molecular cloning, sequencing, and heterologous expression of the vaoA gene from *Penicillium simplicissimum* CBS 170.90 encoding vanillyl-alcohol oxidase, *Journal of Biological Chemistry.* **273**, 7865-7872.
19. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. **(2000)** The Protein Data Bank, *Nucleic Acids Research.* **28**, 235-242.
20. Bewley, M. C., Marohnic, C. C. & Barber, M. J. **(2001)** The structure and biochemistry of NADH-dependent cytochrome b5 reductase are now consistent, *Biochemistry.* **40**, 13574-13582.
21. Bibikov, S. I., Biran, R., Rudd, K. E. & Parkinson, J. S. **(1997)** A signal transducer for aerotaxis in *Escherichia coli*, *Journal of Bacteriology.* **179**, 4075-4079.
22. Biemann, M., Claiborne, A., Ghisla, S., Massey, V. & Hemmerich, P. **(1983)** Oxidation of 2-thioflavins by peroxides, *Journal of Biological Chemistry.* **258**, 5440-5448.
23. Binda, C., Mattevi, A. & Edmondson, D. E. **(2002)** Structure-function relationships in flavoenzyme-dependent amine oxidations: a comparison of polyamine oxidase and monoamine oxidase, *Journal of Biological Chemistry.* **277**, 23973-23976.
24. Binda, C., Newton-Vinson, P., Hubalek, F., Edmondson, D. E. & Mattevi, A. **(2002)** Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders, *Nature Structural Biology.* **9**, 22-26.
25. Blanco, G., Drummond, M., Woodley, P. & Kennedy, C. **(1993)** Sequence and molecular analysis of the *nifL* gene of *Azotobacter vinelandii*, *Molecular Microbiology.* **9**, 869-879.
26. Bonants, P. J., Müller, F., Vervoort, J. & Edmondson, D. E. **(1990)** A ^{31}P-nuclear-magnetic-resonance study of NADPH-cytochrome-P-450 reductase and of the *Azotobacter* flavodoxin/ferredoxin-NADP$^+$ reductase complex, *European Journal of Biochemistry.* **190**, 531-537.
27. Borgstahl, G. E. O., Williams, D. R. & Getzoff, E. D. **(1995)** 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: Unusual fold, active site, and chromophore, *Biochemistry.* **34**, 6278-6287.
28. Bossi, R. T., Negri, A., Tedeschi, G. & Mattevi, A. **(2002)** Structure of FAD-bound L-aspartate oxidase: insight into substrate specificity and catalysis, *Biochemistry.* **41**, 3018-3024.
29. Boulin, C. J., Kempf, R., Gabriel, A. & Koch, M. H. J. **(1986)** Data appraisal, evaluation and display for synchrotron radiation experiments: hardware and software, *Nuclear Instruments and Methods in Physics Research Section A.* **249**, 399-407.
30. Boulin, C. J., Kempf, R., Gabriel, A. & Koch, M. H. J. **(1988)** Data acquisition systems for linear and area X-

ray detectors using delay line readout, *Nuclear Instruments and Methods in Physics Research Section A.* **269**, 312-320.
31. Bradford, M. M. (**1976**) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Analytical Biochemistry.* **72**, 248-254.
32. Brady, A. H. & Beychok, S. (**1969**) Optical activity and conformational studies of pig heart lipoamide dehydrogenase, *Journal of Biological Chemistry.* **244**, 4634-4637.
33. Briggs, W. R., Christie, J. M. & Salomon, M. (**2001**) Phototropins: a new family of flavin-binding blue light receptors in plants, *Antioxidants & Redox Signaling.* **3**, 775-788.
34. Brock, B. J. & Gold, M. H. (**1996**) 1,4-Benzoquinone reductase from basidiomycete *Phanerochaete chrysosporium*: spectral and kinetic analysis, *Archives of Biochemistry and Biophysics.* **331**, 31-40.
35. Brudler, R., Meyer, T. E., Genick, U. K., Devanathan, S., Woo, T. T., Millar, D. P., Gerwert, K., Cusanovich, M. A., Tollin, G. & Getzoff, E. D. (**2000**) Coupling of hydrogen bonding to chromophore conformation and function in photoactive yellow protein, *Biochemistry.* **39**, 13478-13486.
36. Burgess, B. K. & Lowe, D. J. (**1996**) Mechanism of molybdenum nitrogenase, *Chemical Reviews.* **96**, 2983-3012.
37. Carlson, R. & Langerman, N. (**1984**) The thermodynamics of flavin binding to the apoflavodoxin from *Azotobacter vinelandii*, *Archives of Biochemistry and Biophysics.* **229**, 440-447.
38. Casalin, P., Pollegioni, L., Curti, B. & Pilone Simonetta, M. (**1991**) A study on apoenzyme from *Rhodotorula gracilis* D-amino acid oxidase, *European Journal of Biochemistry.* **197**, 513-517.
39. Chakraborty, S. & Massey, V. (**2002**) Reaction of reduced flavins and flavoproteins with diphenyliodonium chloride, *Journal of Biological Chemistry.* **277**, 41507-41516.
40. Champier, L., Sibille, N., Bersch, B., Brutscher, B., Blackledge, M. & Coves, J. (**2002**) Reactivity, secondary structure, and molecular topology of the *Escherichia coli* sulfite reductase flavodoxin-like domain, *Biochemistry.* **41**, 3770-3780.
41. Chang, J. (**1985**) Thrombin specificity; requirement for apolar amino acids adjecent to the thrombin cleavage site of polypeptide substrate, *European Journal of Biochemistry.* **151**, 217-224.
42. Chen, Z., Karaplis, A. C., Ackerman, S. L., Pogribny, I. P., Melnyk, S., Lussier-Cacan, S., Chen, M. F., Pai, A., John, S. W., Smith, R. S., Bottiglieri, T., Bagley, P., Selhub, J., Rudnicki, M. A., James, S. J. & Rozen, R. (**2001**) Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased methylation capacity, with neuropathology and aortic lipid deposition, *Human Molecular Genetics.* **10**, 433-443.
43. Choong, Y. S., Shepherd, M. G. & Sullivan, P. A. (**1975**) Preparation of the lactate oxidase apoenzyme and studies on the binding of flavin mononucleotide to the apoenzyme, *Biochemical Journal.* **145**, 37-45.
44. Christie, J. M., Salomon, M., Nozue, K., Wada, M. & Briggs, W. R. (**1999**) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (*nph1*): binding sites for the chromophore flavin mononucleotide, *Proceedings of the National Academy of Sciences of the United States of America.* **96**, 8779-8783.
45. Christie, J. M., Swartz, T. E., Bogomolni, R. A. & Briggs, W. R. (**2002**) Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function, *The Plant Journal.* **32**, 205-219.
46. Claiborne, A., Massey, V., Fitzpatrick, P. F. & Schopfer, L. M. (**1982**) 2-Thioflavins as active site probes of flavoproteins, *Journal of Biological Chemistry.* **257**, 174-182.
47. Conant, R. & Collins, J. (**1998**) *A field Guide to Reptiles and Amphibians*, Houghton Mifflin Company, New York.
48. Crews, S. T., Thomas, J. B. & Goodman, C. S. (**1988**) The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product, *Cell.* **52**, 143-151.
49. Crosson, S. & Moffat, K. (**2001**) Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 2995-3000.
50. Crosson, S. & Moffat, K. (**2002**) Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch, *Plant Cell.* **14**, 1067-1075.
51. Dabrowski, S. & Kur, J. (**1999**) Cloning, overexpression, and purification of the recombinant His-tagged SSB protein of *Escherichia coli* and use in polymerase chain reaction amplification, *Protein Expression and Purification.* **16**, 96-102.
52. Daugas, E., Nochy, D., Ravagnan, L., Loeffler, M., Susin, S. A., Zamzami, N. & Kroemer, G. (**2000**) Apoptosis-inducing factor (AIF): a ubiquitous mitochondrial oxidoreductase involved in apoptosis, *FEBS Letters.* **476**, 118-123.
53. de Kok, A. & van Berkel, W. J. H. (**1996**) Lipoamide dehydrogenase in *Alpha-Keto Acid Dehydrogenase Complexes* (Patel, M. S., Roche, T. E. & Harris, R. A., eds) pp. 53-70, Birkhäuser Verlag, Basel.
54. Dean, D. R. & Jacobson, M. R. (**1992**) Biochemical Genetics of Nitrogenase in *Biological Nitrogen Fixation* (Stacey, G., Burris, R. H. & Evans, H. J., eds) pp. 763-834, Chapman & Hall, New York.
55. Dean, P. D. & Watson, D. H. (**1979**) Protein purification using immobilised triazine dyes, *Journal of Chromatography.* **165**, 301-319.
56. Dekker, J., Eppink, M. H. M., van Zwieten, R., de Rijk, T., Remacha, A. F., Law, L. K., Li, A. M., Cheung, K. L., van Berkel, W. J. H. & Roos, D. (**2001**) Seven new mutations in the nicotinamide adenine dinucleotide reduced-cytochrome b(5) reductase gene leading to methemoglobinemia type I, *Blood.* **97**, 1106-1114.
57. Dixon, R. (**1998**) The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in gamma-proteobacteria, *Archives of Microbiology.* **169**, 371-380.
58. Dixon, R., Eydmann, T., Henderson, N. & Austin, S. (**1991**) Substitutions at a single amino acid residue in the nitrogen-regulated activator protein NTRC differentially influence its activity in response to phosphorylation, *Molecular Microbiology.* **5**, 1657-1668.
59. Dobbek, H., Gremer, L., Meyer, O. & Huber, R. (**1999**) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine, *Proceedings of the National Academy of Sciences of the United States of America.* **96**, 8884-8889.
60. Dobbek, H., Svetlitchnyi, V., Gremer, L., Huber, R. & Meyer, O. (**2001**) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster, *Science.* **293**, 1281-1285.
61. Duex, P., Rubinstenn, G., Vuister, G. W., Boelens, R., Mulder, F. A. A., Hard, K., Hoff, W. D., Kroon, A. R.,

Crielaard, W., Hellingwerf, K. J. & Kaptein, R. (**1998**) Solution structure and backbone dynamics of the photoactive yellow protein, *Biochemistry.* **37**, 12689-12699.

62. Dym, O. & Eisenberg, D. (**2001**) Sequence-structure analysis of FAD-containing proteins, *Protein Science.* **10**, 1712-1728.
63. Edmondson, D. E. & Tollin, G. (**1971**) Flavin-protein interactions and the redox properties of the Shethna flavoprotein, *Biochemistry.* **10**, 133-145.
64. Edy, V. G., Billiau, A. & De Somer, P. (**1977**) Purification of human Fibroblast Interferon by zinc chelate affinity chromatography, *Journal of Biological Chemistry.* **252**, 5934-5935.
65. Efimov, I., Cronin, C. N. & McIntire, W. S. (**2001**) Effects of noncovalent and covalent FAD binding on the redox and catalytic properties of *p*-cresol methylhydroxylase, *Biochemistry.* **40**, 2155-2166.
66. Eggink, G., Engel, H., Vriend, G., Terpstra, P. & Witholt, B. (**1990**) Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints, *Journal of Molecular Biology.* **212**, 135-142.
67. Ellman, G. L. (**1959**) Tissue sulfhydryl groups, *Archives of Biochemistry and Biophysics.* **82**, 70-77.
68. Entsch, B. & van Berkel, W. J. H. (**1995**) Structure and mechanism of para-hydroxybenzoate hydroxylase, *Faseb Journal.* **9**, 476-483.
69. Eppink, M. H. M., Cammaart, E., van Wassenaar, D., Middelhoven, W. J. & van Berkel, W. J. H. (**2000**) Purification and properties of hydroquinone hydroxylase, a FAD-dependent monooxygenase involved in the catabolism of 4-hydroxybenzoate in *Candida parapsilosis* CBS604, *European Journal of Biochemistry.* **267**, 6832-6840.
70. Eppink, M. H. M., Schreuder, H. A. & van Berkel, W. J. H. (**1997**) Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding, *Protein Science.* **6**, 2454-2458.
71. Eschenbrenner, M., Coves, J. & Fontecave, M. (**1995**) The flavin reductase activity of the flavoprotein component of sulfite reductase from *Escherichia coli*. A new model for the protein structure, *Journal of Biological Chemistry.* **270**, 20550-20555.
72. Eschenbrenner, M., Coves, J. & Fontecave, M. (**1995**) NADPH-sulfite reductase flavoprotein from *Escherichia coli*: contribution to the flavin content and subunit interaction, *FEBS Letters.* **374**, 82-84.
73. Eschrich, K., van Berkel, W. J. H., Westphal, A. H., de Kok, A., Mattevi, A., Obmolova, G., Kalk, K. H. & Hol, W. G. (**1990**) Engineering of microheterogeneity-resistant *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*, *FEBS Letters.* **277**, 197-199.
74. Eydmann, T., Söderbäck, E., Jones, T., Hill, S., Austin, S. & Dixon, R. (**1995**) Transcriptional activation of the nitrogenase promoter in vitro: Adenosine nucleotides are required for inhibition of NIFA activity by NIFL, *Journal of Bacteriology.* **177**, 1186-1195.
75. Farrell, H. M., Jr., Mallette, M. F., Buss, E. G. & Clagett, C. O. (**1969**) The nature of the biochemical lesion in avian renal riboflavinuria. 3. The isolation and characterization of the riboflavin-binding protein from egg albumin, *Biochimica et Biophysica Acta.* **194**, 433-442.
76. Feigin, L. A. & Svergun, D. I. (**1987**) *Structure analysis by small-angle X-ray and neutron scattering*, Plenum, New York.
77. Fiser, A., Do, R. K. & Sali, A. (**2000**) Modeling of loops in protein structures, *Protein Science.* **9**, 1753-1773.
78. Fitzpatrick, P. F., Ghisla, S. & Massey, V. (**1985**) 8-Azidoflavins as photoaffinity labels for flavoproteins, *Journal of Biological Chemistry.* **260**, 8483-8491.
79. Fleischmann, G., Lederer, F., Müller, F., Bacher, A. & Rüterjans, H. (**2000**) Flavin-protein interactions in flavocytochrome b2 as studied by NMR after reconstitution of the enzyme with ^{13}C- and ^{15}N-labelled flavin, *European Journal of Biochemistry.* **267**, 5156-5167.
80. Fosterhartnett, D. & Kranz, R. G. (**1992**) Analysis of the promotors and upstream sequences of *nifA*1 and *nifA*2 in *Rhodobacter capsulatus* - activation requires NtrC but not RpnoN, *Molecular Microbiology.* **6**, 1049-1060.
81. Fraaije, M. W., Kamerbeek, N. M., van Berkel, W. J. H. & Janssen, D. B. (**2002**) Identification of a Baeyer-Villiger monooxygenase sequence motif, *FEBS Letters.* **518**, 43-47.
82. Fraaije, M. W. & Mattevi, A. (**2000**) Flavoenzymes: diverse catalysts with recurrent features, *Trends in Biochemical Sciences.* **25**, 126-132.
83. Fraaije, M. W., van Berkel, W. J. H., Benen, J. A., Visser, J. & Mattevi, A. (**1998**) A novel oxidoreductase family sharing a conserved FAD-binding domain, *Trends in Biochemical Sciences.* **23**, 206-207.
84. Fraaije, M. W., van den Heuvel, R. H. H., van Berkel, W. J. H. & Mattevi, A. (**1999**) Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase, *Journal of Biological Chemistry.* **274**, 35514-35520.
85. Fraaije, M. W., van den Heuvel, R. H. H., van Berkel, W. J. H. & Mattevi, A. (**2000**) Structural analysis of flavinylation in vanillyl-alcohol oxidase, *Journal of Biological Chemistry.* **275**, 38654-38658.
86. Franken, H. D., Rüterjans, H. & Müller, F. (**1984**) Nuclear-magnetic-resonance investigation of ^{15}N-labeled flavins, free and bound to *Megasphaera elsdenii* apoflavodoxin, *European Journal of Biochemistry.* **138**, 481-489.
87. Frosst, P., Blom, H. J., Milos, R., Goyette, P., Sheppard, C. A., Matthews, R. G., Boers, G. J., den Heijer, M., Kluijtmans, L. A., van den Heuvel, L. P. & et al. (**1995**) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase, *Nature Genetics.* **10**, 111-113.
88. Fujii, K., Galivan, J. H. & Huennekens, F. M. (**1977**) Activation of methionine synthase: further characterization of flavoprotein system, *Archives of Biochemistry and Biophysics.* **178**, 662-70.
89. Gabriel, A. & Dauvergne, F. (**1982**) The localization method used at EMBL, *Nuclear Instruments and Methods in Physics Research.* **201**, 223-224.
90. Gadda, G. & Fitzpatrick, P. F. (**1998**) Biochemical and physical characterization of the active FAD-containing form of nitroalkane oxidase from *Fusarium oxysporum*, *Biochemistry.* **37**, 6154-6164.
91. Gatti, D. L., Palfey, B. A., Lah, M. S., Entsch, B., Massey, V., Ballou, D. P. & Ludwig, M. L. (**1994**) The mobile flavin of 4-OH benzoate hydroxylase, *Science.* **266**, 110-114.
92. Genick, U. K., Borgstahl, G. E. O., Ng, K., Ren, Z., Pradervand, C., Burke, P. M., Srajer, V., Teng, T. Y., Schildkamp, W., McRee, D. E., Moffat, K. & Getzoff, E. D. (**1997**) Structure of a protein photocycle intermediate by millisecond time-resolved crystallography, *Science.* **275**, 1471-1475.
93. Genick, U. K., Soltis, S. M., Kuhn, P., Canestrelli, I. L. & Getzoff, E. D. (**1998**) Structure at 0.85 Å resolution of an early protein phytocycle intermediate, *Nature.*

392, 206-209.
94. Ghisla, S. & Massey, V. (**1986**) New flavins for old: artificial flavins as active site probes of flavoproteins, *Biochemical Journal.* **239**, 1-12.
95. Gill, S. C. & Von Hippel, P. H. (**1989**) Calculation of protein extinction coefficients from amino acid sequence data, *Analytical Biochemistry.* **182**, 319-326.
96. Gong, W., Hao, B. & Chan, M. K. (**2000**) New mechanistic insights from structural studies of the oxygen-sensing domain of *Bradyrhizobium japonicum* FixL, *Biochemistry.* **39**, 3955-3962.
97. Gong, W., Hao, B., Mansy, S. S., Gonzalez, G., Gilles, G. M. A. & Chan, M. K. (**1998**) Structure of a biological sensor: A new mechanism for heme-driven signal transduction, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 15177-15182.
98. Gorbunoff, M. J. (**1984**) The interaction of proteins with hydroxyapatite, *Analytical Biochemistry.* **136**, 425-445.
99. Goyette, P., Summer, J. S., Milos, R., Duncan, A. M., Rosenblatt, D. S., Matthews, R. G. & Rozen, R. (**1994**) Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification, *Nature Genetics.* **7**, 195-200.
100. Gremer, L., Kellner, S., Dobbek, H., Huber, R. & Meyer, O. (**2000**) Binding of flavin adenine dinucleotide to molybdenum-containing carbon monoxide dehydrogenase from *Oligotropha carboxidovorans*. Structural and functional analysis of a carbon monoxide dehydrogenase species in which the native flavoprotein has been replaced by its recombinant counterpart produced in *Escherichia coli*, *Journal of Biological Chemistry.* **275**, 1864-1872.
101. Guenther, B. D., Sheppard, C. A., Tran, P., Rozen, R., Matthews, R. G. & Ludwig, M. L. (**1999**) The structure and properties of methylenetetrahydrofolate reductase from *Escherichia coli* suggest how folate ameliorates human hyperhomocysteinemia, *Nature Structural Biology.* **6**, 359-365.
102. Guinier, A. (**1939**) La diffraction des rayons X aux très petits angles, application à l'étude de phénomènes ultramicroscopiques, *Annual Physique.* **12**, 161-237.
103. Haas, E. (**1938**) Isolierung eines neuen gelben Ferments, *Biochemische Zeitschrift.* **298**, 378-390.
104. Habeeb, A. F. S. A. (**1972**) Reaction of protein sulfhydryl groups with Ellman's reagent, *Methods in Enzymology.* **25**, 457-464.
105. Haines, D. C., Sevrioukova, I. F. & Peterson, J. A. (**2000**) The FMN-binding domain of cytochrome P450BM-3: resolution, reconstitution, and flavin analogue substitution, *Biochemistry.* **39**, 9419-9429.
106. Haley, E. E. & Lambooy, J. P. (**1954**) Synthesis of D-riboflavin-2-C^{14} and its metabolism by *Lactobacillus casei*, *Journal of the American Chemical Society.* **76**, 2926-2929.
107. Hall, D. A., Jordan-Starck, T. C., Loo, R. O., Ludwig, M. L. & Matthews, R. G. (**2000**) Interaction of flavodoxin with cobalamin-dependent methionine synthase, *Biochemistry.* **39**, 10711-10719.
108. Hall, D. A., Vander Kooi, C. W., Stasik, C. N., Stevens, S. Y., Zuiderweg, E. R. & Matthews, R. G. (**2001**) Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 9521-9526.
109. Hayashi, M., Nakayama, Y. & Unemoto, T. (**2001**) Recent progress in the Na(+)-translocating NADH-quinone reductase from the marine *Vibrio alginolyticus*, *Biochimica et Biophysica Acta.* **1505**, 37-44.
110. Hayashi, M., Nakayama, Y., Yasui, M., Maeda, M., Furuishi, K. & Unemoto, T. (**2001**) FMN is covalently attached to a threonine residue in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from *Vibrio alginolyticus*, *FEBS Letters.* **488**, 5-8.
111. Hefti, M. H., Hendle, J., Enroth, C., Vervoort, J. & Tucker, P. A. (**2001**) Crystallization and preliminary crystallographic data of the PAS domain of the NifL protein from *Azotobacter vinelandii*, *Acta Crystallographica Section D Biological Crystallography.* **57**, 1895-1896.
112. Hefti, M. H., Milder, F. J., Boeren, S., Vervoort, J. & van Berkel, W. J. H. (**2002**) Preparation and reconstitution of apoflavoproteins using a His-tag based immobilization method in *Flavins and Flavoproteins* (Chapman, S. K., Perham, R. & Scrutton, N. S., eds) pp. 969-973, Rudolf Weber, Berlin.
113. Hefti, M. H., Milder, F. J., Boeren, S., Vervoort, J. & van Berkel, W. J. H. (**2003**) A His-tag based immobilization method for the preparation and reconstitution of apoflavoproteins, *Biochimica et Biophysica Acta.* **1619**, 139-143.
114. Hefti, M. H., Van Vugt - Van der Toorn, C. J. G., Dixon, R. & Vervoort, J. (**2001**) A novel purification method for histidine-tagged proteins containing a thrombin cleavage site, *Analytical Biochemistry.* **295**, 180-185.
115. Hill, S. (**1992**) Physiology of Nitrogen Fixation in free-living Heterotrophs in *Biological Nitrogen Fixation* (Stacey, G., Burris, R. H. & Evans, H. J., eds) pp. 87-134, Chapman & Hall, New York.
116. Hill, S., Austin, S., Eydmann, T., Jones, T. & Dixon, R. (**1996**) *Azotobacter vinelandii* NIFL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch, *Proceedings of the National Academy of Sciences of the United States of America.* **93**, 2143-2148.
117. Hoffman, E. C., Reyes, H., Chu, F. F., Sander, F., Conley, L. H., Brooks, B. A. & Hankinson, O. (**1991**) Cloning of a factor required for activity of the Ah (dioxin) receptor, *Science.* **252**, 954-958.
118. Howard, J. B. & Rees, D. C. (**1996**) Structural basis of biological nitrogen fixation, *Chemical Reviews.* **96**, 2965-2982.
119. Husain, M. & Massey, V. (**1978**) Reversible resolution of flavoproteins into apoproteins and free flavins, *Methods in Enzymology.* **53**, 429-437.
120. Jancarik, J. & Kim, S. H. (**1991**) Sparse Matrix Sampling: a screening method for crystallization of proteins, *Journal of Applied Crystallography.* **24**, 409-411.
121. Jiang, B. H., Rue, E., Wang, G. L., Roe, R. & Semenza, G. L. (**1996**) Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1, *Journal of Biological Chemistry.* **271**, 17771-17778.
122. Joza, N., Susin, S. A., Daugas, E., Stanford, W. L., Cho, S. K., Li, C. Y., Sasaki, T., Elia, A. J., Cheng, H. Y., Ravagnan, L., Ferri, K. F., Zamzami, N., Wakeham, A., Hakem, R., Yoshida, H., Kong, Y. Y., Mak, T. W., Zuniga-Pflucker, J. C., Kroemer, G. & Penninger, J. M. (**2001**) Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death, *Nature.* **410**, 549-554.
123. Kalse, J. F. & Veeger, C. (**1968**) Relation between conformations and activities of lipoamide dehydrogenase. I. Relation between diaphorase and lipoamide dehydrogenase activities upon binding of FAD by the apoenzyme, *Biochimica et Biophysica Acta.* **159**, 244-256.
124. Kanda, M., Brady, F. O., Rajagopalan, K. V. &

Handler, P. **(1972)** Studies of the dissociation of flavin adenine dinucleotide from metalloflavoproteins, *Journal of Biological Chemistry.* **247**, 765-770.

125. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. & Tabata, S. **(2000)** Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti, DNA Research.* 7, 331-338.
126. Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. **(1995)** Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome, *DNA Research.* **2**, 153-166.
127. Kasahara, M., Swartz, T. E., Olney, M. A., Onodera, A., Mochizuki, N., Fukuzawa, H., Asamizu, E., Tabata, S., Kanegae, H., Takano, M., Christie, J. M., Nagatani, A. & Briggs, W. R. **(2002)** Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from *Arabidopsis*, rice, and *Chlamydomonas reinhardtii*, *Plant Physiology.* **129**, 762-773.
128. Kay, S. A. **(1997)** PAS, present, and future: Clues to the origins of circadian clocks, *Science.* **276**, 753-754.
129. Kearney, E. B., Salach, J. I., Walker, W. H., Seng, R. L., Kenney, W., Zeszotek, E. & Singer, T. P. **(1971)** The covalently-bound flavin of hepatic monoamine oxidase. 1. Isolation and sequence of a flavin peptide and evidence for binding at the 8α position, *European Journal of Biochemistry.* **24**, 321-327.
130. Kerby, R. L., Hong, S. S., Ensign, S. A., Coppoc, L. J., Ludden, P. W. & Roberts, G. P. **(1992)** Genetic and physiological characterization of the *Rhodospirillum rubrum* carbon monoxide dehydrogenase system, *Journal of Bacteriology.* **174**, 5284-94.
131. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. **(1993)** Folding and stability of a tryptophan-containing mutant of ubiquitin, *Biochemistry.* **32**, 7054-7063.
132. Kim, J. J., Wang, M. & Paschke, R. **(1993)** Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate, *Proceedings of the National Academy of Sciences of the United States of America.* **90**, 7523-7.
133. Kirkpatrick, S., Gelatt, C. D. J. & Vecci, M. P. **(1983)** Optimization by simmulated annealing, *Science.* **220**, 671-680.
134. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. **(1999)** Rapid automated molecular replacement by evolutionary search, *Acta Crystallographica Section D Biological Crystallography.* **55**, 484-491.
135. Koch, M. H. J. & Bordas, J. **(1983)** X-ray diffraction and scattering on disordered systems using synchrotron radiation, *Nuclear Instruments and Methods in Physics Research.* **203**, 461-469.
136. Komai, H., Massey, V. & Palmer, G. **(1969)** The preparation and properties of deflavo xanthine oxidase, *Journal of Biological Chemistry.* **224**, 1692-1700.
137. Kozin, M. B., Volkov, V. V. & Svergun, D. I. **(1997)** ASSA, a program for the three-dimensional rendering in solution scattering from biopolymers, *Journal of Applied Crystallography.* **30**, 811-815.
138. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. **(2002)** Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine, *Nature.* **416**, 455-459.
139. Laane, C., de Roo, G., van den Ban, E. C. D., Sjauw-En-Wa, M. W., Duyvis, M. G., Hagen, W. R., van Berkel, W. J. H., Hilhorst, R., Schmedding, D. J. M. & Evans, D. J. **(1999)** The role of riboflavin in beer flavour instability : EPR studies and the application of flavin binding proteins, *Journal of the Institute of Brewing.* **105**, 392-397.
140. Laemmli, U. K. **(1970)** Cleavage of structural proteins during the assembly of the head of Bacteriophage T4, *Nature.* **227**, 680-685.
141. Lanfermeijer, F. C., Venema, K. & Palmgren, M. G. **(1998)** Purification of a histidine-tagged plant plasma membrane H^+-ATPase expressed in yeast, *Protein Expression and Purification.* **12**, 29-37.
142. Leclerc, D., Wilson, A., Dumas, R., Gafuik, C., Song, D., Watkins, D., Heng, H. H., Rommens, J. M., Scherer, S. W., Rosenblatt, D. S. & Gravel, R. A. **(1998)** Cloning and mapping of a cDNA for methionine synthase reductase, a flavoprotein defective in patients with homocystinuria, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 3059-3064.
143. Lederer, F., Rüterjans, H. & Fleischmann, G. **(1999)** Flavoprotein resolution and reconstitution in *Flavoprotein Protocols* (Chapman, S. K. & Reid, G. A., eds) pp. 149-155, Humana Press Inc.
144. Lee, H. S., Narberhaus, F. & Kustu, S. **(1993)** In vitro activity of NifL, a signal transduction protein for biological nitrogen fixation, *Journal of Bacteriology.* **175**, 7683-7688.
145. Lei, S., Pulakat, L. & Gavini, N. **(1999)** Genetic analysis of nif regulatory genes by utilizing the yeast two-hybrid system detected formation of a NifL-NifA complex that is implicated in regulated expression of *nif* genes, *Journal of Bacteriology.* **181**, 6535-6539.
146. Léonil, J., Langrené, S., Sicsic, S. & Le Goffic, F. **(1985)** Purification of D-amino acid oxidase apoenzyme by affinity chromatography on Cibacron Blue Sepharose, *Journal of Chromatography.* **347**, 316-319.
147. Leroux, A., Junien, C., Kaplan, J. & Bamberger, J. **(1975)** Generalised deficiency of cytochrome b5 reductase in congenital methaemoglobinaemia with mental retardation, *Nature.* **258**, 619-620.
148. Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R. R., Ponting, C. P. & Bork, P. **(2002)** Recent improvements to the SMART domain-based sequence annotation resource, *Nucleic Acids Research.* **30**, 242-244.
149. Leusen, J. H., Meischl, C., Eppink, M. H. M., Hilarius, P. M., de Boer, M., Weening, R. S., Ahlin, A., Sanders, L., Goldblatt, D., Skopczynska, H., Bernatowska, E., Palmblad, J., Verhoeven, A. J., van Berkel, W. J. H. & Roos, D. **(2000)** Four novel mutations in the gene encoding gp91-phox of human NADPH oxidase: consequences for oxidase assembly, *Blood.* **95**, 666-673.
150. Li, Z. & Crooke, E. **(1999)** Functional analysis of affinity-purified polyhistidine-tagged DnaA protein, *Protein Expression and Purification.* **17**, 41-48.
151. Lin, L. J. & Foster, J. F. **(1975)** Agarose-DTNB column for binding sulfhydryl-containing proteins and peptides, *Analytical Biochemistry.* **63**, 485-490.
152. Little, R., Colombo, V., Leech, A. & Dixon, R. **(2002)** Direct interaction of the NifL regulatory protein with the GlnK signal transducer enables the *Azotobacter vinelandii* NifL-NifA regulatory system to respond to conditions replete for nitrogen, *Journal of Biological Chemistry.* **277**, 15472-15481.
153. Little, R., Reyes-Ramirez, F., Zhang, Y., van Heeswijk, W. C. & Dixon, R. **(2000)** Signal transduction to the

Azotobacter vinelandii NIFL-NIFA regulatory system is influenced directly by interaction with 2-oxoglutarate and the PII regulatory protein, *Embo Journal.* **19**, 6041-6050.

154. Lostao, A., El Harrous, M., Daoudi, F., Romero, A., Parody-Morreale, A. & Sancho, J. (**2000**) Dissecting the energetics of the apoflavodoxin-FMN complex, *Journal of Biological Chemistry.* **275**, 9518-9526.
155. Macheroux, P., Hill, S., Austin, S., Eydmann, T., Jones, T., Kim, S. O., Poole, R. & Dixon, R. (**1998**) Electron donation to the flavoprotein NifL, a redox-sensing transcriptional regulator, *Biochemical Journal.* **332**, 413-419.
156. Maeda, M., Hamada, D., Hoshino, M., Onda, Y., Hase, T. & Goto, Y. (**2002**) Partially folded structure of flavin adenine dinucleotide-depleted ferredoxin-NADP$^+$ reductase with residual NADP$^+$ binding domain, *Journal of Biological Chemistry.* **277**, 17101-17107.
157. Manstein, D. J., Massey, V., Ghisla, S. & Pai, E. F. (**1988**) Stereochemistry and accessibility of prosthetic groups in flavoproteins, *Biochemistry.* **27**, 2300-2305.
158. Manstein, D. J. & Pai, E. F. (**1986**) Purification and characterization of FAD synthetase from *Brevibacterium ammoniagenes*, *Journal of Biological Chemistry.* **261**, 16169-16173.
159. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (**2000**) Comparative protein structure modeling of genes and genomes, *Annual Review of Biophysics and Biomolecular Structure.* **29**, 291-325.
160. Massey, V. (**1990**) A simple method for the determination of redox potentials in *Flavins and Flavoproteins* (Curti, B., Ronchi, S. & Zanetti, G., eds) pp. 859-866, Walter de Gruyter & Co, Berlin.
161. Massey, V. (**2000**) Chemical and biological versatility of riboflavin, *Biochemical Society Transactions.* **28**, 283-296.
162. Massey, V., Claiborne, A., Biemann, M. & Ghisla, S. (**1984**) 4-Thioflavins as active site probes of flavoproteins, *Journal of Biological Chemistry.* **259**, 9667-9678.
163. Massey, V. & Curti, B. (**1966**) A new method of preparation of D-amino acid oxidase apoprotein and a conformational change after its recombination with flavin adenine dinucleotide, *Journal of Biological Chemistry.* **241**, 3417-3423.
164. Massey, V., Ghisla, S. & Moore, E. G. (**1979**) 8-Mercaptoflavins as active site probes of flavoenzymes, *Journal of Biological Chemistry.* **254**, 9640-9650.
165. Massey, V., Ghisla, S. & Yagi, K. (**1986**) 6-Azido- and 6-aminoflavins as active-site probes of flavin enzymes, *Biochemistry.* **25**, 8095-8102.
166. Massey, V., Ghisla, S. & Yagi, K. (**1986**) 6-Thiocyanatoflavins and 6-mercaptoflavins as active-site probes of flavoproteins, *Biochemistry.* **25**, 8103-8112.
167. Massey, V. & Harris, C. M. (**1997**) Milk xanthine oxidoreductase: the first one hundred years, *Biochemical Society Transactions.* **25**, 750-755.
168. Massey, V. & Hemmerich, P. (**1980**) Active-site probes of flavoproteins, *Biochemical Society Transactions.* **8**, 246-257.
169. Masters, B. S. S. (**2000**) Structural variations to accommodate functional themes of the isoforms of nitric oxide synthases in *Nitric Oxide* (Ignarro, L., ed) pp. 91-104, Academic Press, San Diego.
170. Mate, M. J., Ortiz-Lombardia, M., Boitel, B., Haouz, A., Tello, D., Susin, S. A., Penninger, J., Kroemer, G. & Alzari, P. M. (**2002**) The crystal structure of the mouse apoptosis-inducing factor AIF, *Nature Structural Biology.* **9**, 442-446.
171. Mateo, P. L. & Sturtevant, J. M. (**1977**) Thermodynamics of the binding of flavin adenine dinucleotide to D-amino acid oxidase, *Biosystems.* **8**, 247-253.
172. Mathews, F. S. (**1991**) New flavoenzymes, *Current Opinion in Structural Biology.* **1**, 954-967.
173. Matsui, K., Sugimoto, K. & Kasai, S. (**1982**) Thermodynamics of association of egg yolk riboflavin binding protein with 8-substituted riboflavins. Comparison with the egg white protein, *Journal of Biochemistry (Tokyo).* **91**, 1357-1362.
174. Mattevi, A. (**1998**) The PHBH fold: not only flavoenzymes, *Biophysical Chemistry.* **70**, 217-222.
175. Mattevi, A., Fraaije, M. W., Mozzarelli, A., Olivi, L., Coda, A. & van Berkel, W. J. H. (**1997**) Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity, *Structure.* **5**, 907-920.
176. Mattevi, A., Obmolova, G., Kalk, K. H., van Berkel, W. J. H. & Hol, W. G. (**1993**) Three-dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens* at 2.8 Å resolution. Analysis of redox and thermostability properties, *Journal of Molecular Biology.* **230**, 1200-1215.
177. Mattevi, A., Tedeschi, G., Bacchella, L., Coda, A., Negri, A. & Ronchi, S. (**1999**) Structure of L-aspartate oxidase: implications for the succinate dehydrogenase/fumarate reductase oxidoreductase family, *Structure with Folding & Design.* **7**, 745-756.
178. Matthews, B. (**1968**) *Journal of Molecular Biology.* **33**, 491-497.
179. Mattison, K., Oropeza, R. & Kenney, L. J. (**2002**) The linker region plays an important role in the interdomain communication of the response regulator OmpR, *Journal of Biological Chemistry.* **277**, 32714-32721.
180. Mayer, E. J. & Thorpe, C. (**1981**) A method for resolution of general acyl-coenzyme A dehydrogenase apoprotein, *Analytical Biochemistry.* **116**, 227-229.
181. Mayhew, S. G. (**1971**) Studies on flavin binding in flavodoxins, *Biochimica et Biophysica Acta.* **235**, 289-302.
182. Mayhew, S. G. & Ludwig, M. L. (**1975**) Flavodoxins and electron-transferring flavoproteins, *Enzymes.* **12**, 57-118.
183. McClelland, D. A. & Price, N. C. (**1998**) Stopped-flow analysis of the refolding of hen egg white riboflavin binding protein in its native and dephosphorylated forms, *Biochimica et Biophysica Acta.* **1382**, 157-166.
184. Merrick, M., Hill, S., Hennecke, H., Hahn, M., Dixon, R. & Kennedy, C. (**1982**) Repressor properties of the *nifL* gene product of *Klebsiella pneumoniae*, *Molecular & General Genetics.* **185**, 75-81.
185. Mewies, M., McIntire, W. S. & Scrutton, N. S. (**1998**) Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: The current state of affairs, *Protein Science.* **7**, 7-20.
186. Miller, M. S., Benore-Parsons, M. & White, H. B., 3rd. (**1982**) Dephosphorylation of chicken riboflavin-binding protein and phosvitin decreases their uptake by oocytes, *Journal of Biological Chemistry.* **257**, 6818-6824.
187. Miura, R. & Miyake, Y. (**1987**) ^{13}C-NMR studies of porcine kidney D-amino acid oxidase reconstituted with ^{13}C-enriched flavin adenine dinucleotide. Effects of competitive inhibitors, *Journal of Biochemistry (Tokyo).* **101**, 581-589.
188. Miura, R. & Miyake, Y. (**1987**) ^{13}C-NMR studies on the reaction intermediates of porcine kidney D-amino acid oxidase reconstituted with ^{13}C-enriched flavin adenine dinucleotide, *Journal of Biochemistry (Tokyo).* **102**, 1345-1354.

189. Miyano, M., Fukui, K., Watanabe, F., Takahashi, S., Tada, M., Kanashiro, M. & Miyake, Y. **(1991)** Studies on Phe-228 and Leu-307 recombinant mutants of porcine kidney D- amino acid oxidase: expression, purification, and characterization, *Journal of Biochemistry (Tokyo).* **109**, 171-177.
190. Miyatake, H., Kanai, M., Adachi, S. i., Nakamura, H., Tamura, K., Tanida, H., Tsuchiya, T., Iizuka, T. & Shiro, Y. **(1999)** Dynamic light-scattering and preliminary crystallographic studies of the sensor domain of the haem-based oxygen sensor FixL from *Rhizobium meliloti*, *Acta Crystallographica Section D Biological Crystallography.* **55**, 1215-1218.
191. Miyatake, H., Mukai, M., Park, S., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T. & Shiro, Y. **(2000)** Sensory Mechanism of Oxygen Sensor FixL from *Rhizobium meliloti*: Crystallographic, Mutagenesis and Resonance Raman Spectroscopic Studies, *Journal of Molecular Biology.* **301**, 415-431.
192. Monaco, H. L. **(1997)** Crystal structure of chicken riboflavin-binding protein, *Embo Journal.* **16**, 1475-1483.
193. Money, T., Barrett, J., Dixon, R. & Austin, S. **(2001)** Protein-protein interactions in the complex between the enhancer binding protein NIFA and the sensor NIFL from *Azotobacter vinelandii*, *Journal of Bacteriology.* **183**, 1359-1368.
194. Money, T., Jones, T., Dixon, R. & Austin, S. **(1999)** Isolation and properties of the complex between the enhancer binding protein NIFA and the sensor NIFL, *Journal of Bacteriology.* **181**, 4461-4468.
195. Moonen, C. T. & Müller, F. **(1983)** On the mobility of riboflavin 5'-phosphate in *Megasphaera elsdenii* flavodoxin as studied by ^{13}C-nuclear-magnetic-resonance relaxation, *European Journal of Biochemistry.* **133**, 463-470.
196. Moonen, C. T., Vervoort, J. & Müller, F. **(1984)** Reinvestigation of the structure of oxidized and reduced flavin: carbon-13 and nitrogen-15 nuclear magnetic resonance study, *Biochemistry.* **23**, 4859-4867.
197. Moore, E. G., Cardemil, E. & Massey, V. **(1987)** Production of a covalent flavin linkage in lipoamide dehydrogenase. Reaction with 8-Cl-FAD, *Journal of Biological Chemistry.* **253**, 6413-6422.
198. Morais Cabral, J. H., Lee, A., Cohen, S. L., Chait, B. T., Li, M. & Mackinnon, R. **(1998)** Crystal structure and functional analysis of the HERG potassium channel N terminus: A eukaryotic PAS domain, *Cell.* **95**, 649-655.
199. Morris, D. L., Ellis, P. B., Carrico, R. J., Yeager, F. M., Schroeder, H. R., Albarella, J. P., Boguslaski, R. C., Hornby, W. E. & Rawson, D. **(1981)** Flavin adenine dinucleotide as a label in homogeneous colorimetric immunoassays, *Analytical Chemistry.* **53**, 658-665.
200. Müller, F. **(1991)** Free flavins: synthesis, chemical and physical properties in *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed) pp. 1-71, CRC Press Inc., Boca Raton.
201. Müller, F. & van Berkel, W. J. H. **(1982)** A study on *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. A convenient method of preparation and some properties of the apoenzyme, *European Journal of Biochemistry.* **128**, 21-27.
202. Müller, F. & van Berkel, W. J. H. **(1991)** Methods used to reversibly resolve flavoproteins in the constituents apoflavoprotein and prosthetic group in *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed) pp. 261-274, CRC Press Inc., Boca Raton.
203. Müller, F., Voordouw, G., van Berkel, W. J. H., Steennis, P. J., Visser, S. & van Rooijen, P. J. **(1979)** A study of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. Improved purification, relative molecular mass, and amino acid composition, *European Journal of Biochemistry.* **101**, 235-244.
204. Munro, A. W., Coggins, J. R., Lindsay, J. G., Kelly, S. & Price, N. C. **(1996)** Deflavination of cytochrome P450 BM3 by treatment with guanidinium chloride, *Biochemical Society Transactions.* **24**, 19S.
205. Munro, A. W., Kelly, S. M. & Price, N. C. **(1999)** Circular dichroism studies of flavoproteins, *Methods in Molecular Biology.* **131**, 111-123.
206. Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W., Daff, S., Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, C. C., Page, C. C. & Dutton, P. L. **(2002)** P450 BM3: the very model of a modern flavocytochrome, *Trends in Biochemical Sciences.* **27**, 250-257.
207. Munro, A. W. & Noble, M. A. **(1999)** Fluorescence analysis of flavoproteins, *Methods in Molecular Biology.* **131**, 25-48.
208. Murthy, Y. V. S. N. & Massey, V. **(1997)** Synthesis and applications of flavin analogs as active site probes of flavoproteins, *Methods in Enzymology.* **280**, 436-460.
209. Nakayama, Y., Yasui, M., Sugahara, K., Hayashi, M. & Unemoto, T. **(2000)** Covalently bound flavin in the NqrB and NqrC subunits of Na(+)-translocating NADH-quinone reductase from *Vibrio alginolyticus*, *FEBS Letters.* **474**, 165-168.
210. Nambu, J. R., Lewis, J. O., Wharton, K. A. J. & Crews, S. T. **(1991)** The *Drosophila single-minded* gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development, *Cell.* **67**, 1157-1167.
211. Neujahr, H. Y. **(1983)** Effect of anions, chaotropes, and phenol on the attachment of flavin adenine dinucleotide to phenol hydroxylase, *Biochemistry.* **22**, 580-584.
212. Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T. A., Welti, R., Goo, Y. A., Leithauser, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W., Maddocks, D. G., Jablonski, P. E., Krebs, M. P., Angevine, C. M., Dale, H., Isenbarger, T. A., Peck, R. F., Pohlschroder, M., Spudich, J. L., Jung, K. W., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P., Omer, A. D., Ebhardt, H., Lowe, T. M., Liang, P., Riley, M., Hood, L. & DasSarma, S. **(2000)** Genome sequence of *Halobacterium* species NRC-1, *Proceedings of the National Academy of Sciences of the United States of America.* **97**, 12176-12181.
213. Nielsen, P., Rauschenbach, P. & Bacher, A. **(1983)** Phosphates of riboflavin and riboflavin analogs: a reinvestigation by high-performance liquid chromatography, *Analytical Chemistry.* **130**, 359-368.
214. Nishina, Y., Horiike, K., Shiga, K., Miyake, Y. & Yamano, T. **(1977)** Effect of halide anions on the binding of FAD to D-amino acid oxidase and the tryptophanyl fluorescence of the apoenzyme, *Journal of Biochemistry.* **81**, 1455-1463.
215. Olteanu, H. & Banerjee, R. **(2001)** Human methionine synthase reductase, a soluble P-450 reductase-like dual flavoprotein, is sufficient for NADPH-dependent methionine synthase activation, *Journal of Biological Chemistry.* **276**, 35558-35563.
216. Ortiz-Maldonado, M., Aeschliman, S. M., Ballou, D. P. & Massey, V. **(2001)** Synergistic interactions of multiple mutations on catalysis during the hydroxylation reaction of *p*-hydroxybenzoate hydroxylase: studies of the Lys297Met, Asn300Asp, and Tyr385Phe mutants reconstituted with 8-Cl- flavin, *Biochemistry.* **40**, 8705-8716.

217. Ortiz-Maldonado, M., Ballou, D. P. & Massey, V. (**1999**) Use of free energy relationships to probe the individual steps of hydroxylation of *p*-hydroxybenzoate hydroxylase: studies with a series of 8-substituted flavins, *Biochemistry.* **38**, 8124-8137.
218. Ortiz-Maldonado, M., Ballou, D. P. & Massey, V. (**2001**) A rate-limiting conformational change of the flavin in *p*-hydroxybenzoate hydroxylase is necessary for ligand exchange and catalysis: studies with 8-mercapto- and 8-hydroxy-flavins, *Biochemistry.* **40**, 1091-1101.
219. Ostrowski, J., Barber, M. J., Rueger, D. C., Miller, B. E., Siegel, L. M. & Kredich, N. M. (**1989**) Characterization of the flavoprotein moieties of NADPH-sulfite reductase from *Salmonella typhimurium* and *Escherichia coli*, *Journal of Biological Chemistry.* **264**, 15796-15808.
220. Otwinowski, Z. & Minor, W. (**1997**) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology.* **276**, 307-326.
221. Paine, M. J., Garner, A. P., Powell, D., Sibbald, J., Sales, M., Pratt, N., Smith, T., Tew, D. G. & Wolf, C. R. (**2000**) Cloning and characterization of a novel human dual flavin reductase, *Journal of Biological Chemistry.* **275**, 1471-1478.
222. Palfey, B. A., Ballou, D. P. & Massey, V. (**1997**) Flavin conformational changes in the catalytic cycle of *p*-hydroxybenzoate hydroxylase substituted with 6-azido- and 6-aminoflavin adenine dinucleotide, *Biochemistry.* **36**, 15713-15723.
223. Palfey, B. A., Entsch, B., Ballou, D. P. & Massey, V. (**1994**) Changes in the catalytic properties of *p*-hydroxybenzoate hydroxylase caused by the mutation Asn300Asp, *Biochemistry.* **33**, 1545-1554.
224. Pedersen, J., Lauritzen, C., Madsen, M. T. & Dahl, S. W. (**1999**) Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases, *Protein Expression and Purification.* **15**, 389-400.
225. Peelen, S. & Vervoort, J. (**1994**) Two-dimensional NMR studies of the flavin binding site of *Desulfovibrio vulgaris* flavodoxin in its three redox states, *Archives of Biochemistry and Biophysics.* **314**, 291-300.
226. Pellequer, J. L., Wager-Smith, K. A., Kay, S. A. & Getzoff, E. D. (**1998**) Photoactive yellow protein: A structural prototype for the three-dimensional fold of the PAS domain superfamily, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 5884-5890.
227. Perman, B., Srajer, V., Ren, Z., Teng, T. Y., Pradervand, C., Ursby, T., Bourgeois, D., Schotte, F., Wulff, M., Kort, R., Hellingwerf, K. & Moffat, K. (**1998**) Energy transduction on the nanosecond time scale: Early structural events in a xanthopsin photocycle, *Science.* **279**, 1946-1950.
228. Piubelli, L., Caldinelli, L., Molla, G., Pilone, M. S. & Pollegioni, L. (**2002**) Conversion of the dimeric D-amino acid oxidase from *Rhodotorula gracilis* to a monomeric form. A rational mutagenesis approach, *FEBS Letters.* **526**, 43-48.
229. Pollegioni, L. & Pilone, M. S. (**1996**) On the holoenzyme reconstitution process in native and truncated *Rhodotorula gracilis* D-amino acid oxidase, *Archives of Biochemistry and Biophysics.* **332**, 58-62.
230. Pompon, D. & Lederer, F. (**1978**) Binding of Cibacron Blue F3GA to flavocytochrome *b*2 from baker's yeast, *European Journal of Biochemistry.* **90**, 563-569.
231. Ponting, C. P. & Aravind, L. (**1997**) PAS: a multifunctional domain family comes to light, *Current Biology.* **7**, R674-R677.
232. Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. (**1975**) Metal chelate affinity chromatography, a new approach to protein fractionation, *Nature.* **258**, 598-599.
233. Porath, J. & Olin, B. (**1983**) Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron and nickel ions, *Biochemistry.* **22**, 1621-1630.
234. Porod, G. (**1982**) General theory in *Small angle X-ray scattering* (Glatter, O. & Kratky, O., eds) pp. 17-51, Academic Press, London.
235. Pueyo, J. J., Curley, G. P. & Mayhew, S. G. (**1996**) Kinetics and thermodynamics of the binding of riboflavin, riboflavin 5'- phosphate and riboflavin 3',5'-bisphosphate by apoflavodoxins, *Biochemical Journal.* **313**, 855-861.
236. Radjendirane, V., Bhat, M. A. & Vaidyanathan, C. S. (**1991**) Affinity purification and characterization of 2,4-dichlorophenol hydroxylase from *Pseudomonas cepacia*, *Archives of Biochemistry and Biophysics.* **288**, 169-176.
237. Raibekas, A. A. & Massey, V. (**1996**) Glycerol-induced development of catalytically active conformation of *Crotalus adamanteus* L-amino acid oxidase in vitro, *Proceedings of the National Academy of Sciences of the United States of America.* **93**, 7546-7551.
238. Raibekas, A. A. & Massey, V. (**1997**) Glycerol-assisted restorative adjustment of flavoenzyme conformation perturbed by site-directed mutagenesis, *Journal of Biological Chemistry.* **272**, 22248-22252.
239. Reitzer, L. J. (**1996**) Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine, and D-alanine in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F. C., ed) pp. 391-407, ASM Press, Washington D.C.
240. Reyes-Ramirez, F., Little, R. & Dixon, R. (**2001**) Role of Escherichia coli nitrogen regulatory genes in the nitrogen response of the *Azotobacter vinelandii* NifL-NifA complex, *Journal of Bacteriology.* **183**, 3076-3082.
241. Riklin, A., Katz, E., Willner, I., Stocker, A. & Bückmann, A. F. (**1995**) Improving enzyme-electrode contacts by redox modification of cofactors, *Nature.* **376**, 672-675.
242. Roos, D. & Winterbourn, C. C. (**2002**) Lethal Weapons, *Science.* **296**, 669-671.
243. Rossman, M. G., Liljas, A., Brändén, C.-I. & Banaszak, L. J. (**1975**) Evolutionary and structural relationships among dehydrogenases, *Enzymes.* **11**, 61-102.
244. Ryerson, C. C., Ballou, D. P. & Walsh, C. (**1982**) Mechanistic studies on cyclohexanone oxygenase, *Biochemistry.* **21**, 2644-2655.
245. Salomon, M., Eisenreich, W., Duerr, H., Schleicher, E., Knieb, E., Massey, V., Ruediger, W., Müller, F., Bacher, A. & Richter, G. (**2001**) An optomechanical transducer in the blue light receptor phototropin from *Avena sativa*, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 12357-12361.
246. Sambrook, J., Fritsch, E. F. & Maniatis, T. (**1989**) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratory Press.
247. Sánchez, R. & Šali, A. (**1998**) Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 13597-13602.
248. Sanner, C., Macheroux, P., Rüterjans, H., Müller, F. & Bacher, A. (**1991**) ^{15}N and ^{13}C-NMR investigations of glucose oxidase from *Aspergillus niger*, *European Journal of Biochemistry.* **196**, 663-672.

249. Santolini, J., Adak, S., Curran, C. M. & Stuehr, D. J. **(2001)** A kinetic simulation model that describes catalysis and regulation in nitric-oxide synthase, *Journal of Biological Chemistry.* **276**, 1233-1243.
250. Schägger, H. & Von Jagow, G. **(1987)** Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophosesis for the separation of proteins in the range from 1 to 100 kDa, *Analytical Biochemistry.* **166**, 368-379.
251. Schibler, U. **(1998)** New cogwheels in the clockwork, *Nature.* **393**, 620-621.
252. Schmidt, M., Tuominen, N., Johansson, T., Weiss, S. A., Keinanen, K. & Oker-Blom, C. **(1998)** Baculovirus-mediated large-scale expression and purification of a polyhistidine-tagged rubella virus capsid protein, *Protein Expression and Purification.* **12**, 323-330.
253. Schmitz, R. A. **(1997)** NifL of *Klebsiella pneumoniae* carries an N-terminally bound FAD cofactor, which is not directly required for the inhibitory function of NifL, *FEMS Microbiology Letters.* **157**, 313-318.
254. Schmitz, R. A., He, L. & Kustu, S. **(1996)** Iron is required to relieve inhibitory effects of NifL on transcriptional activation by NifA in *Klebsiella pneumoniae*, *Journal of Bacteriology.* **178**, 4679-4687.
255. Schopfer, L. M., Massey, V. & Claiborne, A. **(1981)** Active site probes of flavoproteins. Determination of the solvent accessibility of the flavin position 8 for a series of flavoproteins, *Journal of Biological Chemistry.* **256**, 7329-7337.
256. Schuman Jorns, M. **(1979)** Mechanism of catalysis by the flavoenzyme oxynitrilase, *Journal of Biological Chemistry.* **254**, 12145-12152.
257. Schuman Jorns, M., Wang, B., Jordan, S. P. & Chanderkar, L. P. **(1990)** Chromophore function and interaction in *Escherichia coli* DNA photolyase: Reconstitution of the apoenzyme with pterin and/or flavin derivatives, *Biochemistry.* **29**, 552-561.
258. Scola-Nagelschneider, G. & Hemmerich, P. **(1976)** Synthesis, separation, identification, and interconversion of riboflavin phosphates and their acetyl derivatives: a reinvestigation, *European Journal of Biochemistry.* **66**, 567-577.
259. Sharma, S. K., Evans, D. B., Vosters, A. F., Chattopadhyay, D., Hoogerheide, J. G. & Campbell, C. M. **(1992)** Immobilized metal affinity chromatography of bacterially expressed proteins engineered to contain an alternating-histidine domain, *Methods.* **4**, 57-67.
260. Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R. K. & Ermler, U. **(2000)** Structure of coenzyme F_{420} dependent methylenetetrahydromethanopterin reductase from two methanogenic archaea, *Journal of Molecular Biology.* **300**, 935-950.
261. Sinclair, J. C., Delgoda, R., Noble, M. E. M., Jarmin, S., Goh, N. K. & Sim, E. **(1998)** Purification, characterization, and crystallization of an N-hydroxyarylamine O-acetyltransferase from *Salmonella typhimurium*, *Protein Expression and Purification.* **12**, 371-380.
262. Sippl, M. J. **(1993)** Recognition of errors in three-dimensional structures of proteins, *Proteins.* **17**, 355-362.
263. Smith, D. B. **(1993)** Purification of Glutathione S-Transferase fusion proteins, *Methods in Molecular and Cellular Biology.* **4**, 220-229.
264. Smith, W. W., Burnett, R. M., Darling, G. D. & Ludwig, M. L. **(1977)** Structure of the semiquinone form of flavodoxin from *Clostridum MP*. Extension of 1.8 Å resolution and some comparisons with the oxidized state, *Journal of Molecular Biology.* **117**, 195-225.
265. Söderbäck, E., Reyes Ramirez, F., Eydmann, T., Austin, S., Hill, S. & Dixon, R. **(1998)** The redox- and fixed nitrogen-responsive regulatory protein NIFL from *Azotobacter vinelandii* comprises discrete flavin and nucleotide-binding domains, *Molecular Microbiology.* **28**, 179-192.
266. Soderling, S. H., Bayuga, S. J. & Beavo, J. A. **(1998)** Cloning and characterization of cAMP-specific cyclic nucleotide phosphodiesterase, *Proceedings of the National Academy of Sciences of the United States of America.* **95**, 8991-8996.
267. Souza, E. M., Machado, H. B. & Yates, M. G. **(1995)** Deletion analysis of the promotor region of the *nifA* gene from *Herbaspirillum seropedicae* in *Nitrogen Fixation: Fundamentals and Applications* (Tikhonovich, I. A., Provorov, N. A., Romanov, V. I. & Newton, W. E., eds) pp. 259, Kluwer Academic, Dordrecht.
268. Spencer, R., Fisher, J. & Walsh, C. **(1976)** Preparation, characterization, and chemical properties of the flavin coenzyme analogues 5-deazariboflavin, 5-deazariboflavin 5'-phosphate, and 5-deazariboflavin 5'-diphosphate, 5'->5'-adenosine ester, *Biochemistry.* **15**, 1043-1053.
269. Sprenger, W. W., Hoff, W. D., Armitage, J. P. & Hellingwerf, K. J. **(1993)** The eubacterium *Ectothiorhodospira halophila* is negatively photoactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein, *Journal of Bacteriology.* **175**, 3096-3104.
270. Staal, G. E., Helleman, P. W., de Wael, J. & Veeger, C. **(1969)** Purification and properties of an abnormal glutathione reductase from human erythrocytes, *Biochimica et Biophysica Acta.* **185**, 63-69.
271. Stankovich, M. T. **(1991)** Redox properties of flavins and flavoproteins in *Chemistry and Biochemistry of Flavoenzymes* (Müller, F., ed) pp. 401-425, CRC Press Inc., Boca Raton.
272. Sticha, K. R. K., Sieg, C. A., Bergstrom, C. P., Hanna, P. E. & Wagner, C. R. **(1997)** Overexpression and large-scale purification of recombinant hamster polymorphic arylamine N-acetyltransferase as a dihydrofolate reductase fusion protein, *Protein Expression and Purification.* **10**, 141-153.
273. Stockman, B. J., Westler, W. M., Mooberry, E. S. & Markley, J. L. **(1998)** Flavodoxin from *Anabaena* 7120: uniform nitrigen-15 enrichment and hydrogen-1, nitrogen-15, and phosphorous-31 NMR investigations of the flavin mononucleotide binding site in the reduced and oxidized state, *Biochemistry.* **27**, 136-142.
274. Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S. & Olson, M. V. **(2000)** Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature.* **406**, 959-964.
275. Strittmatter, P. **(1961)** The nature of the flavin binding in microsomal cytochrome b_5 reductase, *Journal of Biological Chemistry.* **236**, 2329-2335.
276. Sturtevant, J. M. & Mateo, P. L. **(1978)** Proposed temperature-dependent conformational transition in D-amino acid oxidase: a differential scanning microcalorimetric study, *Proceedings of the National Academy of Sciences of the United States of America.* **75**, 2584-2587.
277. Sukumar, N., Xu, Y., Gatti, D. L., Mitra, B. & Mathews, F. S. **(2001)** Structure of an active soluble

mutant of the membrane-associated (*S*)-mandelate dehydrogenase, *Biochemistry.* **40**, 9870-9878.
278. Sulkowski, E. (**1989**) The saga of IMAC and MIT, *BioEssays.* **10**, 170-175.
279. Svergun, D. I. (**1992**) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *Journal of Applied Crystallography.* **25**, 495-503.
280. Svergun, D. I. (**1993**) A direct indirect method of small-angle scattering data treatment, *Journal of Applied Crystallography.* **30**, 258-267.
281. Svergun, D. I. (**1999**) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophysical Journal.* **76**, 2879-2886.
282. Svergun, D. I. & Koch, M. H. J., unpublished results.
283. Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. (**1988**) Small-angle scattering data treatment by the regularization method, *Acta Crystallographica Section A.* **44**, 244-251.
284. Swindells, M. B. (**1993**) Classification of doubly wound nucleotide binding topologies using automated loop searches, *Protein Science.* **2**, 2146-2153.
285. Swoboda, B. E. (**1969**) The relationship between molecular conformation and the binding of flavin-adenine dinucleotide in glucose oxidase, *Biochimica et Biophysica Acta.* **175**, 365-379.
286. Tahallah, N., van den Heuvel, R. H., van den Berg, W. A. M., Maier, C. S., van Berkel, W. J. H. & Heck, A. J. R. (**2002**) Cofactor-dependent assembly of the flavoenzyme vanillyl-alcohol oxidase, *Journal of Biological Chemistry.* **277**, 36425-36432.
287. Taylor, B. L. & Zhulin, I. B. (**1999**) PAS domains: Internal sensors of oxygen, redox potential, and light, *Microbiology and Molecular Biology Reviews.* **63**, 479-506.
288. Theorell, H. (**1935**) Das gelbe Oxydationsferment, *Biochemische Zeitschrift.* **278**, 263-290.
289. Thompson, S. T., Cass, R. & Stellwagen, E. (**1976**) A quantitative procedure for recovery of nucleoside phosphate ligands and preparation of apoprotein from holoproteins using blue dextran-Sepharose chromatography, *Analytical Biochemistry.* **72**, 293-296.
290. Tishler, M., Pfister, K., Babson, R. D., Ladenburg, K. & Fleming, A. J. (**1947**) The reaction between o-aminoazo compounds and barbituric acid. A new synthesis of riboflavin, *Journal of the American Chemical Society.* **69**, 1487-1492.
291. Trimmer, E. E., Ballou, D. P. & Matthews, R. G. (**2001**) Methylenetetrahydrofolate reductase from *Escherichia coli*: elucidation of the kinetic mechanism by steady-state and rapid-reaction studies, *Biochemistry.* **40**, 6205-6215.
292. Tu, S. C. & McCormick, D. B. (**1974**) Conformation of porcine D-amino acid oxidase as studied by protein fluorescence and optical rotatory dispersion, *Biochemistry.* **13**, 893-899.
293. Vallon, O. (**2000**) New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure, *Proteins.* **38**, 95-114.
294. van Aalten, D. M. F., Crielaard, W., Hellingwerf, K. J. & Joshua-Tor, L. (**2000**) Conformational substates in different crystal forms of the photoactive yellow protein-correlation with theoretical and experimental flexibility, *Protein Science.* **9**, 64-72.
295. van Berkel, W. J. H., Benen, J. A. E. & Snoek, M. C. (**1991**) On the FAD-induced dimerization of apo-lipoamide dehydrogenase from *Azotobacter vinelandii* and *Pseudomonas fluorescens*: Kinetics of reconstitution, *European Journal of Biochemistry.* **197**, 769-780.
296. van Berkel, W. J. H., Eppink, M. H. M. & Schreuder, H. A. (**1994**) Crystal structure of *p*-hydroxybenzoate hydroxylase reconstituted with the modified FAD present in alcohol oxidase from methylotrophic yeasts: Evidence for an arabinoflavin, *Protein Science.* **3**, 2245-2253.
297. van Berkel, W. J. H., Evans, D. J., Laane, C. & Schmedding, D. J. M. (**1997**) Beer and similar light-sensitive beverages with increased flavour stability and process for producing it in *Patent Application NL 97201547.3* Netherlands.
298. van Berkel, W. J. H. & Müller, F. (**1987**) The elucidation of the microheterogeneity of highly purified *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* by various biochemical techniques, *European Journal of Biochemistry.* **167**, 35-46.
299. van Berkel, W. J. H., Regelink, A. G., Beintema, J. J. & de Kok, A. (**1991**) The conformational stability of the redox states of lipoamide dehydrogenase from *Azotobacter vinelandii*, *European Journal of Biochemistry.* **202**, 1049-1055.
300. van Berkel, W. J. H., van den Berg, W. A. M. & Müller, F. (**1988**) Large-scale preparation and reconstitution of apo-flavoproteins with special reference to butyryl-CoA dehydrogenase from *Megasphaera elsdenii*: Hydrophobic-interaction chromatography, *European Journal of Biochemistry.* **178**, 197-208.
301. van Berkel, W. J. H., Weijer, W. J., Müller, F., Jekel, P. A. & Beintema, J. J. (**1984**) Chemical modification of sulfhydryl groups in *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. Involvement in catalysis and assignment in the sequence, *European Journal of Biochemistry.* **145**, 245-256.
302. van den Heuvel, R. H., Fraaije, M. W. & van Berkel, W. J. H. (**2002**) Redox properties of vanillyl-alcohol oxidase, *Methods in Enzymology.* **353**, 177-186.
303. van Mierlo, C. P. M. & Steensma, E. (**2000**) Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story, *Journal of Biotechnology.* **79**, 281-298.
304. van Mierlo, C. P. M., van der Sanden, B. P., van Woensel, P., Müller, F. & Vervoort, J. (**1990**) A two-dimensional ^{1}H-NMR study on *Megasphaera elsdenii* flavodoxin in the oxidized state and some comparisons with the two-electron-reduced state, *European Journal of Biochemistry.* **194**, 199-216.
305. van Mierlo, C. P. M., van Dongen, W. M. A. M., Vergeldt, F., van Berkel, W. J. H. & Steensma, E. (**1998**) The equilibrium unfolding of *Azotobacter vinelandii* apoflavodoxin II occurs via a relatively stable folding intermediate, *Protein Science.* **7**, 2331-2344.
306. Vermilion, J. L. & Coon, J. M. (**1978**) Identification of the high and low potential flavins of liver microsomal NADPH-cytochrome P-450 reductase, *Journal of Biological Chemistry.* **253**, 8812-8819.
307. Vervoort, J. & Hefti, M. H. (**1999**) NMR of flavoproteins in *Flavoprotein Protocols* (Chapman, S. K. & Reid, G. A., eds) pp. 139-147, Humana Press Inc.
308. Vervoort, J., Müller, F., Lee, J., van den Berg, W. A. M. & Moonen, C. T. W. (**1986**) Identifications of the true carbon-13 nuclear magnetic resonance spectrum of the stable intermediate II in bacterial luciferase, *Biochemistry.* **25**, 8062-8067.
309. Vervoort, J., Müller, F., Mayhew, S. G., van den Berg, W. A., Moonen, C. T. & Bacher, A. (**1986**) A comparative carbon-13, nitrogen-15, and phosphorus-31 nuclear magnetic resonance study on the flavodoxins from *Clostridium MP*, *Megasphaera*

elsdenii, and *Azotobacter vinelandii*, *Biochemistry.* **25**, 6789-6799.

310. Vervoort, J., Müller, F., O'Kane, D. J., Lee, J. & Bacher, A. **(1986)** Bacterial luciferase, a carbon-13, nitrogen-15 and phosphorus-31 nuclear magnetic resonance investigation, *Biochemistry.* **25**, 8067-8075.
311. Vervoort, J., van Berkel, W. J. H., Müller, F. & Moonen, C. T. W. **(1991)** NMR studies on *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* and salicylate hydroxylase from *Pseudomonas putida*, *European Journal of Biochemistry.* **200**, 731-738.
312. Vothknecht, U. C., Kannangara, C. G. & Von Wettstein, D. **(1996)** Expression of catalytically active barley glutamyl tRNA-Glu reductase in *Escherichia coli* as a fusion protein with Glutathione S-Transferase, *Proceedings of the National Academy of Sciences of the United States of America.* **93**, 9287-9291.
313. Wang, L. H., Tu, S. C. & Lusk, R. C. **(1984)** Apoenzyme of *Pseudomonas cepacia* salicylate hydroxylase. Preparation, fluorescence property, and nature of flavin binding, *Journal of Biological Chemistry.* **259**, 1136-1142.
314. Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S. S. & Kim, J. J. **(1997)** Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes, *Proceedings of the National Academy of Sciences of the United States of America.* **94**, 8411-8416.
315. Warburg, O. & Christian, W. **(1938)** Isolierung der prosthetischen Gruppe der D-Aminosäureoxydase, *Biochemische Zeitschrift.* **298**, 150-168.
316. Warmke, J. W. & Ganetzky, B. **(1994)** A family of potassium channel genes related to *eag* in *Drosophila* and mammals, *Proceedings of the National Academy of Sciences of the United States of America.* **91**, 3438-3442.
317. Watanabe, F., Fukui, K., Momoi, K. & Miyake, Y. **(1989)** Site-specific mutagenesis of lysine-204, tyrosine-224, tyrosine-228, and histidine-307 of porcine kidney D-amino acid oxidase and the implications as to its catalytic function, *Journal of Biochemistry (Tokyo).* **105**, 1024-1029.
318. Wellner, D. & Meister, A. **(1960)** Crystalline L-amino acid oxidase of *Crotalus adamanteus*, *Journal of Biological Chemistry.* **235**, 2013-2018.
319. Whitby, L. G. **(1953)** A new method for preparing Flavin-Adenine Dinucleotide, *Biochemical Journal.* **54**, 437-442.
320. White, H. B., 3rd & Merrill, A. H., Jr. **(1988)** Riboflavin-binding proteins, *Annual Review of Nutrition.* **8**, 279-299.
321. Wierenga, R. K., Drenth, J. & Schulz, G. E. **(1983)** Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase, *Journal of Molecular Biology.* **167**, 725-739.
322. Williamson, G., Engel, P. C., Mizzer, J. P., Thorpe, C. & Massey, V. **(1982)** Evidence that the greening ligand in native butyryl-CoA dehydrogenase is a CoA persulfide, *Journal of Biological Chemistry.* **257**, 4314-4320.
323. Woodley, P. & Drummond, M. **(1994)** Redundancy of the conserved His residue in *Azotobacter vinelandii* NifL, a histidine autokinase homologue which regulates transcription of nitrogen fixation genes, *Molecular Microbiology.* **13**, 619-626.
324. Wootton, J. C. & Drummond, M. H. **(1989)** The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins, *Protein Engineering.* **2**, 535-543.
325. Yamada, K., Chen, Z., Rozen, R. & Matthews, R. G. **(2001)** Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase, *Proceedings of the National Academy of Sciences of the United States of America.* **98**, 14853-14858.
326. Yorita, K., Misaki, H., Palfey, B. A. & Massey, V. **(2000)** On the interpretation of quantitative structure-function activity relationship data for lactate oxidase, *Proceedings of the National Academy of Sciences of the United States of America.* **97**, 2480-2485.
327. Yoshida, L. S., Saruta, F., Yoshikawa, K., Tatsuzawa, O. & Tsunawaki, S. **(1998)** Mutation at histidine 338 of gp91(phox) depletes FAD and affects expression of cytochrome b558 of the human NADPH oxidase, *Journal of Biological Chemistry.* **273**, 27879-27886.
328. Zeghouf, M., Fontecave, M., Macherel, D. & Coves, J. **(1998)** The flavoprotein component of the *Escherichia coli* sulfite reductase: expression, purification, and spectral and catalytic properties of a monomeric form containing both the flavin adenine dinucleotide and the flavin mononucleotide cofactors, *Biochemistry.* **37**, 6114-6123.
329. Zhang, J., Martasek, P., Paschke, R., Shea, T., Masters, B. S. S. & Kim, J. J. **(2001)** Crystal structure of the FAD/NADPH-binding domain of rat neuronal nitric-oxide synthase. Comparisons with NADPH-cytochrome P450 oxidoreductase, *Journal of Biological Chemistry.* **276**, 37506-37513.
330. Zhang, Y., Burris, R. H., Ludden, P. W. & Roberts, G. P. **(1996)** Presence of a second mechanism for the posttranslational regulation of nitrogenase activity in *Azospirillum brasilense* in response to ammonium, *Journal of Bacteriology.* **178**, 2948-2953.
331. Zhao, Q., Modi, S., Smith, G., Paine, M., McDonagh, P. D., Wolf, C. R., Tew, D., Lian, L. Y., Roberts, G. C. & Driessen, H. P. **(1999)** Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 Å resolution, *Protein Science.* **8**, 298-306.
332. Zhulin, I. B., Taylor, B. L. & Dixon, R. **(1997)** PAS domain S-boxes in Archaea, bacteria and sensors for oxygen and redox, *Trends in Biochemical Sciences.* **22**, 331-333.