AN ABSTRACT OF THE THESIS OF

Ganapathy N. Sarma for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on October 4, 2005.

Title: Crystallographic Studies of *Pyrenophora tritici-repentis* ToxA.

Abstract approved: Redacted for privacy

P. Andrew Karplus

Tan spot of wheat is an economically significant disease caused by the fungal pathogen, *Pyrenophora tritici-repentis*. Certain races of the fungus secrete Ptr ToxA (ToxA), a 13.2 kDa proteinaceous host-selective toxin that is responsible and sufficient to cause disease in susceptible wheat varieties. Disease symptoms develop only when the *ToxA* gene in the fungus and a single gene in the wheat host are expressed. The understanding of this gene-for-gene interaction could be instrumental towards control of the disease and is also being developed as a model system for understanding host-pathogen interactions. Here, this effort is given a solid structural foundation through crystallographic analysis of the ToxA structure.

The ToxA structure was solved at 1.65 Å resolution using the anomalous signal from inherently present sulfur atoms. The monomeric toxin adopts a β-sandwich fold of two anti-parallel β-sheets composed of four strands each. The mapping of existing mutation data onto the structure reveals that a sequence of Arg-Gly-Asp (RGD) and surrounding residues required for activity are present on a solvent-exposed loop thereby making them potential candidates for recognition events that are required for ToxA activity. Unexpectedly, after a simple circular permutation,

the ToxA structure is topologically identical to the classic mammalian RGD-containing fibronectin type III (FnIII) domain, and furthermore the RGD residues are topologically equivalent. These results support the hypothesis that ToxA, like FnIII, interacts with an integrin-like receptor on the host plant cell surface.

There has been a renewed interest in the method of using the anomalous signal from sulfur atoms to solve protein structures. As a spin-off of the structure solution work, the data were systematically analyzed to study the effects of crystal decay, resolution and data redundancy on the ability to locate the sulfur positions and subsequent phasing of the protein. The analyses show that the choices made about data redundancy and resolution limits may be crucial for the structure determination and that anomalous correlation coefficients are helpful indicators in making these choices.

Crystallographic Studies of *Pyrenophora tritici-repentis* ToxA

by

Ganapathy N. Sarma

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented October 4, 2005

Commencement June 2006

Doctor of Philosophy thesis of Ganapathy N. Sarma presented on October 4, 2005.

APPROVED:

Redacted for privacy

Major Professor, representing Biochemistry and Biophysics

Redacted for privacy

Chair of the Department of Biochemistry and Biophysics

Redacted for privacy

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for privacy

Ganapathy N. Sarma, Author

ACKNOWLEDGEMENTS

Surely I am not the first person who finds it hard to come up with the right words thanking all the people – family, teachers and friends – who have made my life a wonderful experience. In the presence of their warmth, wisdom and kindness, I am merely a humble pilgrim who walks the path reassured that he will never find himself alone. Perhaps the difficulty in writing a few paragraphs expressing my gratitude and appreciation comes from the lack of natural flair to put to paper the right words and phrases. I have therefore, taken a different approach. I have listed these people and written a few words about them. The words I use to thank and describe them are the only substitutes that I can think of. To all the people I mention here (in alphabetical order), I can only make one promise – I will strive to pass on the love and knowledge you have shared with me.

Kalyani	Ammal	0	Paati, name, soul
Vivek	Badrinarayana	11	chemist, classmate
Katja	Becker	5	scientist, pfgr, pfaop, encouragement
Joseph	Beckman	4	thesis
Forum	Bhatt	8	classmate, oasis
Akshay	Bhinge	22	friend
Lynda	Ciuffetti	3	scientist, toxa, advice
Aishwarya	Damodar	7	notes
Vidya	Dinesh	10	kindness, kulfi, 'chalega'
Kathy	Faber	4	warm smiles, homemade cereals
Megan	Faber	4	voracious reader, inspiration
Nathan	Faber	4	buddy
Rick	Faber	4	friend, crystallographer, reluctant optimist
Sonia	Fernandes	8	classmate, bombayiite
Wil	Gamble	4	advice
Deven	Gedia	11	aikg member
Frank	Hays	5	classmate, debates, rhetoric
Susan	Hays	5	apple pies, 'just kidding', encouragement

Shing	Ho	5	thesis, dna course
Raghu	Iyer	11	friend, brother
Shyam	Iyer	6	shakti, Hendrix
S. G.	Iyer	20	thatha, philospher, literature
JaiPrakash	Kamat	11	pesarattu
Alka	Kanhere	8	teacher, enthusiasm, advice
Andy	Karplus	5	advisor, crystallographer, mind-boggling protein backbone chain tracing, open doors, encouragement, second chances, optimism, ...
Sharda	Krishnan	29	sister, voice
Subbulakshmi	Krishnan	22	periya amma
Amrish	Makwana	16	friend, realist
Viola	Manning	3	scientist, toxa, inspiration, discussions
Vinay	Mathias	14	architect, encouragement
Pallavi	Mhatre	8	classmate, kelva
Seema	Namboori	9	advice
P	Nandan	10	botanist, teacher
Satya	Nagini	5	teacher, tb samples
Alka	Nerurkar	8	teacher, tb samples, advisor
Amol	Pargoankar	11	classmate, nane ghat
Swathi	Parmeshwaran	10	swati, laughter, 'kushtum'
Daysen	Pather	8	classmate, 'ffwd'
Prajakta	Pradhan	8	classmate, library
Blaine	Roberts	3	labmate
Anne	Roberts	3	down under, thunder, volleyball
Divya	Sarma	19	sister, fellow naag
Leela	Sarma	29	Amma
Mahesh	Sarma	29	Anna, advise, encouragement
Subramaniam	Sarma	14	Maama, flickers, cricket bats, memories
Meghna	Sawant	8	classmate, 'pleeease'
Michael	Schimerlik	5	thesis, advice, discussions, jazz

Heiner	Schirmer	5	scientist, pfgr, cover, enthusiasm
Nilesh	Shirodkar	11	friend, fellow pilgrim
Kunal	Singh	11	friend, ddlj, 'kewal, kintu, parantu'
Jitesh	Soares	8	classmate, feni
Pradeep	Tempalli	5	driving, food
Neelakantan	V	29	Appa
Hema	Vivek	10	classmate, 'what ya'
Zac	Wood	5	crystallographer, advice
Derek	Youngblood	3	labmate

CONTRIBUTION OF AUTHORS

P. Andrew Karplus was involved in the design, analyses and writing of each manuscript. Viola A. Manning purified mature ToxA for crystallization and also performed the gel filtration studies for Chapter 3. Along with Lynda M. Ciuffetti, she was involved in the manuscript preparation (V. A. M and L. M. C are in the Dept. of Botany and Plant Pathology, Oregon State University, USA). Christine Nickel, Stefan Rahlfs and Marina Fischer expressed, purified and crystallized *Pf*AOP in Appendix 1. Katja Becker was involved in the manuscript preparation of Appendices 1 and 2 (C. N, S. R, M. F and K. B are in the Interdisciplinary Research Center, Giessen University, Germany). Savvas N. Savvides crystallized *Pf*GR and collected diffraction data and Markus Schirmer performed the inhibition studies on *Pf*GR for Appendix 2. R. Heiner Schirmer was involved in the manuscript preparation (S. N. S was a graduate student in Dr. Karplus' lab. M. S and R. H. S are in the Biochemistry Center, Heidelberg University, Germany).

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDICES (Continued)

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

Dedicated to

Amma & Appa

And

To the memory of 'Daadi' Thatha, Subbulakshmi Perima & Mani Mama

Crystallographic Studies of *Pyrenophora tritici-repentis* ToxA

Chapter 1

Thesis Overview

This thesis is the culmination of the work I have carried out over the past five years on two separate projects that both employ X-ray crystallography to further our understanding of structure-function relationships of proteins. The first involves the structural studies of a fungal toxin, Ptr ToxA from *Pyrenophora tritici-repentis*. The second project involves the structural studies of two *Plasmodium falciparum* redox-active proteins, glutathione reductase (*Pf*GR) and anti-oxidant protein (*Pf*AOP). At the request of my thesis committee, in order to present a more cohesive thesis, the first project is presented in the main body of the thesis and the second project is separated and included as appendices.

The crystallographic studies of Ptr ToxA are presented in Chapters 2 through 4. *P. tritici-repentis* is a fungal pathogen that causes the economically significant tan spot disease of wheat. The interaction between the fungus and the host is being developed in Dr. Lynda Ciuffetti's laboratory as a model system for understanding host-pathogen interactions. Chapter 2 provides an introduction to the host-selective toxins of *P. tritici-repentis* and summarizes our present knowledge about ToxA. To extend the boundaries of our understanding of this toxin, we have solved the structure of the 13.2 kDa protein using sulfur single-wavelength anomalous dispersion data. The structure provides atomic level information that was used to rationalize existing biochemical data. The structure solution also laid a foundation on which future mutational and biochemical experiments can now be designed. This work has been accepted for publication in *The Plant Cell* and is included as Chapter 3. In a follow-up

study, the anomalous scattering signal in the ToxA diffraction data was systematically analyzed to learn more about the various parameters that influence the ability to solve the ToxA structure. These analyses are included here as Chapter 4 and will be submitted to *Acta Crystallographica D: Biological Crystallography* in the near future.

The second project, described in the appendices of this thesis, was carried out in collaboration with Drs. Heiner Schirmer and Katja Becker in Germany. The structures of *Pf*AOP, reported in Appendix 1, and *Pf*GR, reported in Appendix 2, were undertaken as part of a program to pursue oxidative stress as a potential avenue for drug development against malaria, one of the most devastating diseases in the tropical regions of the world. In both cases, our collaborators provided purified protein and crystallization conditions and the structures were solved by molecular replacement. The structures of these two *Plasmodial* proteins now provide a framework for structure-based drug design. Each of these studies has been published in the *Journal of Molecular Biology*.

Chapter 2

Introduction

2.1 Tan spot disease of wheat

Tan spot disease of wheat is caused by the fungal pathogen, *Pyrenophora tritici-repentis*. This economically significant disease has been identified in the major wheat-growing countries of the world. In the United States, the disease was first observed in the late-1930s in New York and then in Kansas in the early-1940s [1]. Later, the disease was observed in the Central and the Southern Plains [2]. In 1974, the first serious outbreak of the disease occurred in the United States and since then annual crop losses have ranged from 3 to 50% in the Central Plains of the United States and Canada [2-4]. The geographic extent of this disease has increased since then, with wheat losses reported in other countries including Mexico, Paraguay, Brazil, Argentina, regions in North Africa, Azerbaïjan and Australia [5-7]. The increase in the incidence of tan spot of wheat is now universally considered to be related to a change in agricultural practices: there has been a shift from conventional to minimum to zero tillage and also a trend to move away from residual crop burning [8-11]. The result of these practices is that there is residual crop left on the surface of the soil, resulting in an increased likelihood of the fungus surviving from one season to the next. Furthermore, shorter rotation periods between crops or continuous wheat cultivation instead of crop rotation has also contributed to pathogen survival and thus contributed to the increase in disease incidence [8,12-14]. Finally, growing highly susceptible cultivars also enhanced the spread of this disease and increased crop losses [10,15].

2.2 The pathogen

2.2.1 Taxonomy

The word *Pyrenophora* means the presence of a seed or kernel within the ascocarp, which is the fruiting body of some fungi. Historically, this fungus was first observed and isolated from grasses in the early 20th century [6]. In particular this fungus was observed to infect the aggressive weed, *Agropyron repens* (commonly known as

quackgrass). It has been proposed that the pathogen might have evolved within quackgrass to a virulent form before moving into modern wheat, *Triticum aestivum* [7]. While quackgrass has long been known to be the host of this fungus, wheat has now emerged as a primary host. The hyphenated name, *Pyrenophora tritici-repentis*, thus acknowledges its association with the grass and the economic impact it has on wheat production.

2.2.2 **Symptoms**

The fungus is known to penetrate the epidermis and grow intracellularly through the leaf mesophyll and cause necrosis or chlorosis or both [16-18]. Necrosis is defined as cell death and tissue decay that can be extensive throughout the plant. Chlorosis, on the other hand, is a loss of coloration in the tissue (Figure 2.1). Advanced symptoms are observed only in wheat species that are susceptible to the fungus. On resistant wheat, the lesion size is reduced and no chlorosis or necrosis is observed [19]. Wheat varieties are classified as susceptible and resistant based on whether the fungus can cause disease on that particular variety of wheat.

Early on, necrosis and chlorosis were considered as just two symptoms of a single kind of tan spot disease of wheat and various classification systems based on the lesion size [20], the percent infection [21] or a combination of the both [22] were proposed. However, Lamari and Bernier showed that these symptoms were in fact, distinct from each other [23] and that the expression of each of these symptoms was based on specific interactions between certain fungal isolates and certain wheat varieties [23,24].

2.2.3 **Host range**

Using this information, fungal isolates were originally classified into four pathotypes (or races) based on the ability to cause chlorosis or necrosis in certain wheat cultivars. These four types were necrosis and chlorosis (nec^+ chl^+, pathotype 1), necrosis only (nec^+ chl^-, pathotype 2), chlorosis only (nec^- chl^+, pathotype 3) and no necrosis or chlorosis (nec^- chl^-, pathotype 4) [23]. The majority of the *P. tritici-repentis* isolates that were identified based on their activity on one of the four wheat genotypes were found to belong to pathotype 1. This symptom-based classification was found to be flawed in that it could not accommodate new virulence patterns [25].

Figure 2.1. **Tan spot disease of wheat.** Wheat leaves infected by *Pyrenophora tritici-repentis* showing the characteristic necrotic and chlorotic lesions. (http://greengenes.cit.cornell.edu/gifs/wheat_diseases/023.gif)

Currently there are eight races of the fungus, shown in Table 2.1, that have been characterized based on the responses they induce in three wheat cultivars, Glenlea, Katepwa and 6B365 [7,23,26-28]. Also included in Table 2.1 is a fourth wheat cultivar, Salamouni that is resistant to all known isolates of the fungus. The races of the fungus can have simple or complex patterns of virulence. Races 3 and 5 follow a simple pattern; they produce a single toxin and attack only one cultivar of wheat each causing chlorosis. Race 2 also produces a single toxin and is capable of infecting two wheat cultivars. Races 1, 6, and 7 produce two toxins and are capable of attacking two lines of wheat and so form a more complex virulence pattern. The most virulent of all, race 8 produces three toxins and is capable of attacking the three wheat cultivars (Table 2.1) [7,27,28]. As new isolates are tested, it is expected that the number of races will increase.

Races of *P. tritici-repentis* cause either necrosis or chlorosis in the host wheat cultivars in a highly specific manner, suggesting an involvement of host-specific pathogenicity factors or toxins to produce the symptoms associated with infection [19]. It has been shown that this is indeed the case, as the different races of *P. tritici-repentis* do produce different toxins or combination of toxins that cause symptom development in various wheat cultivars [29,30]. Each toxin is active against certain cultivars of wheat but not others. Such patterns of specificity of a toxin to a single host have been well-documented in other fungi (reviewed by Wolpert *et al.* [31]) and such toxins are called host-selective toxins (HSTs).

The HSTs produced by *P. tritici-repentis* are both proteinaceous and non-proteinaceous in nature and so far, three HSTs – Ptr ToxA, Ptr ToxB, and Ptr ToxC – have been discovered that fully account for the pathogenicity profile observed for the 8 *P. tritici-repentis* races (Table 2.1).

Ptr ToxA: Ptr ToxA (hereafter called as ToxA) was the first proteinaceous HST that was isolated from a fungal pathogen [32-35]. This 13.2 kDa toxin is produced by races 1, 2, 7 and 8 [28] (Table 2.1) and is known to cause necrosis in the Glenlea and Katepwa cultivars of wheat [30]. The gene *ToxA* that encodes this protein is present in only those races known to produce the toxin and not in other races [36,37].

Table 2.1. **Host range of different *Pyrenophora tritici-repentis* races** [a]

	Wheat Varieties			
	Glenlea	Katepwa	6B365	Salamouni
Race 1	N (ToxA)	N (ToxA)	C (ToxC)	R
Race 2	N (ToxA)	N (ToxA)	R	R
Race 3	R	R	C (ToxC)	R
Race 4	R	R	R	R
Race 5	R	C (ToxB)	R	R
Race 6	R	C (ToxB)	C (ToxC)	R
Race 7	N (ToxA)	N (ToxA) C (ToxB)	R	R
Race 8	N (ToxA)	N (ToxA) C (ToxB)	C (ToxC)	R

[a] The toxins produced by the races are in parentheses. The symptoms caused by each race is abbreviated as necrosis, N; chlorosis, C; resistant, R.

It has been clearly shown that transformation of fungal isolates that do not produce this toxin with the *ToxA* gene, results in toxin production and pathogenicity [37]. Since the structural determination of this toxin is one of the primary objectives in this thesis, it will be discussed in greater detail later (see below).

Ptr ToxB: This toxin was the second HST to be identified and characterized from *P. tritici-repentis* [38]. Like ToxA, this toxin is also proteinaceous in nature. However, while ToxA cause necrosis in certain wheat cultivars, ToxB causes chlorosis in sensitive wheat cultivars. Races 5 through 8 produce this toxin and cause chlorosis in the wheat cultivar, Katepwa [27,28] (Table 2.1). The *ToxB* gene has been shown to be present in multiple copies [39]. While the protein is synthesized as a 87-amino acid nascent protein, the N-terminal 23 residues are cleaved off and the mature 64-residue (6.6 kDa) toxin is secreted [27,39]. Interestingly, a cryptic *toxb* gene has been found and characterized in non-pathogenic isolates of the fungus [40]. The presence of this gene is interesting because so far, HST genes have been found only in those races that produce the respective toxin. In the isolates that have the *toxb* gene, no transcription of the toxin is seen and it has been suggested that the genes *ToxB* and *toxb* differ in the transcriptional level [40].

Ptr ToxC: ToxC, the third toxin in *P. tritici-repentis*, is partially characterized and appears to be a polar, low-molecular weight molecule [41]. This toxin, produced by races 1, 3, 6 and 8, uniquely causes chlorosis in the wheat cultivar, 6B365 [24,27,28,42]. Difficulty in purifying this toxin has been the hindrance for its characterization [7].

2.3 **Host-selective toxins**

HSTs, in general, are molecules that act as agents of pathogenecity or virulence. As discussed in the earlier cases, these toxins are absolutely essential for pathogenesis. These molecules are produced by the pathogenic strains of certain fungi and are known to have diverse structures. Before the discovery and characterization of ToxA, all known HSTs were small molecular weight compounds that are generally by-products of metabolic pathways. For example, Victorin, a HST from *Cochliobolus victoriae* [43], causes victoria blight of oats. Similarly, T-toxin produced by

Cochliobolus heterostroplus race T causes Southern corn leaf blight disease [44,45]. Of the few exceptions, ToxA and ToxB are the two proteinaceous toxins from *P. tritici-repentis*.

In the case of *P. tritici-repentis*, as discussed earlier, the fungal races are distinguished by the toxins they produce. In addition to the HST, a single gene in the host plant also determines the specificity of the toxin. Thus for disease occurrence there is a requirement for not only the HST but also the product of a gene present in the host plant.

The concept of the requirement of two gene products – one in the pathogen and one in the host – is not new and has been described before in a different context. Before the discovery of the HSTs, the major paradigm for relationships between pathogen and host was the phenomenon of avirulence.

In the phenomenon of avirulence, resistance by a plant to a pathogen occurs when the resistance (*R*) gene product in the plant directly or indirectly recognizes a pathogen that expresses a corresponding avirulence (*Avr*) gene product. This is an important concept that has been termed as the "gene-for-gene" interaction [46,47]. This specific recognition of the pathogen by the host results in the activation of signal transduction pathways in the host that include protein phosphorylation and the generation of reactive oxygen species [48]. These signals result in the activation of genes that encode plant cell wall proteins, proteinase inhibitors, hydrolytic enzymes, peroxidases and other defense related proteins. These signals also activate the hypersensitive response (HR) that results in the death of those plant cells that are in direct contact with the pathogen [49,50]. This prevents the spread of the pathogen to nearby healthy plant cells. The HR also triggers the release of signals that confer non-specific resistance throughout the plant. This phenomenon is called 'systemic acquired resistance' which when activated provides resistance to a whole range of pathogens for a few days.

Via the HR, the expression of the resistance gene in the host makes it incompatible to the pathogen expressing of the corresponding *Avr* gene product. Hence these avirulence factors in the pathogens have been termed 'agents of

incompatibility'. In this classic concept, the absence of the gene, either the *R* gene in the host or the *Avr* gene in the pathogen results in the disease.

Thus, in the classic "gene-for-gene" hypothesis, resistance is brought about by the presence of a gene in the pathogen and a corresponding gene in the host. The HSTs on the other hand, show a gene-for-gene interaction that is the mirror image of that seen in avirulence [31]. In the case of HSTs, sensitivity is brought about by the presence of a gene each in the host and the pathogen. In the absence of one or both genes, resistance results.

2.4 ***Pyrenophora tritici-repentis*** **ToxA**

2.4.1 **Isolation and initial characterization**

ToxA was the first HST isolated that was shown to be a protein. Three groups, working independently of one another, isolated and carried out the initial characterization of ToxA [32-34]. The proteins were named Ptr necrosis toxin [32], Ptr toxin[33], and ToxA [34]. All the proteins had an approximate molecular weight of 13 to 14 kDa and showed similar heat stability properties; when heated and then gradually cooled, they retained activity, however, when heated in the presence of a reducing agent and then cooled, the proteins lost activity. This suggested the presence of a disulfide that is important for the proper folding of the protein [51]. Sequence analyses showed that these proteins were identical and revealed the presence of two cysteine residues. In an effort to standardize the nomenclature, the name Ptr ToxA was adopted[52]. In all three cases, the activity of the toxin was confirmed by infiltrating susceptible wheat cultivars; the development of necrosis, similar to when the fungus infects, has proven to be a reliable assay for testing the activity of this toxin [32-34].

2.4.2 **Structure-function studies**

HSTs have long known to be secondary metabolites of biochemical pathways. Secondary metabolites are low-molecular weight compounds synthesized by the fungus that are not essential for the basic metabolism. ToxA, on the other hand, results from the expression of a single gene and only a single copy of the *ToxA* gene is present in the fungal genome. The protein is expressed as a pre-pro-protein [36,37] (Figure 2.2). The pre-region or the N-terminal peptide (residues 1-22) has been shown

to be involved in the secretion of the protein [37]. The pre-region is followed by the 4.3 kDa N-domain (pro-region; residues 23-60) and the 13.2 kDa C-domain (residues 61-178). The N- and C-domains together form the pro-protein and the C-domain is the mature toxin that is secreted by the fungus. After the pre- and the pro-regions are cleaved, ToxA undergoes post-translational modification with the cyclization of the terminal Gln to form a pyroglutamate. The presence of the pyroglutamate in the N-terminus of ToxA was confirmed by mass spectrometric analysis and enzymatic treatment with pyroglutamate amino peptidase [34,51]. However, it is unlikely that this modification is essential for activity of the mature toxin since recombinant pro-ToxA has been shown to be completely active.

A heterologous expression system in *Escherichia coli* was developed to obtain various ToxA constructs for further studies [51]. ToxA (C-domain) and pro-ToxA (NC-domain) were both expressed in this system as His-tagged proteins. When the activities of these recombinant proteins were compared with natural ToxA, it was seen that recombinant ToxA was only 10 to 20% as active but recombinant pro-ToxA was equally active. This result suggested that the N-domain (pro-region) facilitated the proper folding of the protein, possibly by positioning the two cysteines for disulfide formation [51]. The necessity of the pro-region for the proper folding of the toxin is reminiscent of the requirement of the pro-region of insulin for proper positioning of the cysteine residues for disulfide formation and therefore the correct folding of the hormone [54]. Further, the expression system was also used to produce green fluorescent protein (GFP) tagged pro-ToxA with GFP attached to the N-terminus of pro-ToxA (GFP-ToxA). Using leaf infiltration studies, it was seen that the GFP-ToxA construct showed complete toxic activity by causing necrosis in susceptible wheat varieties [55]. The expression and activity of the GFP-ToxA provided a tool for visualizing the internalization and also the site of ToxA action.

Some biophysical analyses were also carried out to glean structural information about the protein. Circular dichroism (CD) spectroscopy revealed that the protein is mostly composed of β-strands (~ 40%) with a low fraction (~ 8%) of α-helical residues [35], and gel filtration revealed that it behaved as a monomer [34,37,56].

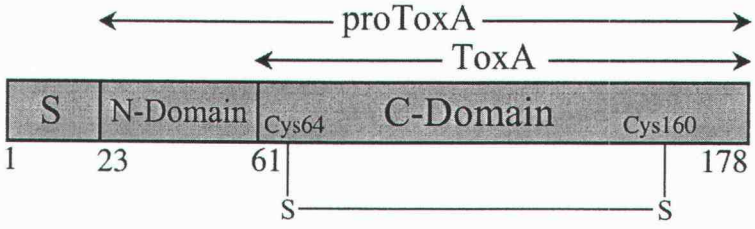


Figure 2.2. **The domain structure of pre-pro-ToxA.** The pre-domain or the signal sequence (S), the pro-region (N-domain), and the mature toxin (C-domain) are labeled. The N- and C-termini are labeled and so are the first residues of the N- and C-domains. The position of the disulfide formed in mature ToxA is also indicated. (This figure is reproduced from Sarma *et al.* [53]; Copyright American Society of Plant Biologists).

2.4.3 Structural information from mutation data

A search of the sequence database did not yield any potential homologs. So, in the absence of atomic-level structural information about the toxin, the sequence of ToxA was compared to other protein sequences using the comparison tool, PROSITE[57] to obtain any information on motifs or residues that could be important for function. PROSITE identified two potential myristolyation sites [37], six potential phosphorylation sites [35,37], and a potential Arg-Gly-Asp (RGD) cell attachment motif[35,37,58] that, in a manner analogous to the mammalian systems, could be involved in adhesion to or entry into the plant cell [37,58,59]. In order to probe their functions, the Gly residues that represent putative myristolyation sites, the Thr residues that represent the potential phosphorylation sites, and the RGD residues along with the surrounding residues were mutated to Ala and their activity tested using leaf infiltration assays. Of the 20 mutations that were carried out, 7 mutants showed wild-type activity, 2 were partially active while the remaining 11 showed no activity. When mutated to Ala, the two Gly residues, potential myristolytion sites, showed partial or no activity. Of the six potential phosphorylation sites, only two mutants showed partial or no activity, suggesting that they could be phosphorylated in the plant [58,59]. The results of the mutation analyses confirmed that the RGD and the surrounding residues were essential for activity of the toxin [59]. However, in the absence of structural information, it is not possible to know if the loss in activity is due to disruption of the structure or due to a specific disruption of interaction with a specific host protein.

2.4.4 ToxA internalization

To study the ToxA site and mechanism of action, sensitive wheat varieties were infiltrated with GFP-ToxA. Surprisingly, not only is ToxA recognized, but is also internalized by the host cell [55] (Figure 2.3). Such import of a protein from the extracellular space across the plant cell membrane is unprecedented. Further, it appears that ToxA associates with the chloroplasts of the host cells (Figure 2.3). The internalization process is very robust since it accommodates co-transport of a 28 kDa GFP that is attached to the N-terminus of a GFP-ToxA fusion protein.

(a)

(b)

Figure 2.3. **Internalization and localization of GFP-ToxA in ToxA sensitive wheat.** Fluorescence microscopy on whole leaf tissue of ToxA insensitive (a) and sensitive (b) wheat leaves treated with GFP-ToxA shows that GFP-ToxA is localized to the chloroplasts only in sensitive leaves. (This figure is reproduced from Manning *et al.* [55]; Copyright American Society of Plant Biologists).

The interaction of an RGD sequence and surrounding residues with integrin receptors has been well documented in the case of animal cells [60-63]. Internalization of proteins that interact with integrin receptors has also been seen [64-67]. Although no integrin or integrin-like proteins have been isolated from plants yet, it has been shown that RGD-containing peptides disrupt the interaction between the plant cell wall and the plasma membrane [68,69].

2.5 **ToxA as a model system for studying host-pathogen interactions**

The unprecedented phenomenon of internalization followed by the localization of ToxA could serve as a model for understanding the mechanisms involved in disease susceptibility of plants when exposed to a pathogen. The available mutation data also provides several clues as to mechanism of ToxA action. Another major advantage of adopting ToxA as a model system for studying host-pathogen interactions lies in the fact that ToxA is a protein that is expressed by a single gene. With the heterologous expression system in place and the availability of GFP-ToxA as a tool, the toxin is now poised to become a model system. The work reported in this thesis is to provide a firm structural foundation for the study of ToxA mechanisms of action.

2.6 **References**

1. Barrus, M. F. (1942). Yellow-spot disease of wheat in New York State. *Plant Disease and Reproduction* 26, 246-248.

2. Hunger, R. M. & Brown, D. A. (1987). Colony color, growth, sporulation, fungicide sensitivity, and pathogenicity of *Pyrenophora tritici-repentis. Plant Dis.* 71.

3. Connors, I. L. (1939). Yellow leaf blotch. *Plant Dis. Surv.* 19, 12-14.

4. Tekauz, A. (1974). Distribution, severity and relative importance of leaf spot disease in hexaploid, tetraploid, and disploid wheat. *Can. Plant Dis. Surv.* 56, 36-40.

5. Ciuffetti, L. M. & Tuori, R. P. (1999). Advances in the characterization of the *Pyrenophora tritici-repentis*-wheat interaction. *Phytopathology* 89, 444-449.

6. De Wolf, E. D., Effertz, R. J., Ali, S. & Francl, L. J. (1998). Vistas of tan spot disease. *Can. J. Plant Pathol.* 20, 349-370.

7. Strelkov, S. E. & Lamari, L. (2003). Host-parasite interaction in tan spot [*Pyrenophora tritici-repentis*] of wheat. *Can. J. Plant Pathol.* 25, 339-349.

8. Sutton, J. C. & Vyn, T. J. (1990). Crop sequences and tillage practices in relation to diseases of winter wheat in Ontaria. *Can. J. Plant Pathol.* 12, 358-368.

9. Bockus, W. W. & Claasen, M. M. (1992). Effects of crop rotation and residue management practices on severity of tan spot of winter wheat. *Plant Dis.* 76, 633-636.

10. Rees, R. G. & Platz, G. J., Eds. (1992). Tan spot and its control-some Australian experiences. Advances in Tan Spot Research: Proceedings of the 2nd International Tan Spot Workshop. Edited by Francl, L. J., Krupinsky, J. M. & McMullen, M. P. Fargo, North Dakota: North Dakota State University.

11. Bailey, K. L. (1996). Diseases under conservation tillage system. *Can. J Plant. Pathol.* 76, 635-639.

12. Freebairn, D. M. (1986). Stubble: The key to success. *Queensl. Agric. J.* 7-8, 197-195.

13. Raymond, P. J., Bockus, W. W. & Norman, B. L. (1985). Tan spot of winter wheat: Procedures to determine host response. *Phytopathology* 75, 686-690.

14. Rees, R. G., Platz, G. J. & Mayer, R. J. (1982). Yield losses in wheat from yellow spot: Comparison of estimates derived from single tillers and plots. *Aust. J. Agric. Res.* 33, 899-908.

15. Diaz de Ackermann, M. & Kohli, M. M. (1998). Research on *Pyrenophora tritici-repentis* tan spot of wheat in Uruguay. In *Helminthosporium Blights of Wheat: Spot blotch and tan spot* (Duveiller, E., Dubin, H. J., Reeves, J. & McNab, A., eds.). International Maize and Wheat improvement center, Mexico City, Mexico.

16. Dushnicky, L. G., Ballance, G. M., Sumner, M. J. & MacGregor, A. W. (1996). Penetration and infection of susceptible and resistant wheat cultivars by a necrosis toxin-producing isolate of *Pyrenophora tritici-repentis*. *Can. J. Plant Pathol.* 18, 392-402.

17. Larez, C. R., Hosford, J., R. M. & Freeman, T. P. (1986). Infection of wheat and oats by *Pyrenophora tritici-repentis* and initial characterization of resistance. *Phytopathology* 76, 931-938.

18. Loughman, R. & Deverall, B. J. (1986). Infection of resistant and susceptible cultivars of wheat by *Pyrenophora tritici-repentis*. *Plant Pathology* 35, 443-450.

19. Hosford, J., R. M. (1971). A form of *Pyrenophora trichostoma* pathogenic to wheat and other grasses. *Phytopathology* 61, 28-32.

20. Cox, D. J. & Hosford, J., R. M. (1987). Resistant winter wheats compared at differing growth stages and leaf positions for tan spot severity. *Plant Dis.* 71, 883-886.

21. Nagle, B. J., Frohberg, R. C. & Hosford, J., R. M. (1982). *Tan spot of wheat and related diseases workshop, Fargo, ND.*

22. da Luz, W. C., Hosford, J., R. M. & Freeman, T. P. (1986). Infection of wheat and oats by *Pyrenophora tritici-repentis* and initial characterization of resistance. *Phytopathology* 76, 931-938.

23. Lamari, L. & Bernier, C. C. (1989). Virulence of isolates of *Pyrenophora tritici-repentis* on 11 wheat cultivars and cytology of the differential host reactions. *Can. J. Plant Pathol.* 11, 284-290.

24. Lamari, L. & Bernier, C. C. (1991). Genetics of tan necrosis and extensive chlorosis in tan spot of wheat caused by *Pyrenophora tritici-repentis*. *Phytopathology* 91, 1092-1095.

25. Lamari, L., Sayoud, R., Boulif, M. & Bernier, C. C. (1995). Identification of a new race of *Pyrenophora tritici-repentis*: Implications for the current pathotype classification system. *Can. J. Plant Pathol.* 17, 312-318.

26. Lamari, L., Ballance, G. M., Orolaza, N. P. & Kowatsch, R. (1995). In planta production and antibody neutralization of the Ptr necrosis toxin from *Pyrenophora tritici-repentis*. *Phytopathology* 85, 333-338.

27. Strelkov, S. E., Lamari, L., Sayoud, R. & Smith, R. B. (2002). Comparative virulence of chlorosis-inducing races of *Pyrenophora tritici-repentis*. *Can. J. plant Pathol.* 24, 29-35.

28. Lamari, L., Strelkov, S., Yahyaoui, A., Orabi, J. & Smith, R. B. (2003). The identification of two new races of *Pyrenophora tritici-repentis* from the host center of diversity confirms a one-to-one relationship in tan spot of wheat. *Phytopathology* 93, 391-396.

29. Tomas, A. & Bockus, W. W. (1987). Cultivar-specific toxicity of culture filtrates of *Pyrenophora tritici-repentis*. *Phytopathology* 77, 1337-1340.

30. Lamari, L. & Bernier, C. C. (1989). Toxin of *Pyrenophora tritici-repentis*: Host-specificity, significance of disease, and inheritance of host reaction. *Phytopathology* 79, 740-744.

31. Wolpert, T. J., Dunkle, L. D. & Ciuffetti, L. M. (2002). Host-selective toxins and avirulence determinants: what's in a name? *Annu. Rev. Phytopathol.* 40, 251-285.

32. Ballance, G. M., Lamari, L. & Bernier, C. C. (1989). Purification and characterization of a host-selective necrosis toxin from *Pyrenophora tritici-repentis*. *Physiol. Mol. Plant Pathol.* 35, 203-213.

33. Tomas, A., Feng, G. H., Reeck, G. R., Bockus, W. W. & Leach, J. E. (1990). Purification of a cultivar-specific toxin from *Pyrenophora tritici-repentis*, causal agent of tan spot of wheat. *Mol. Plant-Microbe Interact.* 3, 221-224.

34. Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. (1995). Purification and immunological characterization of toxic components from cultures of *Pyrenophora tritici-repentis*. *Mol. Plant-Microbe Interact.* 8, 41-48.

35. Zhang, H., Francl, L. J., Jordahl, J. G. & Meinhardt, S. W. (1997). Structural and physical properties of a necrosis-inducing toxin from *Pyrenophora tritici-repentis*. *Phytopathology* 87, 154-160.

36. Ballance, G. M., Lamari, L., Kowatsch, R. & Bernier, C. C. (1996). Cloning, expression and occurrence of the gene encoding the Ptr necrosis toxin from *Pyrenophora tritici-repentis*. *Mol. Plant Pathol. Online*, publication/1996/1209.ballance.

37. Ciuffetti, L. M., Tuori, R. P. & Gaventa, J. M. (1997). A single gene encodes a selective toxin causal to the development of tan spot of wheat. *Plant Cell* 9, 135-144.

38. Strelkov, S. E., Lamari, L. & Ballance, G. M. (1999). Characterization of a host-specific protein toxin (Ptr ToxB) from *Pyrenophora tritici-repentis*. *Mol. Plant-Microbe Interact.* 12, 728-732.

39. Martinez, J. P., Ottum, S. A., Ali, S., Francl, L. J. & Ciuffetti, L. M. (2001). Characterization of the *ToxB* Gene from *Pyrenophora tritici-repentis*. *Mol. Plant-Microbe Interact.* 14, 675-677.

40. Martinez, J. P., Oesch, N. W. & Ciuffetti, L. M. (2004). Characterization of the multiple copy host-selective toxin gene, *ToxB*, in pathogenic and nonpathogenic isolates of *Pyrenophora tritici-repentis*. *Mol. Plant-Microbe Interact.* in press.

41. Effertz, R. J., Meinhardt, S. W., Anderson, J. A., Jordahl, J. G. & Francl, L. J. (2002). Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an insensitivity locus in wheat. *Phytopathology* 92, 527-533.

42. Gamba, F. M., Lamari, L. & Brule-Babel, A. L. (1998). Inheritance of race-specific necrotic and chlorotic reactions induced by *Pyrenophora tritici-repentis* in hexaploid wheats. *Can. J. Plant Pathol.* 20, 401-407.

43. Meehan, F. & Murphy, H. C. (1946). A new *Helminthosporium* blight of oats. *Science* 104, 413-414.

44. Levings, C. S., Rhoads, D. M. & Siedow, J. N. (1995). Molecular interactions of *Bipolaris maydis* T-toxin and maize. *Can. J. Plant Pathol.* 73, 483-489.

45. Rhoads, D. M., Brunner-Neuenschwander, B., Levings, C. S. & Siedow, J. N. (1998). Characterization of the interaction between fungal pathotoxins and Urf13, the cms-T maize mitochondrial T-toxin receptor. In *Molecular genetics of host-specific toxins in plant disease* (Kohmoto, K. & Yoder, O. C., eds.), pp. 355-365. Kluwer, Dordrecht, Netherlands.

46. Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 9, 275-296.

47. Dangl, J. L. & Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826-33.

48. Bolwell, G. P. (1999). Role of active oxygen species and NO in plant defence responses. *Curr. Opin. Plant Biol.* 2, 287-94.

49. Heath, M. C. (2000). Hypersensitive response-related death. *Plant Mol. Biol.* 44, 321-34.

50. Lam, E., Kato, N. & Lawton, M. (2001). Programmed cell death, mitochondria and the plant hypersensitive response. *Nature* 411, 848-53.

51. Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. (2000). Heterologous expression of functional Ptr ToxA. *Mol. Plant-Microbe Interact.* 13, 456-464.

52. Ciuffetti, L. M., Francl, L. J., Ballance, G. M., Bockus, W. W., Lamari, L., Meinhardt, S. W. & Rasmussen, J. B. (1998). Standardization of toxin nomenclature in the *Pyrenophora tritici-repentis*/wheat interaction. *Can. J. Plant Pathol.* 20, 421-424.

53. Sarma, G. N., Manning, V. A., Ciuffetti, L. M. & Karplus, P. A. (2005). Structure of Ptr ToxA: an RGD-containing host-selective toxin from *Pyrenophora tritici-repentis. Plant Cell* Published online on October 7, 2005; 10.1105/tpc.105.034918.

54. Docherty, K. & Steiner, D. F. (1982). Post-translational proteolysis in polypeptide hormone biosynthesis. *Annu. Rev. Physiol.* 44, 625-38.

55. Manning, V. A. & Ciuffetti, L. M. (2005). Localization of Ptr ToxA produced by *Pyrenophora tritici-repentis* reveals protein import into wheat mesophyll cells. *Plant Cell* Published online on September 30, 2005; 10.1105/tpc.105.035063.

56. Meinhardt, S. W., Zhang, H., Effertz, R. J. & Francl, L. J. (1998). Characterization of additional peaks of necrosis activity from *Pyrenophora tritici-repentis. Can. J. Plant Pathol.* 20, 436-437.

57. Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. & Bucher, P. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform.* 3, 265-274.

58. Meinhardt, S. W., Cheng, W., Kwon, C. Y., Donohue, C. M. & Rasmussen, J. B. (2002). Role of the arginyl-glycyl-aspartic motif in the action of Ptr ToxA produced by *Pyrenophora tritici-repentis. Plant Physiol.* 130, 1545-1551.

59. Manning, V. A., Andrie, R. M., Trippe, A. F. & Ciuffetti, L. M. (2004). Ptr ToxA requires multiple motifs for complete activity. *Mol. Plant-Microbe Interact.* 17, 491-501.

60. Xiao, T., Takagi, J., Coller, B. S., Wang, J. H. & Springer, T. A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* 432, 59-67.

61. Takagi, J., Strokovich, K., Springer, T. A. & Walz, T. (2003). Structure of integrin alpha5beta1 in complex with fibronectin. *EMBO J.* 22, 4607-15.

62. Springer, T. A. & Wang, J. H. (2004). The three-dimensional structure of integrins and their ligands, and conformational regulation of cell adhesion. *Adv. Protein Chem.* 68, 29-63.

63. Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L. & Arnaout, M. A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 296, 151-5.

64. Memmo, L. M. & McKeown-Longo, P. (1998). The alphavbeta5 integrin functions as an endocytic receptor for vitronectin. *J. Cell Sci.* 111 (Pt 4), 425-33.

65. Hynes, R. O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell* 110, 673-87.

66. Cherny, R. C., Honan, M. A. & Thiagarajan, P. (1993). Site-directed mutagenesis of the arginine-glycine-aspartic acid in vitronectin abolishes cell adhesion. *J. Biol. Chem.* 268, 9725-9729.

67. Marjomaki, V., Pietiainen, V., Matilainen, H., Upla, P., Ivaska, J., Nissinen, L., Reunanen, H., Huttunen, P., Hyypia, T. & Heino, J. (2002). Internalization of echovirus 1 in caveolae. *J. Virol.* 76, 1856-65.

68. Canut, H., Carrasco, A., Galaud, J. P., Cassan, C., Bouyssou, H., Vita, N., Ferrara, P. & Pont-Lezica, R. (1998). High affinity RGD-binding sites at the plasma membrane of *Arabidopsis thaliana* links the cell wall. *Plant J.* 16, 63-71.

69. Mellersh, D. G. & Heath, M. C. (2001). Plasma membrane-cell wall adhesion is required for expression of plant defense responses during fungal penetration. *Plant Cell* 13, 413-424.

Chapter 3

Structure of Ptr ToxA: an RGD-containing host-selective toxin from *Pyrenophora tritici-repentis* *

Ganapathy N. Sarma, Viola A. Manning, Lynda M. Ciuffetti, and P. Andrew Karplus

Published online in *The Plant Cell* (October 7, 2005), 10.1105/tpc.105.034918

3.1 Abstract

Tan spot of wheat, caused by the fungus *Pyrenophora tritici-repentis*, has significant agricultural and economic impact. Ptr ToxA (ToxA), the first discovered proteinaceous host-selective toxin, is produced by certain *P. tritici-repentis* races, and is necessary and sufficient to cause cell death in sensitive wheat cultivars. We present here the high-resolution crystal structure of ToxA in two different crystal forms, providing four independent views of the protein. ToxA adopts a single domain, β-sandwich fold of novel topology. Mapping of the existing mutation data onto the structure supports the hypothesized importance of an Arg-Gly-Asp (RGD) and surrounding sequence. Its occurrence in a single, solvent-exposed loop in the protein suggests that it is directly involved in recognition events required for ToxA action. Furthermore, the ToxA structure reveals a surprising similarity with the classic mammalian RGD-containing domain, the fibronectin type III (FnIII) domain: the two topologies are related by circular permutation. The similar topologies and the positional conservation of the RGD-containing loop raises the possibility that ToxA is distantly related to mammalian FnIII proteins and that to gain entry it binds to an integrin-like receptor in the plant host.

3.2 Introduction

Pyrenophora tritici-repentis is a fungal pathogen that causes tan spot of wheat in sensitive cultivars. This disease is found throughout the major wheat-growing regions of the world and is characterized by either necrotic lesions or chlorosis or both. As for a number of fungi, disease is associated with the production of host-selective toxins (HSTs; reviewed in ref. 1, 2). *P. tritici-repentis* produces various HSTs each of which uniquely leads to progression of disease on wheat cultivars that are sensitive to that particular toxin (reviewed in ref. 3-5). Of the three HSTs thus far characterized from *P. tritici-repentis*, two – Ptr ToxA [6-9] and Ptr ToxB [10] – have been shown to be proteins, whereas Ptr ToxC is a low molecular weight compound [11].

---

* This chapter includes figures that are published as supplemental information in the *The Plant Cell* article.

For Ptr ToxA (hereafter called ToxA), the first protein HST to be isolated [6-9], sensitivity is conditioned by the presence of a single host locus *Tsn1*, on the 5BL chromosome [12-15] and the *ToxA* gene in the fungus [16,17]; disease occurs only when both the genes are present.

As a host-selective toxin, not only is ToxA required for toxicity but infiltrating leaves of sensitive wheat cultivars with ToxA alone results in symptom development [6-8]. As shown by Manning and Ciuffetti [18], green fluorescent protein tagged ToxA (GFP-ToxA) exerts its toxic effect via internalization into sensitive wheat mesophyll cells. After internalization, it appears to localize to chloroplasts, but the mechanisms involved in cell death remain unknown. Because ToxA itself plays a pivotal role in toxicity, its structure is of great interest.

The *ToxA* gene encodes a pre-pro-protein [16,17] (Figure 3.1). The pre-region (i.e. the signal peptide; residues 1-22) targets the protein to the secretory pathway. The pro-region (also known as the N-domain; residues 23-60) is cleaved off prior to secretion of the mature, 13.2 kDa ToxA (i.e., the C-domain; residues 61-178) [8,17]. Evidence suggests that the N-domain, much like the pro-peptide of insulin, is required during folding for the proper formation of the disulfide in the C-domain that stabilizes the active, native conformation of ToxA [19].

A database search for proteins similar to ToxA did not yield any potential homologs. In the absence of structural or sequence conservation information to guide mutagenesis experiments, sequence motif searches were carried out to find potential functional sites. The searches revealed putative sites of phosphorylation (Thr residues 115, 123, 126, 132, 139, and 167) [9,17], of myristoylation (Gly residues 62 and 93) [17] and of cell attachment (an RGD-containing motif; residues Asn136 through Phe147)[9,17,20]. Mutagenesis studies have identified a number of residues, including the RGD motif, that may be important for function [18-20], but without structural information it was not possible to conclusively deduce which residues were important for structural reasons and which might be directly involved in recognition events. Here, we present the high-resolution, crystal structure of natural, mature ToxA, and correlate the structure with biochemical results.

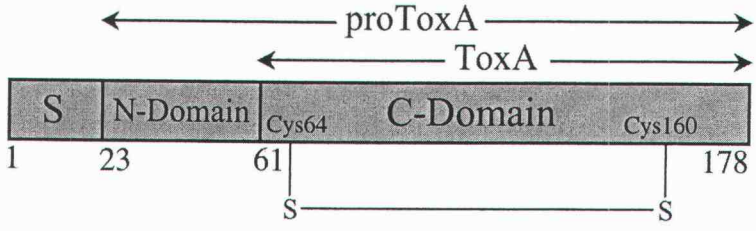


Figure 3.1. **The functional domain structure of nascent ToxA.** The pre-domain or the signal sequence (S), the pro-region (N-domain), and the mature toxin (C-domain) are labeled. The N- and C-termini are labeled and so are the first residues of the N- and C-domains. The position of the disulfide formed in mature ToxA is also indicated. The complete sequence of the mature toxin is shown in Figure 3.2c.

Table 3.1. **Data collection and refinement statistics** [a] *(continued)*

Data collection and phasing	Sulfur SAD – form-I (in-house)	Native – form-I (synchrotron)	Native – form-II (synchrotron)	Native – form-II (in-house)
Resolution	100-1.90 (1.97-1.90)	100-1.65 (1.71-1.65)	100-1.90 (1.97-1.90)	100-3.00 (3.11-3.00)
Reflections (total/ unique)	808412/ 12883	147079/ 19469	191032/ 31109	60556/ 8241
Completeness	99.9 (100.0)	99.9 (99.8)	99.4 (100.0)	99.9 (99.9)
R_{meas} [b]	8.2 (50.7)	8.0 (50.2)	6.8 (56.8)	10.3 (39.5)
$R_{mrgd\text{-}F}$ [b]	2.8 (12.5)	7.9 (40.9)	7.2 (43.2)	8.9 (27.5)
I/ σ	62.9 (56.0)	13.9 (2.6)	11.3 (2.2)	19.0 (5.4)
FOM/ Δφ (°) – Mlphare [c]	0.2/ 66			
FOM/ Δφ (°) – DM [c]	0.7/ 51			

Refinement	Form-I (synchrotron)	Form-II (synchrotron)
Resolution (Å)	100-1.65 (1.69-1.65)	100-1.90 (1.95-1.90)
Number of reflections	19430	31030
Number of amino acids	103	315
Number of solvent atoms	105	165
Number of Ni^{+2} ions	-	8
Number of SO_4^{2-} ions	2	-
Total number of atoms	968	2668

Table 3.1. **Data collection and refinement statistics** [a] *(continued)*

Refinement	Form-I (synchrotron)	Form-II (synchrotron)
Average B (all atoms) (Å^2)	28	45
R_{cryst} (%)	15.7 (21.6)	18.3 (24.5)
R_{free} (%)	18.3 (26.3)	22.6 (31.3)
Rmsd bonds (Å)	0.012	0.008
Rmsd angles (degrees)	1.435	1.097
ϕ, ψ – core region (%) [d]	100	98

[a] Numbers in parentheses correspond to values in the highest resolution shell.
[b] R_{meas} is the multiplicity weighted merging R-factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data [67].
[c] Figure of merit, average absolute phase difference between experimental (Mlphare/ DM) and final refined phases.
[d] Ramachandran plot quality as defined by Kleywegt and Jones [68]. In chain B of the form-II structure, Asn77 (ϕ, ψ = -111°, 73°) and Asn127 (ϕ, ψ = -135°, -159°) lie in the non-core region.

3.3 Results and Discussion

3.3.1 Structure solution

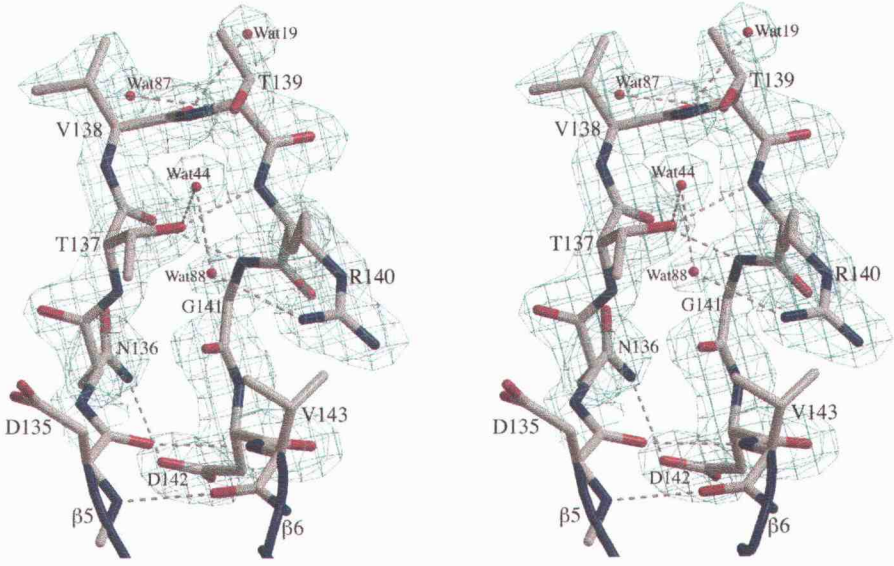
The structure of mature ToxA was solved in two different crystal forms. The structure in form-I crystals (with one molecule in the asymmetric unit) was determined at 3.0 Å resolution using in-house, sulfur single-wavelength anomalous dispersion (SAD) data collected to 1.9 Å resolution. Refinement against 1.65 Å resolution data collected at a synchrotron (see Methods) led to a model with a final R_{free} of 18.3% and good geometry (Table 3.1). The form-II crystal structure (with three molecules in the asymmetric unit) was determined by molecular replacement using the form-I structure as the search model. The structure was refined at 1.9 Å resolution to a final R_{free} of 22.6%. In both form-I and -II crystals, ToxA is present as a trimer (the trimer in form-I is generated by the crystallographic 3-fold axis). Having four independently refined molecules provides insight into the variability of structures. Based on Luzzati analysis[21,22], the estimated coordinate accuracies of the well-ordered parts of the models are 0.2 Å and 0.3 Å for the form-I and form-II structures respectively. The root-mean-square deviation (rmsd) between the Cα atoms of the three chains of form-II is ca. 0.3 Å. The rmsd's between the form-I structure and the three monomers of form-II are ca. 0.6 Å indicating small differences exist. Even though crystal form-I is at a higher resolution, we have prepared figures using the form-II structure because the structures are very similar and the form-II structure has a better ordered RGD-containing loop.

The majority of the main chain in both crystal forms has clear electron density (Figure 3.2a). The exception is about 15 residues near the N-terminus that, due to disorder, have little or no density and have not been modeled (see Methods). One last feature requiring comment is the disulfide bond between the Cys64 and Cys160 [19]. The electron density seen for Cys64 and Cys160 in form-II shows that the sulfurs are unexpectedly 3.5 Å apart in a non-bonded interaction (and in form-I no electron density is seen for Cys64). A control data set collected on our in-house X-ray source (Table 3.1) reveals an intact disulfide, so the cleavage is an artifact of synchrotron data collection as has been seen for other proteins [23,24]. For this reason, we will treat it in this paper as a disulfide.

Figure 3.2. **Tertiary structure of ToxA.** (a) Electron density map quality and the RGD-loop structure. A stereo diagram of the $2F_o–F_c$ electron density map contoured at 1.0 ρ_{rms} shows clear density for the main chain atoms and most of the side chain atoms of the RGD loop of chain B in the form-II crystals. The residues are labeled, and four water sites observed in all four molecules of ToxA are shown. Hydrogen bonds are indicated by broken gray lines. This figure was prepared using BobScript [25] and Raster3D [26]. (b) The overall structure of the ToxA monomer. A stereo ribbon diagram of ToxA monomer emphasizing the β-sandwich fold with the secondary structural elements labeled and colored (β-strands in green and α-helix in blue). The Cα positions of the RGD residues are denoted by semi-transparent spheres and the four ordered N-terminal residues (including Pca61) are shown as ball-and-stick models. The separation of the Sγ atoms of Cys64 and Cys160, actually disulfide bonded in the native protein, is indicated by a dotted yellow line. The possible path of the unmodeled residues is drawn as broken lines. The figure was prepared using Molscript [27] and Raster3D [26]. (c) The topology and main chain hydrogen-bonding of ToxA. A schematic representation of the ToxA fold is shown with the secondary structural elements labeled. The two β-sheets are shaded green and labeled while the residues of the α-helix are shaded blue and boxed. Amino acids are numbered and indicated by standard three-letter code. The residues not visible in one or more chains of ToxA are shown as broken circles. Main chain hydrogen bonds are indicated by arrows drawn pointing from the amide nitrogen to the carbonyl oxygen. A β-bulge present in strand β6 involves Phe147 which has unusual β-sheet torsion angles (ϕ = -86°, ψ = 47°). (d) Backbone mobilities of ToxA. The average main chain B-factors of form-I (thick), and form-II chains A (thin), B (dashed) and C (dashed-dotted) are plotted against residue numbers. Locations of the secondary structure elements and the RGD tripeptide are shown below the plot. As expected, loops and regions not stabilized by crystallographic packing interactions have B-factors higher than average, while β-strands forming the protein core have lower B-factors. For the loops with B-factors approaching 90 $Å^2$, the most mobile residues have weak main chain density (<1 ρ_{rms}) but the chain path is still easily interpretable.

Figure 3.2 (continued)

(a)



(b)

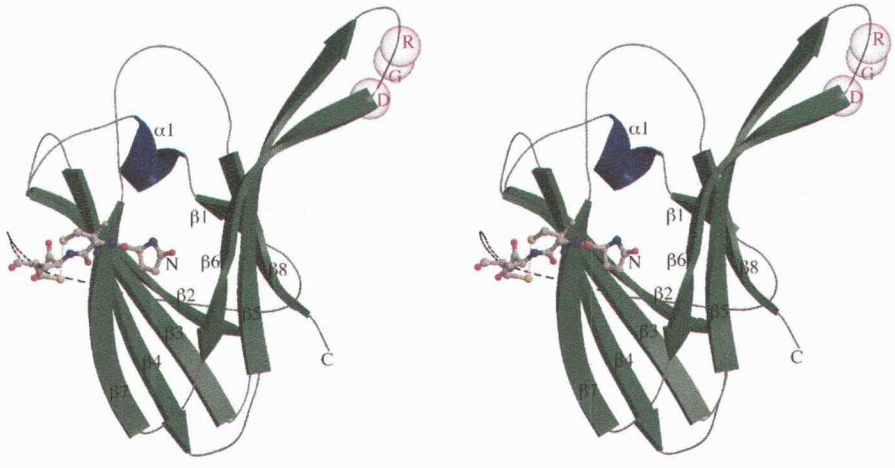


Figure 3.2 (continued)

(c)

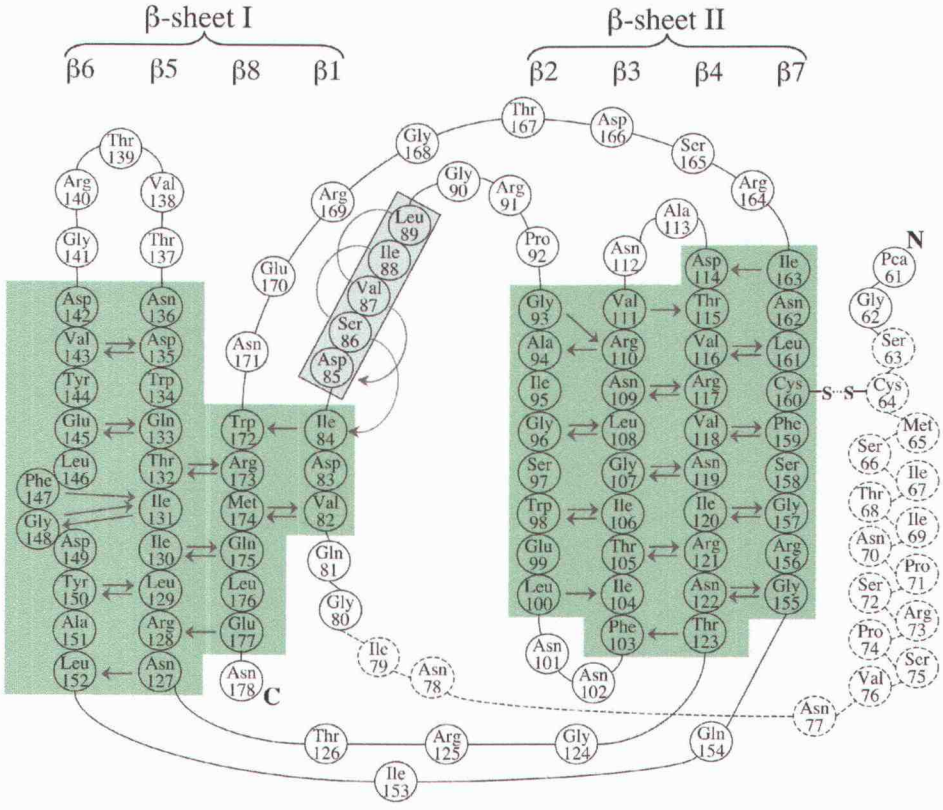
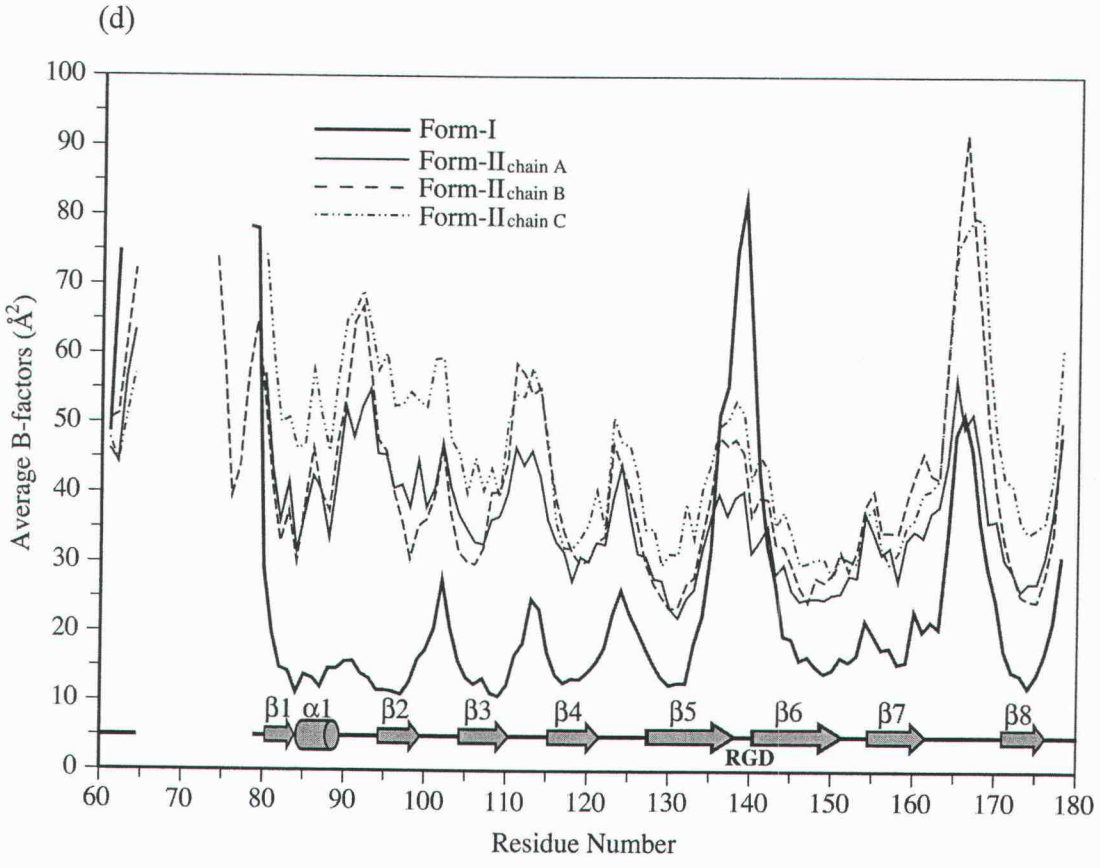


Figure 3.2 (continued)



3.3.2 **Overall structure**

ToxA is a single domain protein having a β-sandwich fold with two anti-parallel β-sheets composed of four strands each enclosing the hydrophobic core (Figure 3.2b). At the N-terminal region of the mature protein that includes pyroglutamate-61 (Pca61) [19], only the first two residues in form-I and first four in form-II show interpretable density (Figure 3.3), which are followed by 15-16 disordered residues before the first β-strand. The strands are numbered sequentially from β1 to β8 with sheet I composed of strands β1, β8, β5 and β6 and sheet II containing strands β2, β3, β4 and β7. A single one-turn α-helix (α1) is present between β1 and β2. The topologies of the sheets are shown in Figures 2.2b and 2.2c. As seen in Figure 3.2d, the β-strands form a rather rigid protein core (low B-factors) and the loops are more mobile (high B-factors).

Among the seven loops connecting the secondary structural elements, three are tight turns (3-4 residues) connecting adjacent anti-parallel strands (loops β2-β3, β3-β4, and β6-β7), and the other four are longer (4-6 residues), bridging strands and helices on opposite sides of the sandwich. The RGD residues (140-142) hypothesized to be functionally important are located in the β5-β6 loop (Figure 3.2b). In the form-I, the loop conformation is influenced by the binding of a sulfate ion to the side chain of Arg140 and main chain amides, but it is nevertheless highly mobile (Figure 3.2d). In form-II, Arg140 interacts with another chain in the trimer (see below) resulting in lower mobility when compared to the cubic form (Figure 3.2d).

3.3.3 **Trimer structure**

Although ToxA is present as a strongly interacting trimer in both the crystal forms (Figure 3.4a), during gel filtration it runs at an apparent M_r of 13 kDa (see Methods; Figure 3.5), suggesting that it is a monomer in solution [8,28]. Sedimentation equilibrium results were consistent with the existence of a monomer-trimer equilibrium in solution with an association constant (K_a) $\approx 1 \times 10^9$ M^{-2} (see Methods; Figure 3.6a and b). This K_a implies that at ToxA concentrations showing toxicity during infusion experiments (~ 1 μM) [19], pure ToxA would be > 99% in the monomeric form.

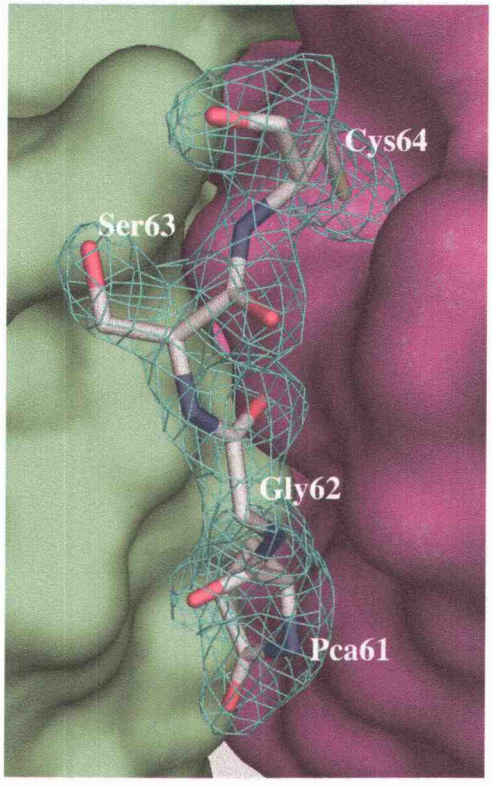


Figure 3.3. **N-terminal residues of ToxA.** The $2F_o - F_c$ electron density map contoured at 1.0 ρ_{rms} shows unambiguous density for residues 61 through 64 situated at the trimer interface in the form-II crystals. The residues, including the pyroglutamate 61 (Pca61), are labeled and the orientation and the coloring scheme are identical to that of Figure 3.4a. The figure was generated using Pymol (www.pymol.org).

Figure 3.4. **The crystallographic trimer of ToxA.** (a) A ribbon diagram of the ToxA trimer as seen in the asymmetric unit of the form-II structure looking down the 3-fold axis is shown. The three chains are labeled and shown in green, blue and purple. The four N-terminal residues and the RGD residues are shown as stick models. (b) The two trimerization surfaces of ToxA. For this image, the 'green' molecule from (a) is separated from the others and rotated 90° around the vertical. The narrow spine-like interface interacting with the purple molecule (colored purple) is facing the reader. The broad vise-shaped interface interacting with the blue molecule (colored blue) faces down. For trimer assembly, the blue surface of one molecule meshes with the purple surface of another. Residues having ≥ 40 Å^2 surface area buried at the interface are labeled. The 3-fold axis is indicated and labeled. This figure was generated using Pymol (www.pymol.org).

Figure 3.4 (continued)

(a)

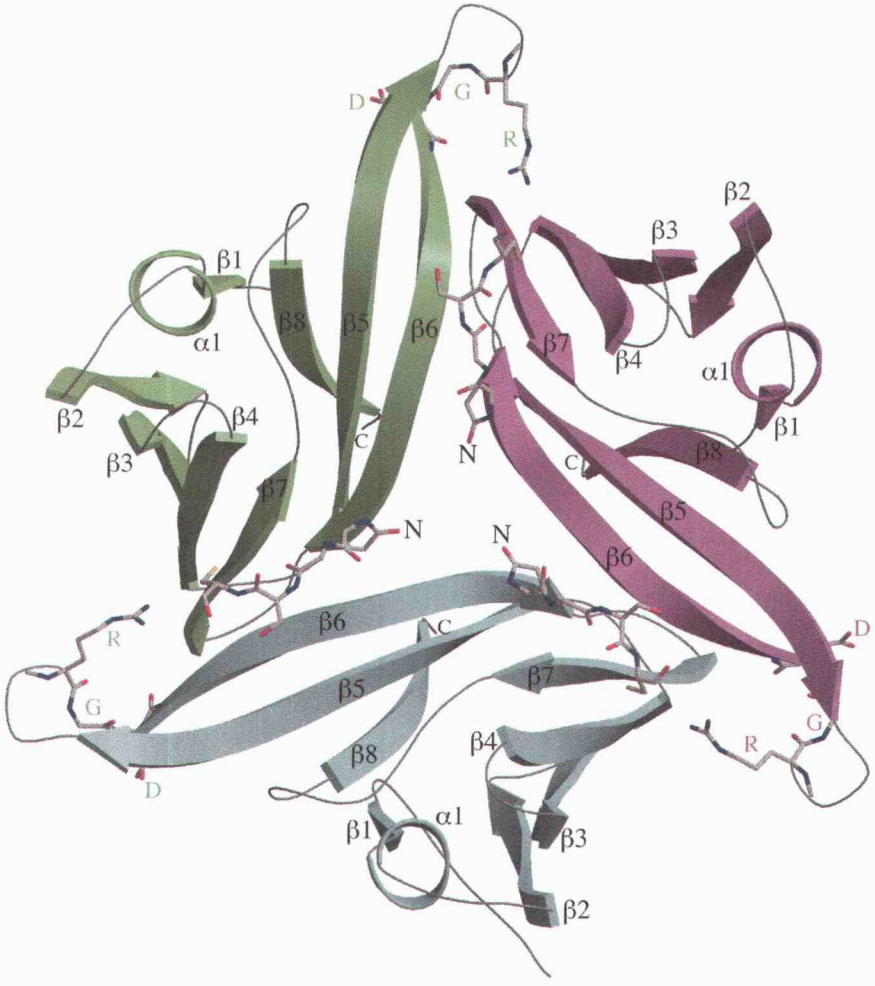
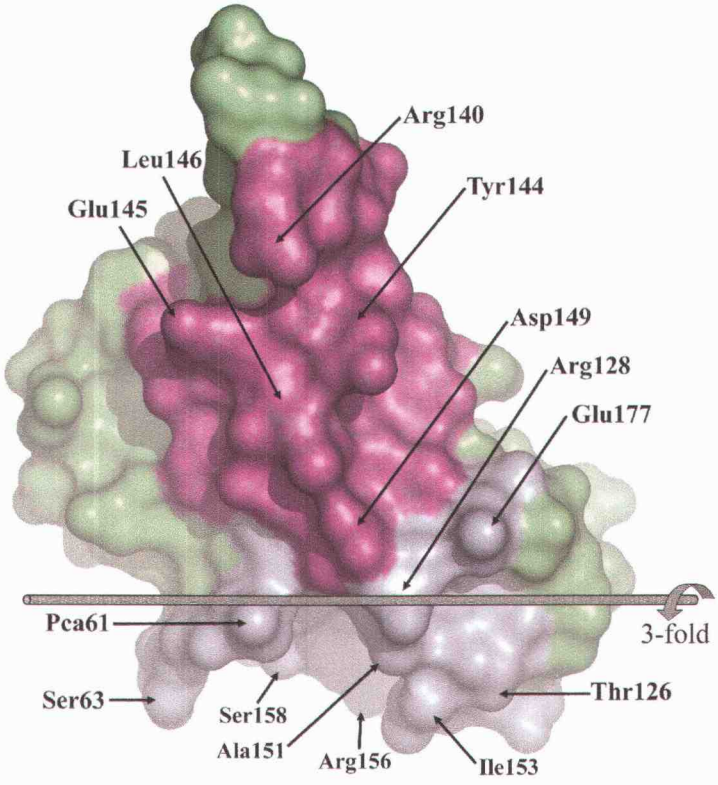


Figure 3.4 (continued)

(b)



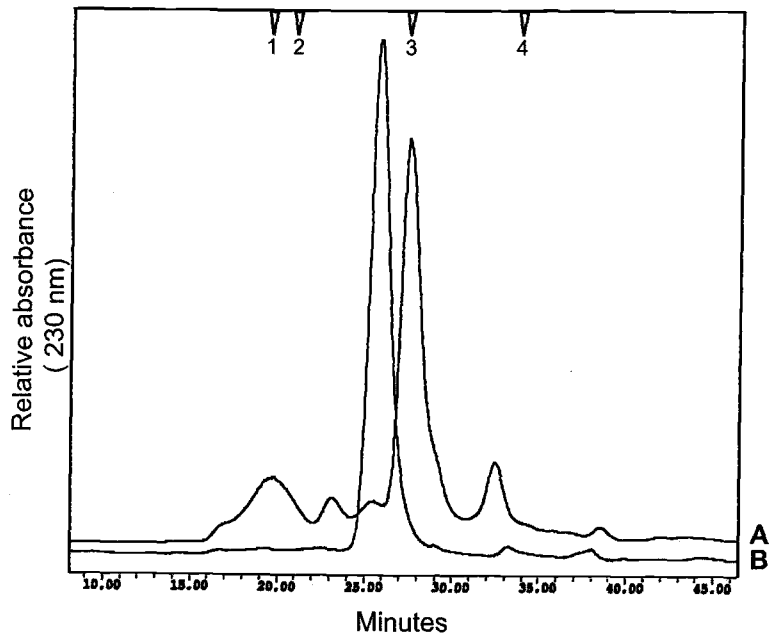
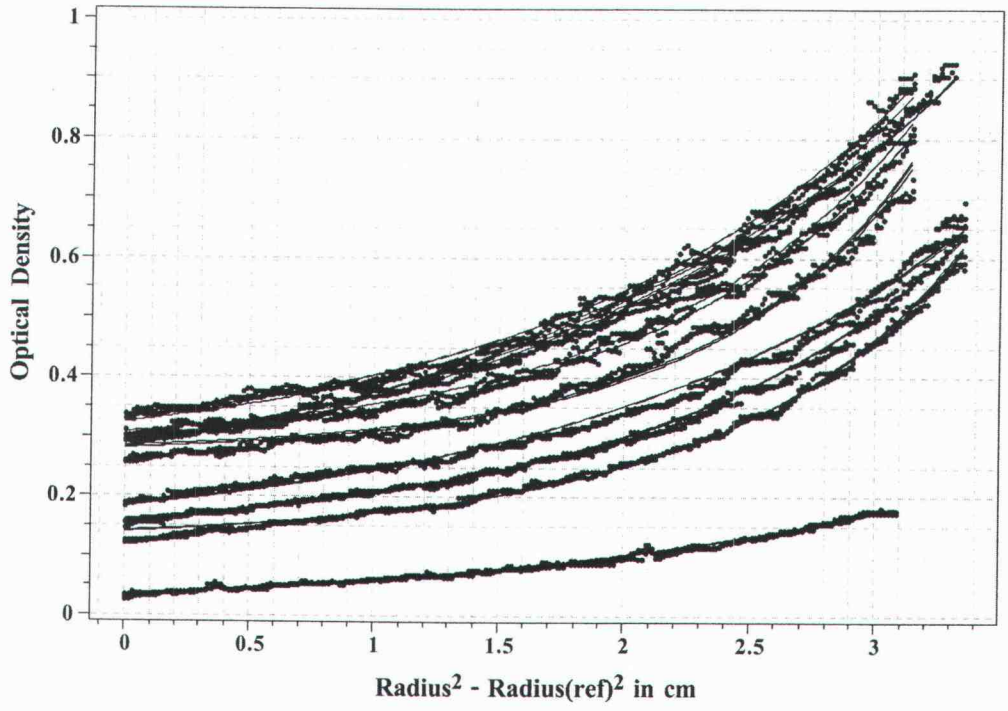


Figure 3.5. **Gel filtration analysis.** Size-exclusion chromatography elution profiles of ToxA (curve A) and His-tagged proToxA (curve B) shows the existence of predominantly monomeric species. Arrows above the profiles indicate the elution time of the standard proteins: Peak 1-Bovine serum albumin (67 kDa), Peak 2-Ovalbumin (43 kDa), Peak 3-Rnase A (13.7 kDa), and Peak 4-Aprotinin (6.5 kDa).

Figure 3.6. **Sedimentation equilibrium analysis.** (a) The 30 data sets resulting from three independent scans of four different ToxA concentrations at three different speeds are shown as dots and the curves are the best global fit of all of the data to a monomer-trimer model. (b) The residuals of the fits for each of the 30 data sets. Similarly, (c) shows the 9 data sets from three independent scans of proToxA at three different speeds with the data fit to a single-component ideal model, and (d) shows the residuals of the fit.

Figure 3.6 (continued)

(a)



(b)

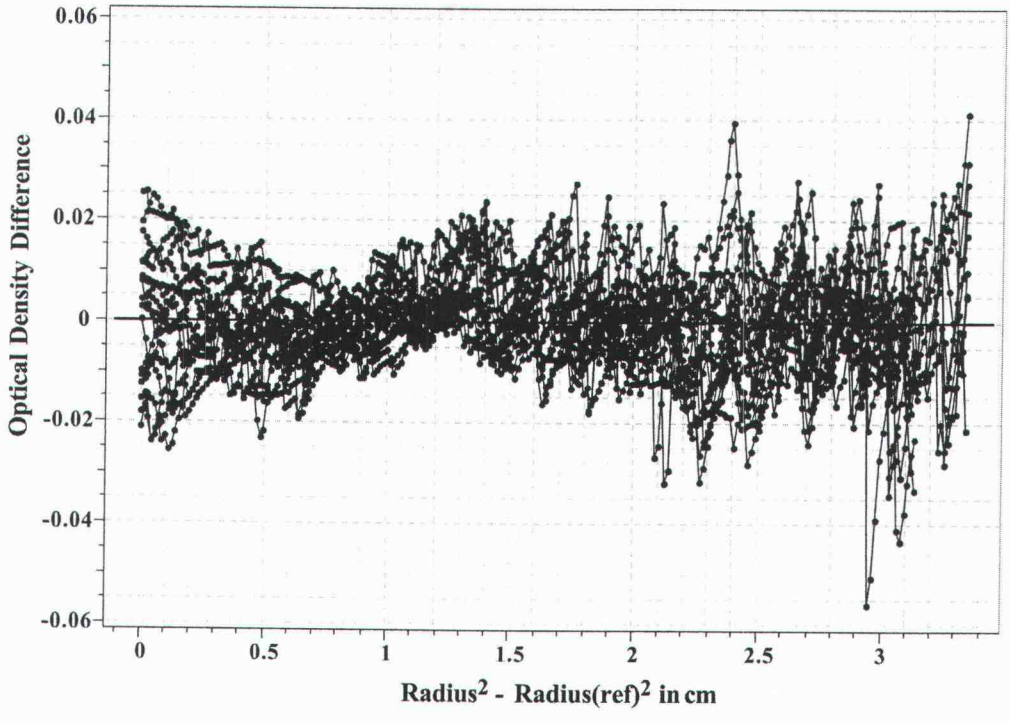
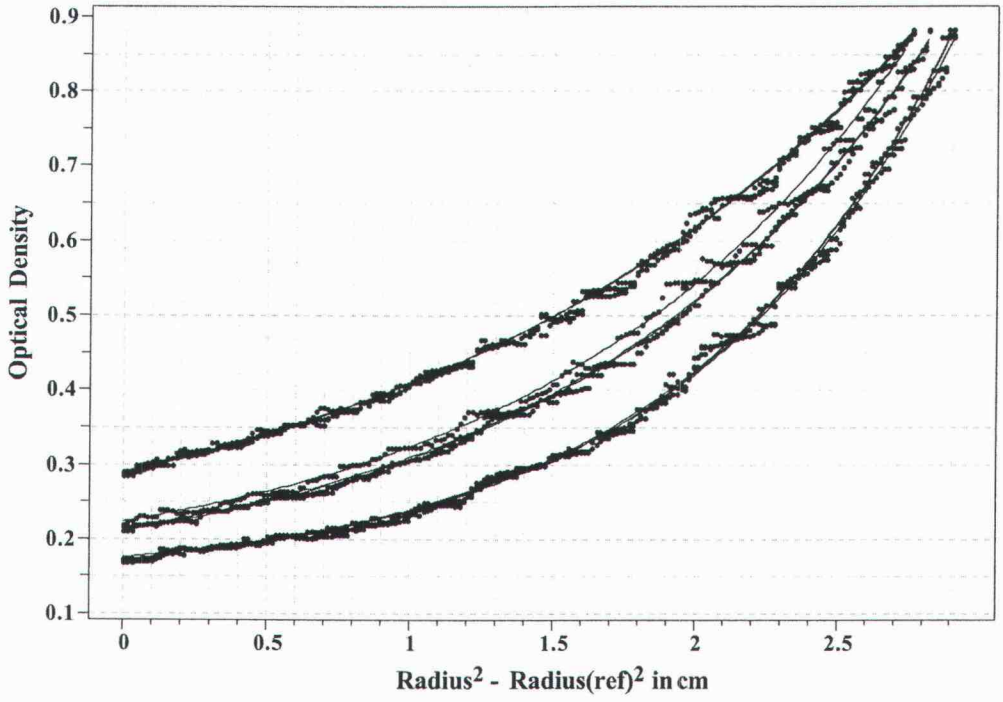
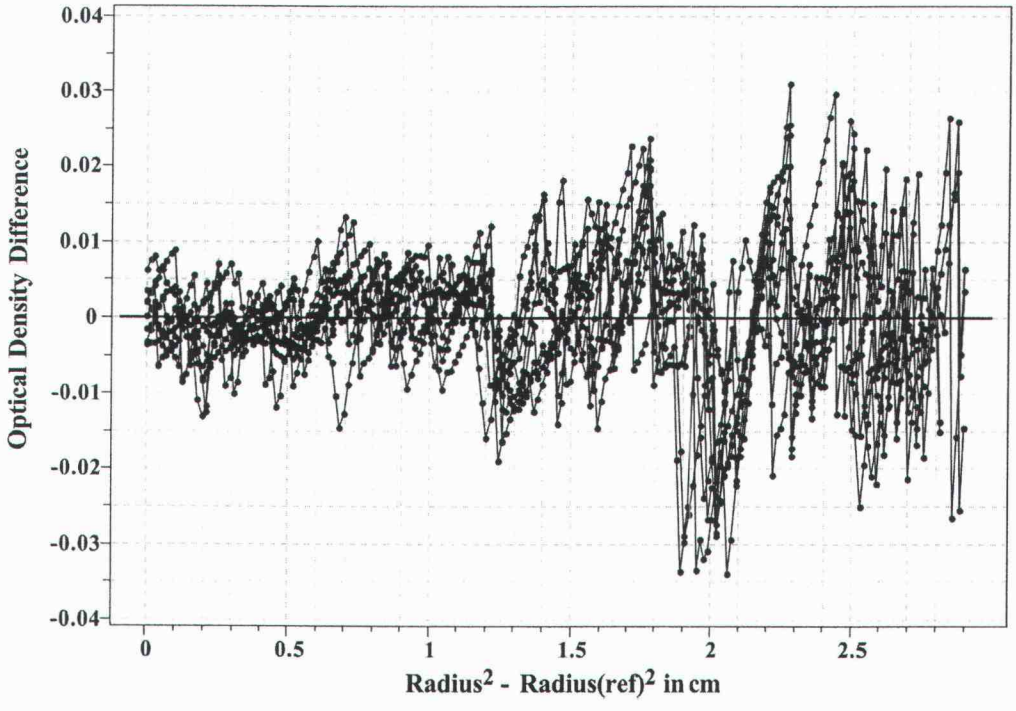


Figure 3.6 (continued)

(c)



(d)



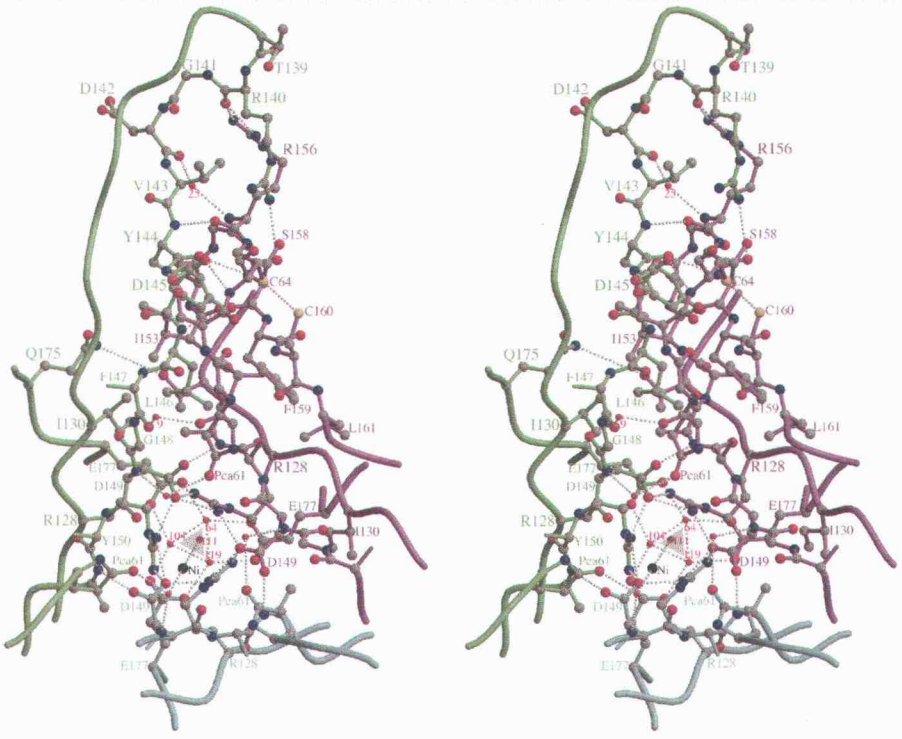


Figure 3.7. **The trimerization interface of ToxA.** A stereo representation showing the interactions at one of the trimer interfaces. The orientation and the coloring scheme are identical to that of Figure 3.4a and the 3-fold axis is highlighted by a gray triangle. The residues involved in the trimer interactions are represented as ball-and-stick models. To maintain clarity the side chain atoms of Phe147, Tyr150, Leu152 and Gln154 are not shown and residues Thr126, Ala151, Leu152, Gln154, Gly155, Gly157, waters 31 and 41 are not labeled. Arg128, Asp149, and Glu177 from all three chains form a symmetrical hydrogen-bonding network with water molecules and a Ni^{+2} at the 3-fold. The burial of many hydrophobic residues also contributes to the stability of the trimer interface. Hydrogen bonds bridging the different monomers are represented by dotted gray lines. Residues having ≥ 40 Å^2 surface area buried at the trimer interface are (in parentheses are buried surface areas in Å^2): Pca61 (62), Ser63 (47), Thr 126 (55), Arg128 (65), Arg140 (91), Tyr144 (123), Glu145 (74), Leu146 (97), Asp149 (48), Ala151 (55), Ile153 (117), Arg156 (115), Ser158 (55) and Glu177 (74). The figure was prepared using Molscript [27] and Raster3D [26].

Furthermore, the recombinant proToxA molecule that is fully active behaves as a stable monomer according to both gel filtration and sedimentation equilibrium (Figures 2.5 and 2.6c,d). Because ToxA would appear to act as a monomer, it is important to describe the trimer interactions in order to understand how the monomer structure may differ from the trimer that is seen in the crystal. Also, since the interaction of ToxA with other proteins may enhance trimer formation, it is still possible that ToxA acts as a trimer sometime during its function.

The overall shape of the trimer is that of a three-blade pinwheel (Figure 3.4a). The trimer interface buries ca. 1200 $Å^2$ of surface area per monomer and involves a concave broad surface of ToxA binding like a vise around a narrow spine of another monomer that is created largely by β6 (Figure 3.4b). In all, 37 residues participate to create a network of salt bridges near the three-fold (involving Arg128, Asp149, and Glu177 from all three chains), a β-sheet interaction between strand β6 (residues Val143 through Leu146) of one chain and strand β7 (Arg156 through Phe159) of a neighboring chain, and the burial of many non-polar side chains (see Figure 3.7 for details). In the form-II structure, hydrogen bonds involving the side chains of Arg140 and Arg156 further tie down the tip of the RGD-containing loop (Asn136 through Val143).

What are the implications of this for the monomer structure? First, because the β-ribbon involving the ends of the strands β5 and β6 (Trp134 – Asn136 and Asp142 – Tyr144) extend beyond the main β-sandwich core (see Figures 2.2b and 2.2b), we predict that in the absence of the interchain vise-like interactions stabilizing this conformation, these residues will form a much more mobile and larger (up to 11-residue) loop. A second expectation is that the N-terminal region (residues 61-64) will also be much less ordered in a monomer.

3.3.4 **Rationalization of existing mutant data**

Combining the mature ToxA structure with existing mutagenesis data allows us to gain insight into ToxA structure-function relationships. To date, published mutation analyses of ToxA include 20 single mutations [18-20] that targeted the residues in the RGD-containing segment (Asn136 through Phe147) and residues that

Table 3.2. **Correlation of ToxA mutations with structure.** [a]

Mutation	Structural location	Accessible side chain area (Å^2)	
Active			
S86G	α1	60	This study [b]
G96A	β2	11	Meinhardt *et al.* [20]
T115A	β5	32	Manning *et al.* [20]
T123A	β4-β5 loop	65	Manning *et al.* [20]
T126A	β4-β5 loop	61	Manning *et al.* [20]
V138A	β5-β6 loop	124	Manning *et al.* [20]
T167A	β7-β8 loop	15	Manning *et al.* [20]
G168A	β7-β8 loop	34	Manning *et al.* [20]
Impaired [c]			
Structurally disruptive			
G93A	β2	25	This study [b]
T132A	β5	3	Manning *et al.* [20]
F147A	β6	33	Manning *et al.* [20]
Structurally nondisruptive			
G62A	N-terminus loop	26	This study [b]
C64G	N-terminus loop	15	Tuori *et al.* [19]
N136A	β5	73	Manning *et al.* [20]
T137A	β5	45	Tuori *et al.* [19]
T139A	β5-β6 loop	92	Manning *et al.* [20]
R140A	β5-β6 loop	137	Manning *et al.* [20]
G141A	β5-β6 loop	21	Manning *et al.* [20], Meinhardt *et al.* [20]
D142E	β6	83	Manning *et al.* [20], Meinhardt *et al.* [20]
V143A	β6	25	Manning *et al.* [20]
Y144A	β6	109	Manning *et al.* [20]
E145A	β6	91	Manning *et al.* [20]
L146A	β6	66	Manning *et al.* [20]

Table 3.2. (continued)

[a] All reported mutants were created and tested in the context of proToxA, and behaved well during expression and purification.
[b] Residues Gly62 and Gly93 are potential myristoylation sites; S86G was a PCR induced mutation.
[c] T132A and L146A are partially active; all other mutants in this category were scored as inactive.

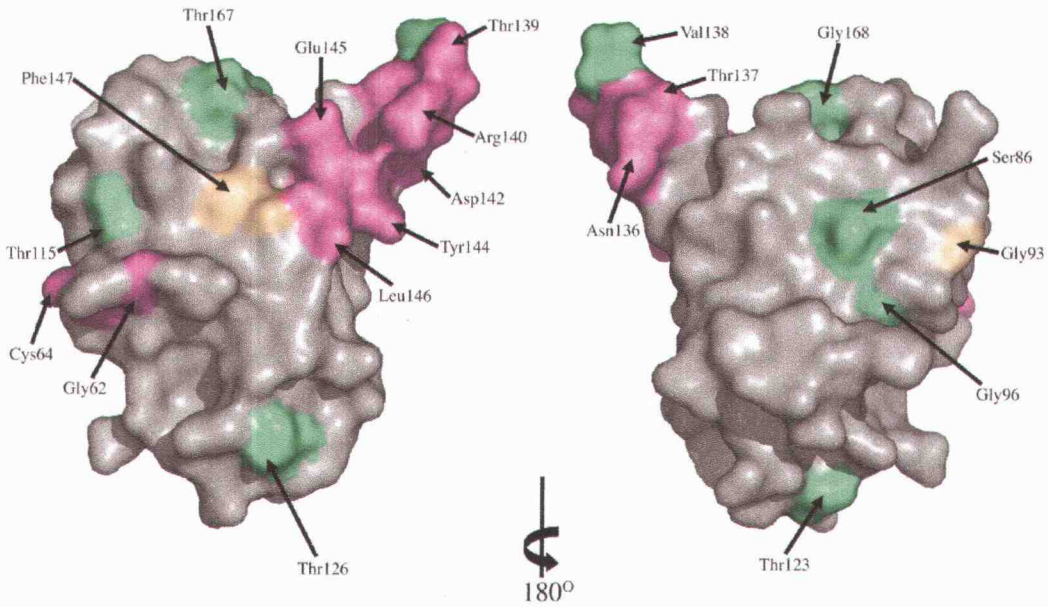


Figure 3.8. **Mapping of site-directed mutagenesis results onto the ToxA structure.** The molecular surface of ToxA is colored to highlight the sites of fully active mutants (green), partially active or inactive mutants that are expected to disrupt the structure (yellow) and partially active or inactive mutants that are expected to be non-disruptive (purple). All residues except Thr132, Gly141, and Val143 are visible in these views. The surface on the left is in the same orientation as Figure 3.2b while the surface on the right is rotated 180° around the vertical to show the other side of the protein. The figure was generated using Pymol (www.pymol.org).

represented potential phophorylation sites (Thr residues 115, 123, 126, 132, 139, and 167) or myristoylation sites (Gly residues 62 and 93) (Table 3.2). In addition to these mutants, we present here data for a further three mutated residues (see Table 3.2 and Methods). Of all the mutants, eight are active, two are partially active, and 13 are inactive. All of the mutation sites have unambiguous density in one or both crystal forms and Figure 3.8 shows how these mutation sites map onto the structure.

Eight residues (Ser86, Gly96, Thr115, Thr123, Thr126, Val138, Thr167 and Gly168) when mutated to Ala or Gly resulted in variants that retained toxicity. The side chain of each of these residues is solvent exposed, where the mutation would not be expected to disrupt ToxA folding or stability. Their unperturbed activity indicates that these surface regions (green in Figure 3.8) are not crucially involved in receptor recognition or other interactions required for ToxA function.

In the case of partially and fully inactive variants, the location of each mutated residue allows us to distinguish whether that mutation would be expected to destabilize the protein (yellow in Figure 3.8) or not (purple in Figure 3.8). If destabilization is unlikely then that site is implicated as a potential point of direct involvement in receptor recognition or another interaction required for ToxA function. Among the partially and completely inactive mutants, three (Gly93, Thr132 and Phe147) are in environments where the mutation would be expected to disrupt the structure. The structure reveals that Gly93, located at the start of strand β2, adopts main chain torsion angles (ϕ, ψ = 105°, -22°) that are unfavorable for non-Gly residues, so the mutation to Ala would destabilize the local conformation. Thr132, one of the potential phosphorylation sites, is mostly buried in the ToxA monomer by nearby side chains of Trp134 and Gln175, and we expect mutation to Ala would disrupt these interactions and destabilize and/or alter ToxA structure. We conclude Thr132 plays a structural role, but since it is partially exposed we cannot rule out that it may play a direct role, such as to be phosphorylated *in planta*. For Phe147, the side chain is buried in the core of the monomer where mutation would likely disturb the structure.

The remaining twelve residues have partly or fully exposed side chains where, based on the folded protein structure, the mutations would not be expected to disrupt the structure. As seen in Figure 3.8, these sites are mostly located on one protrusion of the fold, which is the RGD-containing loop between β5 and β6. The only exceptions are Gly62 and Cys64. Interestingly, it is possible that these two residues represent a special case where the mutations would be compatible with the folded conformation, but they alter the folding process so that the native fold is never reached. Evidence that Cys64 is an important player in the folding process is that the disulfide involving Cys64 is properly formed during *in vivo* folding of proToxA, but does not properly reform during *in vitro* refolding of reduced ToxA [19]. Thus it is plausible that mutations of Cys64 and its neighbors (like Gly62) would interfere with folding.

The clustering of the inactive, non-disruptive mutations on the RGD containing loop is consistent with the original hypothesis [9,17-20] that ToxA may interact with plant cell integrin-like receptors in a manner analogous to mammalian proteins such as vitronectin and fibronectin. In mammals, recognition of RGD residues is highly specific and exquisite in a variety of contexts [29-32]. ToxA is similar in that even a subtle Asp142Glu mutation results in loss of recognition, but it is different in that not just the RGD but also many surrounding residues are also required for recognition. The interest of the RGD residues for ToxA function is further highlighted by the fact that interaction in animal cells of some RGD-containing proteins with integrins results in their internalization. In particular, this route is used by certain viruses as a cell entry mechanism [33-37]. While this phenomenon has not yet been seen in plants, that RGD-containing peptides are known to disrupt the plasma membrane-cell wall continuum [38-40] provides circumstantial evidence for the existence of integrin-like proteins in plants.

3.3.5 Comparison with mammalian RGD-proteins: a possible evolutionary relationship

In order to gain further insight from the ToxA structure, a search for structurally similar proteins was carried out using the program DALI [41]. The top matches were all proteins with β-sandwich folds but the Z-scores (highest 4.2) were well below the cutoff of 6.0 suggested by DALI as indicative of a reliable structural homolog. Interestingly, among the top scores were known RGD-containing mammalian proteins that are based on a fibronectin type III fold (FnIII) – also a β-sandwich structure. Therefore, the structure FnIII was compared to that of ToxA. A study of the topology diagrams revealed that the proteins could be made to have a similar topology through a simple circular permutation of ToxA (Figure 3.9a). The circular permutation event connects the N- and C-termini of ToxA and new termini are formed by the cleavage of the β6-β7 loop. In this comparison, strands β2 and β3 of ToxA are insertions relative to FnIII, and strand A of FnIII is an insertion relative to ToxA (Figure 3.9). Thus, after circular permutation, ToxA and FnIII can be aligned using a common six-stranded core of the same topology.

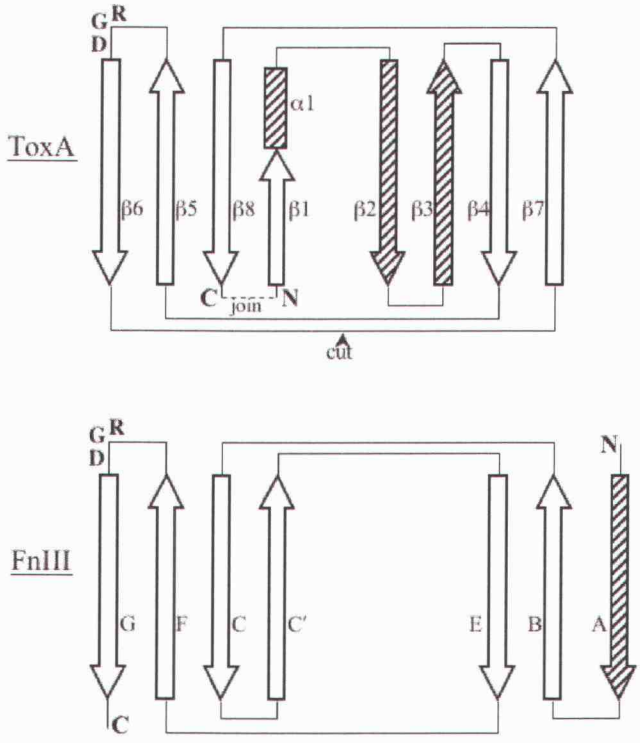
A new DALI search with a circularly permuted ToxA yielded the RGD-containing structures of the tenth domain of fibronectin [39] ($FnIII_{10}$) and the third domain of tenascin [42,43] ($FnIII_3$) among the top matches (Z-scores ≈ 7.5). Since the structures of $FnIII_{10}$ and $FnIII_3$ are highly similar (rmsd of 1.1 Å for 73 C_α atoms), the former is used as a representative for comparisons with ToxA. The structural overlay of ToxA with $FnIII_{10}$ is shown in Figure 3.9b, and a particularly striking feature of the overlay is that the RGD containing loops are at a common position in the structure (in the β5-β6 loop) (Figure 3.9).

Structural similarity after circular permutation combined with the spatial conservation of key functional residues is a well-documented phenomenon [44,45] that generally leads to the proposal that the proteins in question are homologs even in the face of no significant overall sequence similarity (e.g. Murzin *et al.* [46]). The match of the position of the RGD-containing loop could easily be a result of convergent evolution, but nevertheless, the similar fold and positional conservation of the RGD-containing loop in the six-stranded circularly permuted ToxA and FnIII plus their

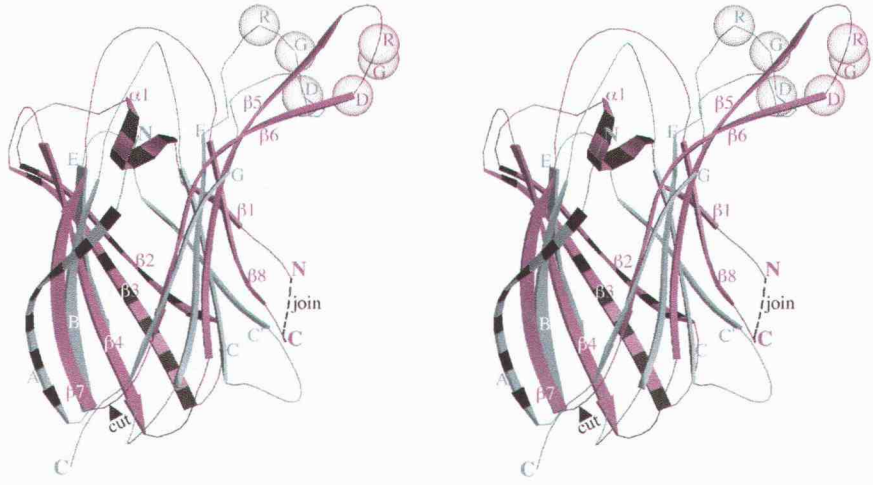
Figure 3.9. **Circular permutation of ToxA.** (a) Topological comparison of ToxA and the tenth domain of fibronectin ($FnIII_{10}$) shown as a representative for the FnIII domain using the strand nomenclature proposed by Leahy *et al.* [42]. Open arrows and rectangles represent elements that are common to both the structures and striped ones represent elements unique to each structure. The positions of the conserved RGD residues are also indicated. (b) Stereoview of the ToxA (pink) – $FnIII_{10}$ (blue) comparison, with the unique elements of each chain indicated (striped). The Cα positions of RGD residues are shown as semi-transparent spheres. The view is identical to that of Figure 3.2b. The secondary structural elements are labeled in both panels. The connection of the N- and C-termini (dashed line) and the cleaving of the β6-β7 loop (solid arrowhead) define the changes involved in the circular permutation.

Figure 3.9

(a)


ToxA
G
D
R
α1
β6
β5
β8
β1
β2
β3
β4
β7
C
join
N
cut
FnIII
G
D
R
N
G
F
C
C′
E
B
A
C

(b)


α1
β5
β6
β1
β2
β8
β3
β4
β7
N
join
C
cut
C

similar functions in terms of binding to a cell surface receptor provides an attractive case for the existence of a common ancestor, especially given the evidence that the plant cell wall is the functional equivalent of the mammalian extracellular matrix [47,48]. Isolation of the plant protein that serves as a ToxA receptor and structural and functional characterization of fibronectin-like proteins from the cell wall, such as the putative one recently isolated [49], may provide clues to the evolutionary origins and mode of action of ToxA.

3.4 Methods

3.4.1 Purification and crystallization

ToxA was purified from culture filtrates as previously described [8]. A procedural modification involved dialyzing the protein against Epure water (Barnstead) instead of 30 mM TRIS, pH 9.0 and the concentration was determined using an extinction coefficient $\varepsilon_{280} = 18\ mM^{-1}cm^{-1}$ [9]. Prior to crystallization, the protein was diluted to 1.5 mg/ml in 0.01 M HEPES, pH 7.0 and stored at 4 °C.

Two crystal forms of ToxA were grown at room temperature using the hanging-drop vapor diffusion method. Form-I crystals exhibited cubic morphology and grew from a 1:1 drop using a reservoir solution containing 0.5 M $(NH_4)_2SO_4$, 15% dioxane, and 0.1 M MES, pH 6.5. Crystals grew over two weeks to a final size of ca. $(0.15\ mm)^3$. For data collection, the cubic crystals were flash-frozen in liquid nitrogen after a 2 min incubation in a cryo-protectant consisting of the reservoir plus 15% glycerol. Form-II crystals were obtained by mixing equal volumes of the protein stock with a reservoir of 7.5% PEG 400 (v/ v), 0.2 M $NiCl_2$, and 0.1 M HEPES-Na, pH 7.0. These tetragonal-shaped crystals appeared within 24 hours and grew in one week to a size of ca. 0.15 x 0.15 x 0.10 mm^3. Divalent cations were essential for crystallization with Co^{+2}, Ca^{+2}, Cd^{+2} ions able to substitute for Ni^{+2}. For data collection, the form-II crystals were progressively soaked for 5 min each in a step-wise manner through solutions of 15, 22.5 and 30% v/v (final concentration) PEG 400 plus reservoir and then flash-frozen in liquid nitrogen.

3.4.2 **Data collection**

For structure determination, highly redundant, single-wavelength anomalous dispersion (SAD) data were collected using a form-I crystal and in-house, Cu-Kα radiation (Rigaku RU-H3R rotating anode operating at 50 kV, 100 mA and a R-Axis IV image plate detector; $\lambda = 1.54$ Å, $\Delta\phi = 1°$, 540 15 min images) (Table 3.1). For structure refinements, 1.65 Å resolution native data were collected at beamline 8.2.1 of the Advanced Light Source (ALS; Berkeley National Laboratory; high resolution pass: $\lambda = 1.0$ Å, $\Delta\phi = 1°$, 40 20 s images; and low resolution pass: $\lambda = 1.0$ Å, $\Delta\phi = 1°$, 40 5 s images) for the form-I crystals and 1.90 Å data ($\lambda = 1.0$ Å, $\Delta\phi = 1°$, 160 15 s images) were collected for the form-II crystals. A low-resolution, in-house data set extending to 3 Å ($\lambda = 1.54$ Å, $\Delta\phi = 1°$, 99 15 min images) was also collected on a form-II crystal to confirm the presence of a disulfide bond. All data sets were processed using the HKL suite of programs [50] (Table 3.1).

The form-I crystals belong to the cubic space group $P2_13$ with unit cell dimensions a = b = c = 78.2 Å with one molecule in the asymmetric unit and a solvent content of 60%. Form-II crystals belong to the tetragonal space group $P4_122$ with unit cell dimensions of a = b = 74.5 Å and c = 137.3 Å, and have three molecules in the asymmetric unit with a solvent content of 50%.

3.4.3 **Structure determination and refinement**

The 13.2 kDa ToxA structure was determined from form-I crystals at 3.0 Å resolution using the weak anomalous signal arising from inherently present sulfur atoms. The expected Bijvoet ratio [51] from the four sulfur atoms – two Met and two Cys – present in the 13.2 kDa protein at 1.54 Å wavelength is ca. 0.8%. It was later observed that two of the four sulfurs (one Cys and one Met) were disordered, but the presence of a single well-ordered sulfate ion in the asymmetric unit led to three sulfurs contributing to the phasing of the 13.2 kDa protein.

In order to determine the low-resolution cutoff for structure solution, the 540-image data set was split into two consecutive 270-image data sets (images 1 through 270 and 271 through 540) and the anomalous correlation coefficients between the two sets were calculated using the program XPREP (Bruker Analytical X-ray Systems). At 3 Å resolution the correlation coefficients fell below 20% and this was chosen as a

cutoff. The program SHELXD [52] was used to locate the positions of the sulfurs. The 100 search trials resulted in a clear solution (ca. 50% success rate) with a correlation coefficient of 39% (all data) [53]. Heavy atom refinement (positions and occupancies) and phasing were carried out using the program MLPHARE [54]. Density modification by DM [55] resulted in experimental maps that were of superior quality. A more detailed analysis assessing how much diffraction data was required to solve the structure will be presented elsewhere (Sarma, G. N and Karplus, P. A, in preparation).

Prior to model building and refinement, 10% of the data were set aside for cross-validation. The automated model building program ARP/wARP [56] was used to extend the 3 Å phases to 1.65 Å and to build ca. 90% of the model. Further the program was used to pick waters using the following criteria: (1) a peak of $\geq 3\ \rho_{rms}$ in the F_o–F_c map and a peak of $\geq 1\ \rho_{rms}$ in the $2F_o$–F_c map, and (2) a distance of ≥ 2.5 Å and ≤ 3.5 Å to nearby hydrogen bond donor or acceptor. Manual rebuilding was carried out using O [57]. TLS refinement as incorporated in REFMAC [58] followed by positional and individual B-factor refinements resulted in a final R/R_{free} of 15.7%/ 18.3% respectively. The final model has 103 residues (61-62 and 78-178), 105 waters, 2 sulfates, with alternate conformations seen for Ser86, Ile95, Arg117, Ile131, Gln154, Arg173, Glu177, and for waters 28, 35, 37, 40, 55, 62, 64, 65. Residues 63 through 77 did not have any visible density and were not modeled.

For form-II structure solution, a crystallographic trimer generated using the form-I monomer (without waters) was used as a search model for molecular replacement at 4 Å using the direct Patterson search as implemented in the program CNS [59]. Cross-rotation and translation searches yielded a clear solution with a correlation coefficient of 0.550 versus 0.114 for the next best solution. The trimer was refined to 1.9 Å using REFMAC [58] with 10% of the data set aside for cross-validation. No NCS restraints were used during refinement. Waters were picked (using the criteria reported above) and alternate conformations were modeled for Ile95 in chain A, Arg113 in chain B, and waters 94 and 126. In one of the chains (chain B), a 6-residue stretch (residues 74-79) forms an additional β-strand that interestingly lies on a crystallographic two-fold axis and shows weak but clear main chain electron density. Thus, in the crystal, this loop shares the space with the exact same segment from another trimer in the crystal.

Accordingly, it was modeled with half occupancy (Figure 3.10). The weak density for the loop is consistent with the high temperature factors for this region of the protein (Figure 3.2d). Strong electron density peaks for Ni^{+2} ions were seen on the crystallographic 2-fold axes. Additional Ni^{+2} ions were located using anomalous difference maps calculated from another data set optimized for Ni anomalous signal (λ = 1.487 Å; data not shown). Occupancies for the Ni ions not on special positions were estimated by carrying out test refinement runs with occupancies of 0.2, 0.4, 0.6 and 0.8, and for each Ni, the occupancy that yielding a B-factor most similar to the surrounding protein atoms was chosen. The eight Ni^{+2} ions ranged in occupancy from 0.2 to 0.5. Unsuccessful attempts to solve the structure of the form-II crystal using Ni multiple wavelength anomalous dispersion (data not shown) could be explained by the low occupancies and high temperature factors of the Ni^{+2} ions.

Iterative cycles of model building and refinement yielded an R/R_{free} of 20.8%/ 26.4%. TLS refinement, with each monomer as a rigid body, lowered the R and R_{free} to 18.3% and 22.6% respectively. The final model has 315 amino acids (103 chain A, 109 chain B and 103 chain C), 165 waters and 8 Ni^{+2} ions with residues 65-79 (chains A and C) and residues 65-73 (chain B) not modeled due to weak electron density.

Since discretely ordered solvent molecules are an integral part of the structure and function of the protein, the preferred hydration sites [60] in each crystal form were numbered according to their electron density strength (Water 1 being the highest) from 1 to 105 in form-I and to 165 in form-II. In forms-I and -II crystals, respectively, 84 and 141 of the sites are present in the first solvent shell (making at least one hydrogen bond with the protein).

3.4.4 **Structural comparison and analyses**

Secondary structure assignments were made using DSSP [61] and also by visual inspection. Hydrogen bonding geometries were calculated using HBPLUS [62]. Structure-based sequence alignments were done using HCORE (Karplus, P.A, unpublished) with a distance cutoff of 3.0 Å.

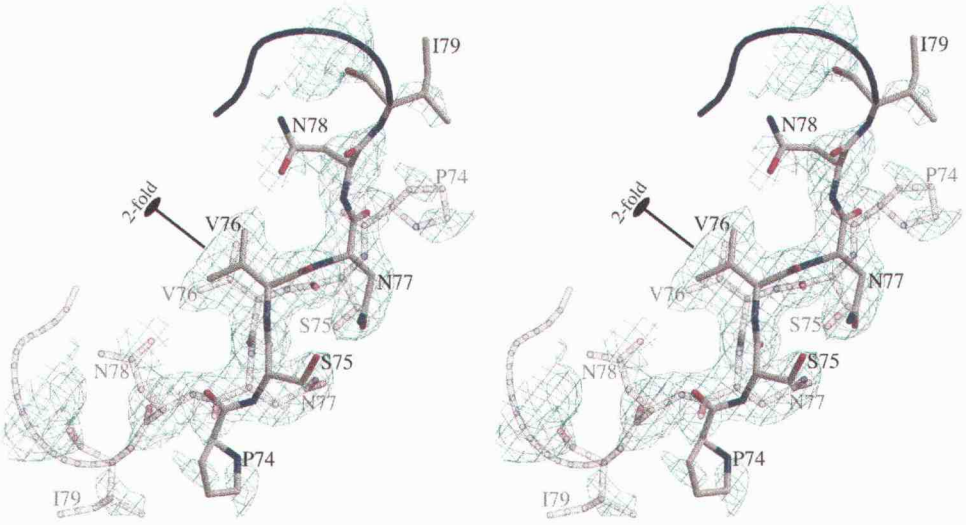


Figure 3.10. **Residues on a crystallographic 2-fold axis.** A stereo diagram showing the $2F_o - F_c$ electron density contoured at 0.7 ρ_{rms} at the crystallographic 2-fold axis near chain B of the form-II crystals. The residues that have been modeled with half-occupancies are indicated as stick models and labeled. The identical residues from the crystallographically related molecule that share the space in the crystal are shown as a semi-transparent model and also labeled. One chain makes a parallel interaction with strand β2 and the other is anti-parallel. The way this disorder works is that half of the molecules in the crystal have one of the chains ordered and half have the other chain ordered. This means that the crystallographic symmetry is broken locally, but is maintained for the crystal as a whole. The 2-fold axis is shown for reference. The figure was prepared using BobScript [25] and Raster3D [26].

3.4.5 Site-directed mutagenesis

Of the three mutants newly reported in this study, G62A and G93A, were generated by site-directed mutagenesis using the following primers (one of the complementary pair): G62A – 5′ CTACAGGAGCGGCAGGCAAGCTGCATGTCAATC 3′; G93A – 5′ CTCGGACGACTTGCTGCTATAGGCTCGTGG 3′. The S86G mutation was generated inadvertently during PCR cloning of the proToxA into pTRC99-HPro [19]. All mutant proteins were expressed in *Escherichia coli* and assayed for bioactivity as described by Manning *et al.* [20].

3.4.6 Gel filtration

The existence of the ToxA monomer in solution (0.5 mg/ml in 100 mM NaCl, 10 mM sodium phosphate, pH 7.0) was confirmed by gel-filtration chromatography using a Superdex-75 10/300 GL size-exclusion column (Amersham Biosciences) at a flow rate of 0.5 ml/min. Using identical conditions, proToxA at 0.17 mg/ml was also chromatographed. Fractions were collected every minute and activity was assayed by infiltrating peak fractions into the sensitive wheat cultivar, Katepwa. For ToxA, the major active fraction and over 80% of the A_{230} absorbing material ran at an elution volume corresponding to a molecular weight of ca. 13 kDa and for proToxA, > 95% of the absorbing material ran at an elution volume corresponding to a molecular weight of ca. 19 kDa (Figure 3.5).

3.4.7 Analytical ultracentrifugation

Sedimentation equilibrium analytical ultracentrifugation experiments were performed with a Beckman Optima XL-A analytical ultracentrifuge at 20 °C using a four-hole AN-60Ti rotor and double sector charcoal/ Epon-filled centerpieces. Buffer densities and viscosity corrections were made according to data published by Laue *et al.* [63]. The partial specific volume of ToxA was estimated from the protein sequence to be 0.727 cm^3/g, using the method of Perkins [64].

The centrifugation experiments were carried out according to the protocol described by Ausio *et al.* [65] in a solution containing 100 mM NaCl, 50mM sodium phosphate, pH 7.0. Four 120 µL samples of ToxA were sedimented to equilibrium: 3.4

μM and 10.1 μM samples were run at 22000 and 30000 rpm, and 7.8 μM and 23.4 μM were run at 22000, 26000 and 30000 rpm. Scans were collected with absorbance optics at wavelengths between 230 and 280 nm. The radial step size was 0.001 cm, and each *c versus r* datapoint was the average of 15 independent measurements. Wavelengths were chosen so that no points exceeded an absorbance of 1. Equilibrium data spanning this concentration range were examined by global fitting using UltraScan software for the Linux operating system (www.ultrascan.uthscsa.edu). Equilibrium data were fit to multiple models using global fitting of all data [66]. The monomer-trimer equilibrium fit resulted in a monomer of molecular mass of ca. 13.4 kDa, which is in good agreement with the molecular mass of 13.2 kDa calculated from the ToxA sequence. Fitting the data to monomer-dimer and monomer-dimer-trimer model also resulted in reasonable fits giving monomer molecular weight of 12.9 and 13.6 kDa, respectively. The monomer-trimer model had the best statistics, the lowest systematic errors based on inspection of the residual patterns, the monomer molecular weight is closest to the true value, and is consistent with the formation of trimer in the crystal.

For proToxA, one sample at a concentration of 33.8 μM was analyzed. The speeds and scans were done as for ToxA. The data fit well to a single-component ideal model that yielded a molecular mass of 19.7 kDa, exactly matching the calculated molecular mass of a monomer.

3.4.8 **Coordinates**

The coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org/pdb/) as entries 1ZLD (form-I) and 1ZLE (form-II).

3.4.9 **Acknowledgements**

We would like to thank Dr. Rick Faber for help in data collection and useful discussions, and Drs. Sonia Anderson and Dean Malencik (Oregon State University) for performing the analytical ultracentrifugation experiments. The invaluable help rendered by the staff at the ALS (beamline 8.2.1), particularly Dr. Christine Trame is also appreciated. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National

Laboratory. This work was supported by a National Science Foundation grant (MCB-0488665) to L.M.C. and P.A.K. and in part by a USDA-CSREES-NRICGP grant (2001-35319-10017) to L.M.C. This work was made possible in part by the Proteins and Nucleic Acids Core facility of the Environmental Health Sciences Center at Oregon State University (NIEHS grant P30 ES00210).

3.5 References

1. Wolpert, T. J., Dunkle, L. D. & Ciuffetti, L. M. (2002). Host-selective toxins and avirulence determinants: what's in a name? *Annu. Rev. Phytopathol.* 40, 251-285.

2. Walton, J. D. (1996). Host-selective toxins: agents of compatibility. *Plant Cell* 8, 1723-1733.

3. De Wolf, E. D., Effertz, R. J. & Francl, L. J. (1998). Vistas of tan spot research. *Can. J. Plant Pathol.* 20, 349-444.

4. Ciuffetti, L. M. & Tuori, R. P. (1999). Advances in the characterization of the *Pyrenophora tritici-repentis*-Wheat interaction. *Phytopathology* 89, 444-449.

5. Strelkov, S. E. & Lamari, L. (2003). Host-parasite interaction in tan spot [*Pyrenophora tritici-repentis*] of wheat. *Can. J. Plant Pathol.* 25, 339-349.

6. Ballance, G. M., Lamari, L. & Bernier, C. C. (1989). Purification and characterization of a host-selective necrosis toxin from *Pyrenophora tritici-repentis*. *Physiol. Mol. Plant Pathol.* 35, 203-213.

7. Tomas, A., Feng, G. H., Reeck, G. R., Bockus, W. W. & Leach, J. E. (1990). Purification of a cultivar-specific toxin from *Pyrenophora tritici-repentis*, causal agent of tan spot of wheat. *Mol. Plant-Microbe Interact.* 3, 221-224.

8. Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. (1995). Purification and immunological characterization of toxic components from cultures of *Pyrenophora tritici-repentis*. *Mol. Plant-Microbe Interact.* 8, 41-48.

9. Zhang, H., Francl, L. J., Jordahl, J. G. & Meinhardt, S. W. (1997). Structural and physical properties of a necrosis-inducing toxin from *Pyrenophora tritici-repentis*. *Phytopathology* 87, 154-160.

10. Strelkov, S., Lamari, L., Ballance, G. M. & Orolaza, N. P. (1998). Isolation and mode of action of PTR chlorosis toxin from *Pyrenophora tritici-repentis*. In *Molecular Genetics of Host-specific Toxins in Plant Disease* (Kohmoto, K. & Yoder, O. C., eds.), pp. 137-138. Kluwer Academic Publishers, Netherlands.

11. Effertz, R. J., Meinhardt, S. W., Anderson, J. A., Jordahl, J. G. & Francl, L. J. (2002). Identification of a chlorosis-inducing toxin from *Pyrenophora tritici-repentis* and the chromosomal location of an insensitivity locus in wheat. *Phytopathology* 92, 527-533.

12. Faris, J. D., Anderson, J. A., Francl, L. J. & Jordahl, J. G. (1996). Chromosomal location of a gene conditioning insensitivity in wheat to a necrosis-inducing culture filtrate from *Pyrenophora tritici-repentis*. *Phytopathology* 86, 459-463.

13. Anderson, J. A., Effertz, R. J., Faris, J. D., Francl, L. J., Meinhardt, S. W. & Gill, B. S. (1999). Genetic analysis of sensitivity to a *Pyrenophora tritici-repentis* necrosis-inducing toxin in durum and common wheat. *Phytopathology* 89, 293-297.

14. Gamba, F. M., Lamari, L. & Brule-Babel, A. L. (1998). Inheritance of race-specific necrotic and chlorotic reactions induced by *Pyrenophora tritici-repentis* in hexaploid wheats. *Can. J. Plant Pathol.* 20, 401-407.

15. Stock, W. S., Brule-Babel, A. L. & Penner, G. A. (1996). A gene for resistance to a necrosis-inducing isolate of *Pyrenophora tritici-repentis* located on 5BL of *Triticum aestivum* cv. Chinese spring. *Genome* 39, 598-604.

16. Ballance, G. M., Lamari, L., Kowatsch, R. & Bernier, C. C. (1996). Cloning, expression and occurrence of the gene encoding the Ptr necrosis toxin from *Pyrenophora tritici-repentis*. *Mol. Plant Pathol. Online*, publication/1996/1209.ballance.

17. Ciuffetti, L. M., Tuori, R. P. & Gaventa, J. M. (1997). A single gene encodes a selective toxin causal to the development of tan spot of wheat. *Plant Cell* 9, 135-144.

18. Manning, V. A. & Ciuffetti, L. M. (2005). Localization of Ptr ToxA produced by *Pyrenophora tritici-repentis* reveals protein import into wheat mesophyll cells. *Plant Cell*. Published online on September 30, 2005; 10.1105/tpc.105.035063.

19. Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. (2000). Heterologous expression of functional Ptr ToxA. *Mol. Plant-Microbe Interact.* 13, 456-464.

20. Meinhardt, S. W., Cheng, W., Kwon, C. Y., Donohue, C. M. & Rasmussen, J. B. (2002). Role of the arginyl-glycyl-aspartic motif in the action of Ptr ToxA produced by *Pyrenophora tritici-repentis*. *Plant Physiol.* 130, 1545-1551.

21. Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* 5, 802-810.

22. Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure* 4, 897-904.

23. Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* 56 (Pt 3), 328-41.

24. Dunlop, K. V., Irvin, R. T. & Hazes, B. (2005). Pros and cons of cryocrystallography: should we also collect a room-temperature data set? *Acta Crystallogr. D Biol. Crystallogr.* 61, 80-7.

25. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* 55 (Pt 4), 938-40.

26. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* 277, 505-524.

27. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Cryst.* 24, 946-950.

28. Meinhardt, S. W., Zhang, H., Effertz, R. J. & Francl, L. J. (1998). Characterization of additional peaks of necrosis activity from *Pyrenophora tritici-repentis*. *Can. J. Plant Pathol.* 20, 436-437.

29. Xiong, J. P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S. L. & Arnaout, M. A. (2002). Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 296, 151-5.

30. Xiao, T., Takagi, J., Coller, B. S., Wang, J. H. & Springer, T. A. (2004). Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. *Nature* 432, 59-67.

31. Springer, T. A. & Wang, J. H. (2004). The three-dimensional structure of integrins and their ligands, and conformational regulation of cell adhesion. *Adv. Protein Chem.* 68, 29-63.

32. Takagi, J., Strokovich, K., Springer, T. A. & Walz, T. (2003). Structure of integrin alpha5beta1 in complex with fibronectin. *EMBO J.* 22, 4607-15.

33. Hynes, R. O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell* 110, 673-87.

34. Memmo, L. M. & McKeown-Longo, P. (1998). The alphavbeta5 integrin functions as an endocytic receptor for vitronectin. *J. Cell Sci.* 111 (Pt 4), 425-33.

35. Marjomaki, V., Pietiainen, V., Matilainen, H., Upla, P., Ivaska, J., Nissinen, L., Reunanen, H., Huttunen, P., Hyypia, T. & Heino, J. (2002). Internalization of echovirus 1 in caveolae. *J. Virol.* 76, 1856-65.

36. Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). The structure of the human adenovirus 2 penton. *Mol Cell.* 17, 121-35.

37. Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. (1993). Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* 73, 309-19.

38. Mellersh, D. G. & Heath, M. C. (2001). Plasma membrane-cell wall adhesion is required for expression of plant defense responses during fungal penetration. *Plant Cell* 13, 413-424.

39. Canut, H., Carrasco, A., Galaud, J. P., Cassan, C., Bouyssou, H., Vita, N., Ferrara, P. & Pont-Lezica, R. (1998). High affinity RGD-binding sites at the plasma membrane of *Arabidopsis thaliana* links the cell wall. *Plant J.* 16, 63-71.

40. Senchou, V., Weide, R., Carrasco, A., Bouyssou, H., Pont-Lezica, R., Govers, F. & Canut, H. (2004). High affinity recognition of a Phytophthora protein by Arabidopsis via an RGD motif. *Cell Mol. Life Sci.* 61, 502-9.

41. Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* 26, 316-9.

42. Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 258, 987-91.

43. Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E. & Ely, K. R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.* 236, 1079-92.

44. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-40.

45. Jung, J. & Lee, B. (2001). Circularly permuted proteins in the protein structure database. *Protein Sci.* 10, 1881-6.

46. Murzin, A. G. (1998). Probable circular permutation in the flavin-binding domain. *Nat. Struct. Biol.* 5, 101.

47. Kohorn, B. D. (2000). Plasma membrane-cell wall contacts. *Plant Physiol.* 124, 31-8.

48. Baluska, F., Samaj, J., Wojtaszek, P., Volkmann, D. & Menzel, D. (2003). Cytoskeleton-plasma membrane-cell wall continuum in plants. Emerging links revisited. *Plant Physiol.* 133, 482-91.

49. Pellenc, D., Schmitt, E. & Gallet, O. (2004). Purification of a plant cell wall fibronectin-like adhesion protein involved in plant response to salt stress. *Protein Expr. Purif.* 34, 208-14.

50. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* 276, 307-26.

51. Hendrickson, W. A. & Teeter, M. M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* 290, 107-113.

52. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1772-9.

53. Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). In-house measurement of the sulfur anomalous signal and its use for phasing. *Acta Crystallogr. D Biol. Crystallogr.* 59, 688-96.

54. Otwinowski, Z. (1991). Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds.), pp. 80-6. Warrington: Daresbury Laboratory.

55. Cowtan, K. D. & Zhang, K. Y. (1999). Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* 72, 245-70.

56. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458-63.

57. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47 (Pt 2), 110-9.

58. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240-255.

59. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54 (Pt 5), 905-21.

60. Karplus, P. A. & Faerman, C. (1994). Ordered water in macromolecular structure. *Current Opinion in Structural Biology* 4, 770-776.

61. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-637.

62. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238, 777-93.

63. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds.), pp. 90-125. Royal Society of Chemistry, Cambridge, England.

64. Perkins, S. J. (1986). Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering matchpoints and 280-nm

absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* 157, 169-80.

65. Ausio, J., Malencik, D. A. & Anderson, S. R. (1992). Analytical sedimentation studies of turkey gizzard myosin light chain kinase and telokin. *Biophys. J.* 61, 1656-63.

66. Russell, T. R., Demeler, B. & Tu, S. C. (2004). Kinetic mechanism and quaternary structure of Aminobacter aminovorans NADH:flavin oxidoreductase: an unusual flavin reductase with bound flavin. *Biochemistry* 43, 1580-90.

3.6 **Footnotes**

Abbreviations used: rmsd, root-mean-square deviation; ρ_{rms}, root-mean-square electron density of map often reported as σ; FnIII, fibronectin type III; HSTs, host-selective toxins; ToxA, Ptr ToxA; rpm, revolutions per minute.

Chapter 4

In-house, low resolution, *de novo* sulfur SAD phasing of Ptr ToxA

Ganapathy N. Sarma and P. Andrew Karplus

Formatted for submission

4.1 Introduction

The first structure to be solved using the anomalous signal of sulfur atoms intrinsic to a protein was that of crambin in 1981 [1]. Six cysteine sulfur atoms contributed to the phasing of this 46-residue protein based on anomalous data collected using a Cu-Kα source. Further novel structures were not determined by sulfur SAD (single wavelength anomalous dispersion) until the late 1990s. Now, as recently reviewed by Dauter [2], advances in sub-structure solution, phasing algorithms and the availability of synchrotron beamlines have renewed the interest in S-SAD.

Sulfur anomalous data can be collected using synchrotrons or the fixed wavelength, in-house Cu-Kα or Cr-Kα source. While the anomalous signals are comparable in-house and at a synchrotron, collecting S-SAD data at a synchrotron is still the preferred choice. This is possibly because of the variable wavelength, the shorter times required to collect a highly redundant dataset, and the expectation of higher data quality. The wavelength range normally used to solve structures is 1.5 to 2.0 Å [3-12]. However, there is a trade-off: the intense and longer wavelength radiation at the synchrotrons enhance the anomalous signal but at the same time, result in faster crystal decay due to greater absorption of the lower energy X-rays. This problem of rapid crystal decay is partly alleviated by using the weaker intensity, in-house Cu-Kα X-ray source. Novel and test structures have been solved using this approach [13-18].

The first modern structure solution using in-house sulfur anomalous data used a fairly low-resolution cutoff for determining the sulfur sub-structure [19]. Contrary to these data, it has been now been generally accepted that higher resolution data is needed to solve the sulfur sub-structure with the majority of the structures mentioned above solved at resolutions better than 2.0 Å. More recently, this dogma is being questioned with structures being solved at lower resolution [6-8,15,20]. The expected Bijvoet ratio [1] in nearly all these cases were in the range of 0.9 – 3.1%. Simulations carried out using idealized data by Wang [21] suggested that a single disulfide per 120 residues with a Bijvoet ratio of 0.6 % would be enough to solve the structure. Thus in all the cases involving in-house sulfur SAD phasing, the Bijvoet ratios were much higher than would have been theoretically required, indicating that these proteins have a high sulfur content. Additionally, in all the cases, the exposure time per image

(where reported) was in the range of 5 – 120 seconds, which indicates that these crystals diffracted very well.

We have recently solved the structure of a 13.2 kDa protein, Ptr ToxA from *Pyrenophora tritici-repentis* using laboratory-based S-SAD data [22]. The anomalous signal of 0.68% [1] from three sulfur sites, 1 Met, 1 Cys and 1 sulfate ion, is only slightly higher than the limit of 0.6% predicted by Wang [21] as a theoretical minimum required to solve a protein structure using the anomalous signal from sulfurs. In solving the structure of Ptr ToxA, we were surprised to find that we were unable to solve the structure using sulfur SAD data collected at a synchrotron at $\lambda = 1.75$ Å, but were able to solve it using data collected using the laboratory Cu-Kα source. To further understand this unexpected result, we undertook a systematic analysis of the data quality requirements and the effects of resolution on sub-structure solution and phasing using only the sulfur anomalous signal. Our results reveal that phasing success is a complex function of data redundancy, quality and resolution cutoff.

4.2 Materials and Methods

4.2.1 Purification and crystallization

The purification of mature ToxA, crystallization and flash-freezing procedures have been described elsewhere [22]. The protein crystallizes in the cubic space group $P2_13$ with unit cell dimensions of $a = b = c = 78.2$ Å, a solvent content of 60% and one molecule in the asymmetric unit. The crystals had cubic morphology and dimensions of ca. $(0.15\ mm)^3$.

4.2.2 Data collection

As summarized in Table 4.1, two single wavelength anomalous diffraction (SAD) datasets were collected from frozen crystals in an attempt to solve the structure using the anomalous signal from the sulfur atoms. For both datasets, no special data collection strategies, such as collecting Bijvoets close in time, were employed.

The first dataset (SYNC), extending to 2.3 Å resolution, was collected from a single ToxA crystal using $\lambda = 1.75$ Å X-rays at beamline 8.2.1 of the Advanced Light Source (ALS; Berkeley National Laboratory; Table 4.1).

Table 4.1. **Data collection statistics** [a]

	SYNC	LAB
Data collection		
Number of images	82	1260
$\Delta\phi$	1°	1°
Wavelength (λ)	1.75 Å	1.54 Å
Exposure times per image	30 sec	20 min
Resolution limits	100 – 2.30 Å (2.38 – 2.30)	100 – 1.90 Å (1.97 – 1.90)
Unique observations [b, c]	7326	12754
Multiplicity [c]	9.5 (9.4)	112.3 (98.8)
Completeness (%)	100 (100)	99.9 (100)
Average I/ σ	13.6 (5.5)	47.4 (14.2)
Mosaicity	0.22°	0.42°
R_{meas} [d]	13.8 (43.3)	11.2 (66.8)
$R_{mrgd\text{-}F}$ [d]	8.9 (23.9)	2.5 (11.7)

[a] Numbers in parentheses correspond to values in the highest resolution shells.
[b] $|F|/\sigma|F| \geq 0$ cutoff was used for data processing.
[c] Bijvoets were not merged.
[d] *Rmeas* is the multiplicity weighted merging *R*-factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data [30].

The second dataset (LAB), extending to 1.9 Å resolution, was collected from a crystal that was not fully fresh, but had been exposed to synchrotron radiation (ALS; λ = 1.75 Å; ~ 13 min total exposure). For the in-house data collection, the crystal was rotated around the ϕ-axis to collect a total of 1260 images (~ 17 days of total exposure) using Cu-Kα radiation (Rigaku RU-H3R rotating anode operating at 50 kV, 100 mA and a Raxis IV image plate detector) (Table 4.1). After the first 360 images, the ϕ angle was offset by 0.5° for the next 360 images in order to fully collect reflections that were collected as partials in the earlier scan and the offset was repeated every 360 images. With the data represented as seven sequential 180° datasets, the offset strategy was such that the images of the 1st and the 5th, the 2nd and 6th, and the 3rd and 7th 180°-image sets covered an identical oscillation range (see Table 4.2). The exposure time per image was a compromise designed to collect highly redundant, medium resolution data even though longer exposure times could have yielded higher quality single pass data.

As reported in the Results section, the data were partitioned in various ways to create a series of test datasets (Tables 4.2 and 4.3). Each of these datasets was processed using the HKL suite of programs [23]. The unmerged diffraction data is available on request from P. Andrew Karplus.

4.2.3 **The reference model**

As described by Sarma *et al.* [22], the 13.2 kDa ToxA structure was determined at 3.0 Å resolution using the sulfur anomalous signal found in the first 540 images of the LAB data. The automated model building program ARP/wARP [24] was used to extend the 3.0 Å phases to 1.65 Å, and refinement yielded the final model with an R/ R_{free} of 15.7%/ 18.3% respectively [22]. This model (PDB entry 1ZLD) was subsequently used to calculate phases that are used here as the reference "true" phases.

4.2.4 **Sub-structure solution and phase determination**

For the systematic phasing analyses reported here, the various anomalous datasets were prepared for sub-structure solution using the program XPREP (Bruker Analytical X-ray Systems). Searches for the three sulfur positions were carried out with the program SHELXD [25] using 1000 search cycles. An approximate number of 'correct' solutions was estimated by choosing a correlation coefficient cutoff partway

Table 4.2. **Unique 180° datasets** [a]

	A	B	C	D	E	F	G
Oscillation range	0° – 180°	180° – 360°	0.5° – 180.5°	180.5° – 360.5°	0° – 180°	180° – 360°	0.5° – 180.5°
Unique observations	12578	12605	12542	12587	12479	12479	12348
Multiplicity [b]	21.0 (18.9)	21.0 (19.1)	21.0 (18.5)	21.0 (18.8)	20.7 (17.0)	20.8 (17.3)	20.0 (12.5)
Average I/ σ [c]	34.7 (8.0)	33.5 (7.7)	30.4 (6.0)	31.1 (6.1)	28.0 (4.5)	28.3 (4.3)	25.1 (2.8)
R_{meas} [d] (%)	7.3 (46.5)	7.0 (45.4)	7.8 (60.5)	7.8 (59.8)	8.8 (76.2)	8.6 (80.1)	9.8 (102.0)
$R_{mrgd\text{-}F}$ [d] (%)	4.5 (19.0)	4.2 (19.8)	5.3 (28.5)	5.0 (26.0)	6.4 (36.1)	6.3 (39.0)	8.1 (55.0)

[a] All data were processed to 1.90 Å resolution with the numbers in the parentheses indicating values in the highest resolution shell (1.97-1.90 Å).
[b] Bijvoets were not merged.
[c] $|F|/\sigma|F| \geq 0$ cutoff was used for data processing.
[d] R_{meas} is the multiplicity weighted merging R-factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data [30].

Table 4.3. **Data sets with increasing redundancies** [a]

	"30"	"60"	"90"	"120"	"150"	A	A-B	A-C	A-D	A-E	A-F	A-G
No. images	30	60	90	120	150	180	360	540	720	900	1080	1260
Unique obv.	12246	12473	12504	12531	12547	12578	12689	12719	12743	12749	12753	12754
Multiplicity[b]	3.5	7.0	10.5	14.0	17.5	21.0	42.0	62.9	83.9	102.8	110.4	112.3
	(3.1)	(6.3)	(9.4)	(12.6)	(15.7)	(18.9)	(37.5)	(56.0)	(74.9)	(90.9)	(99.4)	(98.8)
Average I/ σ [c]	17.3	23.8	25.5	28.6	31.6	34.7	42.4	50.7	52.2	52.9	50.6	47.4
	(3.9)	(5.4)	(6.2)	(6.7)	(7.3)	(8.0)	(11.2)	(12.8)	(14.0)	(14.6)	(14.9)	(14.2)
R_{meas} [d] (%)	5.6	5.8	6.3	6.8	6.9	7.3	7.6	8.2	8.7	9.5	10.3	11.2
	(33.5)	(36.2)	(40.0)	(44.4)	(44.4)	(46.5)	(46.0)	(50.7)	(53.0)	(57.3)	(62.0)	(66.8)
$R_{mrgd\text{-}F}$ [d] (%)	8.5	6.6	5.7	5.1	4.8	4.5	3.1	2.8	2.5	2.4	2.4	2.5
	(35.4)	(28.8)	(24.6)	(22.6)	(22.1)	(19.0)	(14.2)	(12.5)	(11.4)	(11.2)	(10.9)	(11.7)

[a] The data was processed to 1.90 Å with the numbers in the parentheses indicating values in the highest resolution shell (1.97-1.90 Å).
[b] Bijvoets were not merged.
[b] $|F|/\sigma|F| \geq 0$ cutoff was used for data processing.
[c] R_{meas} is the multiplicity weighted merging R-factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data [30].

between the group of 'correct' solutions with significantly higher correlation coefficients and the group of 'incorrect' solutions with lower correlation coefficients. For each analysis the three sulfur positions were refined (positions and occupancies) and phasing was carried out using the program MLPHARE [26]. The phase ambiguity was resolved using the solvent flattening procedure of DM [27]. For all additional data analyses, programs from the CCP4 suite [28] and CNS [29] were used.

4.3 Results

4.3.1 Datasets

In order to systematically study the effects of data quality on sub-structure solution and phasing, the 1260 oscillation images in the LAB dataset were divided into seven consecutive 180-image datasets labeled A through G, each with ca. 20-fold redundancy (Table 4.2). A further eleven datasets were created to study the effect of redundancy. As an extension of the above mentioned nomenclature, six datasets were labeled A-B (360 images), A-C (540 images) and so on to A-G (1260) images (Table 4.3). The five additional datasets, created to test the lower limit of data quality that would allow a structure solution, are designated as "30", "60", "90", "120" and "150" representing the number of images in each set (Table 4.3). The SYNC dataset (Table 4.1) is not a major focus of this work and was simply treated as a single set.

4.3.2 Crystal decay

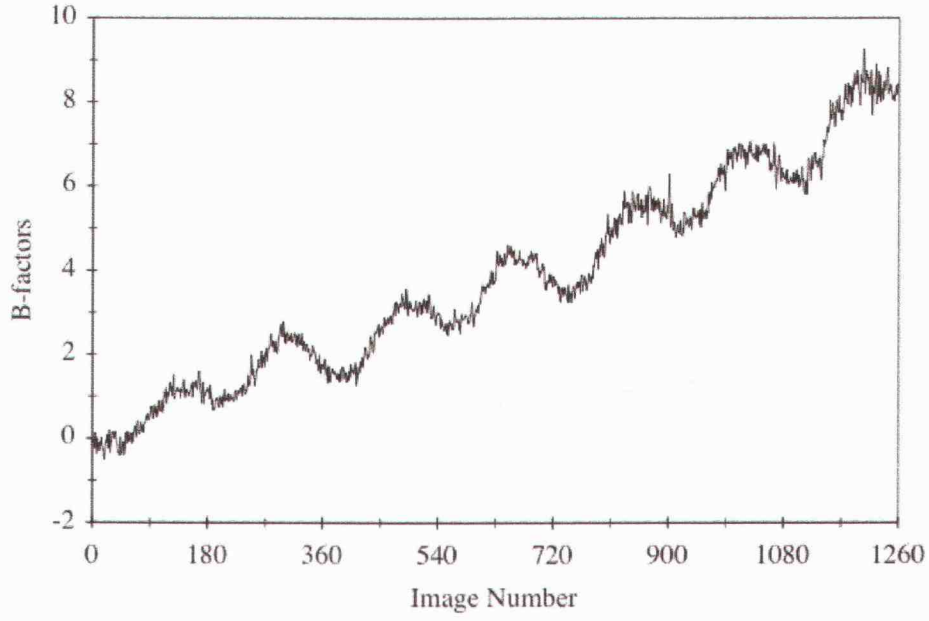
Despite the crystals being frozen, decay was evident during collection of the LAB data as a visible loss in resolution between the first and last image. Quantitatively, the gradual crystal decay can also be seen in the increasing scaling B-factors as a function of image number (Figure 4.1a), and the gradual decrease of the average signal-to-noise ratio (I/σ) of the reduced data for datasets A through G (Figure 4.1b). As expected, the R_{meas} and $R_{mrgd\text{-}F}$ values (both overall and in the highest resolution bin) also show an increase from dataset A to dataset G (see Table 4.2).

For an estimate on how crystal decay impacted data accuracy, *R*-factors were calculated between datasets B through G and dataset A as a function of resolution (Figure 4.1c).

Figure 4.1. **Quantifying crystal decay during data collection.** (a) Scaling B-factors increase gradually as a function of image number. This gradual increase is indicative of crystal decay with increased exposure. (b) Average I/ σ of the reduced data is plotted as a function of resolution for datasets A (blue), B (violet), C (orange), D (cyan), E (purple), F (red), and G (green). (c) *R*-factors between each dataset and dataset A plotted against resolution reflect how decay impacts reduced data accuracy. *R*-factors between datasets B and A (violet), C and A (orange), D and A (cyan), E and A (purple), F and A (red), G and A (green) are shown.

Figure 4.1. (continued)

(a)



(b)

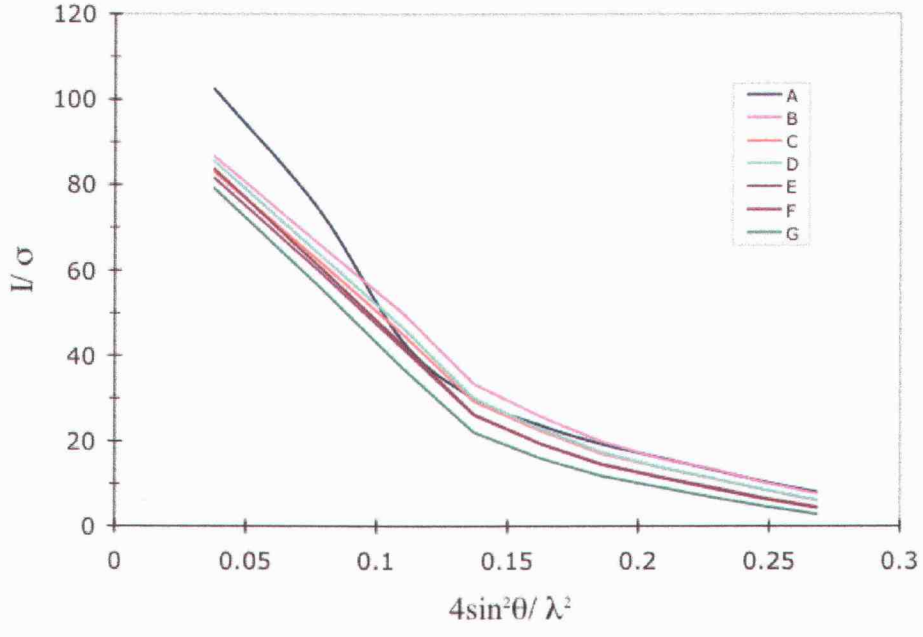
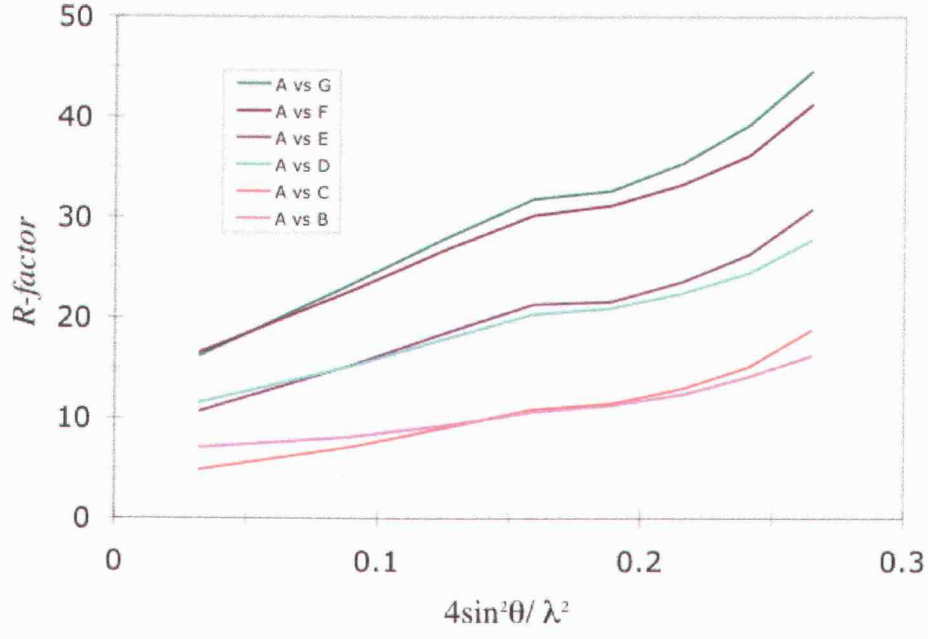


Figure 4.1. (continued)

(c)



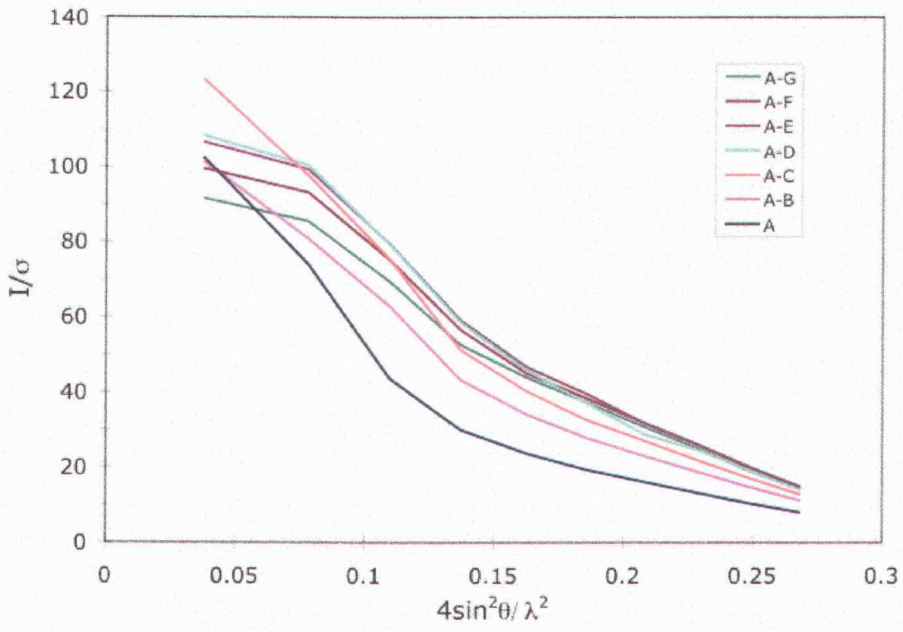
Dataset A was used as a reference since it was collected when the crystal was the freshest. The *R*-factors were the lowest between dataset B and A (overall/ highest resolution bin = 7.0%/ 16.2%) and were the highest between G and A (overall/ highest resolution bin = 16.1%/ 44.6%). Interestingly, the curves were paired such that the B *versus* A *R*-values are very similar to those of C *versus* A and so on (Figure 4.1c). A plausible explanation is that datasets A, C, E and G were collected over the same oscillation ranges (counting the 0.5° offset) and so have minimal systematic differences related to crystal geometry (e.g. absorption) or detector imperfections (e.g. non- homogeneity). Thus, the *R*-values calculated using these datasets are mostly influenced by crystal decay. On the other hand, the comparison of dataset A with datasets B, D and F will reflect decay plus systematic errors and this will shift the *R*-factors to higher values for those datasets. Because systematic errors may be less resolution-dependent, this would explain why the *R*-factors for datasets C, E and G are higher at low resolution.

The behavior of the datasets of increasing redundancy (A through A-G) also shows the effects of decay (Figure 4.2). When the indicators of reduced data quality – I/ σ and R_{mrgd} – are plotted as a function of resolution, the I/ σ increases from dataset A to dataset A-D after which the values get worse despite the increase in redundancy (Figure 4.2a). Similarly, the $R_{mrgd\text{-}F}$ (for which lower values indicate better data [30]) decreases from dataset A to A-D after which no further improvement is seen (Figure 4.2b). According to these statistical measures, increased redundancy results in better quality data up through A-D, after which the overall data quality is negatively impacted by merging of subsequent data.

4.3.3 Assessment of anomalous signal

For SAD phasing, the quality of the anomalous signal is most important so it is of interest to see how this is impacted by the trade-off between increasing redundancy and crystal decay. Even without phases, for any single dataset it is possible to assess the strength of the anomalous signal by splitting the dataset in half and calculating the anomalous correlation coefficients between the two halves. The individual datasets A through G show a trend of decreasing correlation indicating that decay does impact the quality of the anomalous signal (Figure 4.3a). The correlation remains fairly high at

(a)



(b)

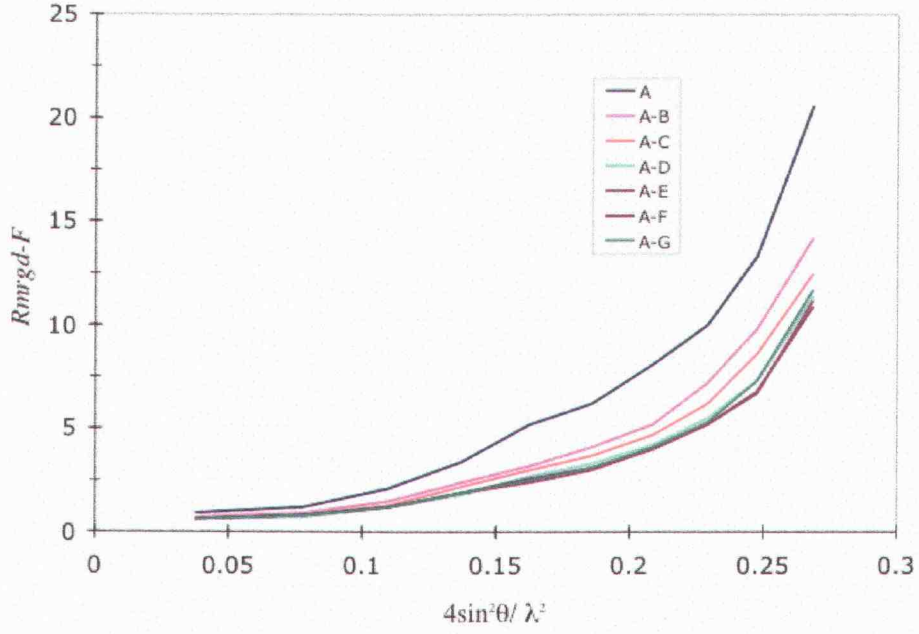
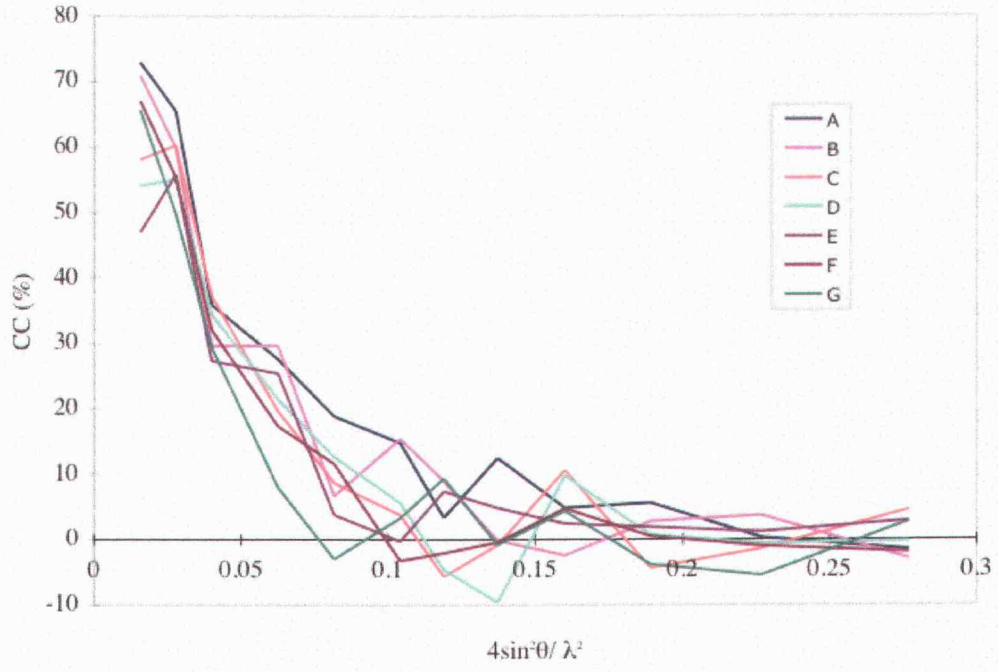


Figure 4.2. **Variation in reduced data quality indicators as a function of redundancy.** I/ σ (a) and $R_{mrgd\text{-}F}$ (b), are plotted as a function of resolution for the datasets A (blue), A-B (violet), A-C (orange), A-D (cyan), A-E (purple), A-F (red), and A-G (green). $R_{mrgd\text{-}F}$ and I/ σ [30] reflect the quality of the final reduced data.

(a)



(b)

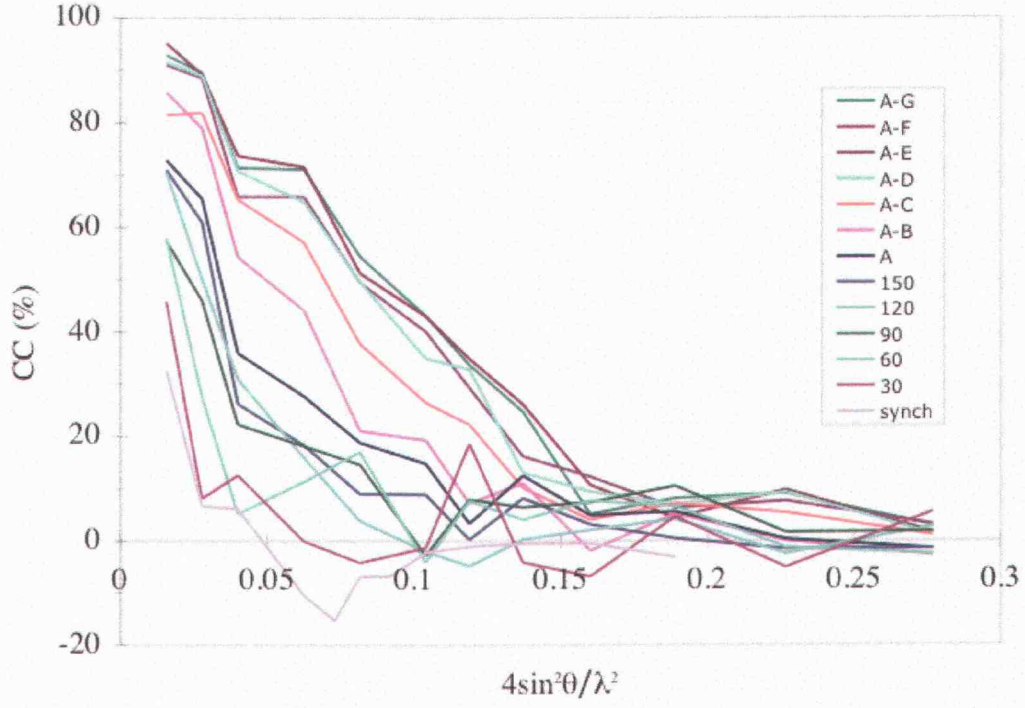


Figure 4.3. **Phase independent assessment of the anomalous signal.** Correlation coefficients (%) between the anomalous differences for (a) datasets A through G and (b) datasets "30" through A-G are plotted against resolution. The curves are colored as in Figures 4.1b and 4.2.

low resolution, but the resolution at which it drops below 20% decreases from near 3.5 Å for dataset A to near 4.5 Å for G. The trend in the anomalous correlations for datasets with increasing redundancies is dramatic. The anomalous signal steadily increases from dataset "30" to A-D. Despite further increases in redundancy, no further increase is seen for datasets between A-D and A-G (Figure 4.3b). The anomalous signal for the synchrotron is even weaker than that of dataset "30" with the correlation coefficients falling below 10% at around 6.0 Å resolution suggesting little anomalous information is present in the synchrotron data.

Given known phases (from the model) a more accurate assessment of the anomalous signal present in each dataset is provided by peak heights in anomalous difference Fourier maps (Table 4.4). While this could be done as a function of resolution, we have done analysis at 3.0 Å to observe the trends. In all cases the top three peaks correspond to the three sulfur sites used for the original structure solution[22]. For all datasets except dataset G, the fourth largest peak in the map corresponds to a fourth sulfur site (a second sulfate ion) in the structure. In the best cases, this peak is over twice as large as the highest noise peaks. We missed this fourth sulfur in the original structure solution [22], as we only searched for three sites.

Consistent with other indicators of data quality, the sulfur peak heights show a steady decrease with increased crystal decay (datasets A to G). The only notable exception is peak 3 calculated using dataset C (Table 4.4); we do not have any explanation for this. Similarly the trend of peak heights as a function of data redundancy (datasets "30" through A-G) match the other data quality statistics as the peak heights increase steadily from dataset "30" to A-D or A-E and then decrease dramatically to dataset A-G. By this measure, dataset A-G has even lower anomalous signal than dataset A alone! A comparison of these results with the anomalous correlation coefficients suggests that the correlation coefficient measure reflects accurately the increase in anomalous signal, but is not as sensitive to the loss in anomalous signal due to the addition of decayed data.

Table 4.4 also reports the top six peaks found using the synchrotron data. There is no clear separation between the noise and the signal peaks, consistent with the inability of the synchrotron data to guide solution of sulfur sub-structure.

Table 4.4. **Anomalous peak heights calculated at 3.0 Å resolution** [a]

	Peak1 (Met174)	Peak2 (Sulfate501)	Peak3 (Cys160)	Peak4 (Sulfate502)	Peak5 (noise)	Peak6 (noise)
A	14.4	12.5	11.5	5.8	3.6	3.5
B	14.3	11.4 [b]	10.1 [b]	6.6	4.0	3.4
C	14.9	10.6	6.8	6.5	4.1	3.9
D	14.0	10.1	9.2	5.9	3.7	3.5
E	12.9	8.5	8.4	5.0	3.9	3.7
F	10.9	9.0	7.4	5.1	4.0	3.8
G	11.5	8.0	7.2	3.8	3.6	3.5
"30"	7.6	7.0	5.2	4.6	4.1	3.8
"60"	10.3	8.3	7.8	6.0	4.2	3.9
"90"	12.4	10.5	9.9	5.5	3.6	3.5
"120"	13.4	11.2	10.8	6.0	3.7	3.4
"150"	14.2	12.3	11.9	5.9	3.8	3.7
A-B	17.2	13.7 [b]	13.2 [b]	7.4	3.7	3.3
A-C	18.6	14.1	12.9	7.7	3.6	3.4
A-D	19.4	14.2	13.0	7.6	3.9	3.7
A-E	19.3	13.8	12.9	7.8	3.7	3.5
A-F	17.1	12.2	10.8	7.3	4.2	3.7
A-G	14.3	9.0	8.6	5.5	4.7	4.1
Synch [c]	(5.0)	(4.2)	(4.1)	(4.0)	(4.0)	(3.9)

[a] 3.0 Å was chosen because it was the resolution at which the original structure solution was carried out [22].
[b] The Sulfate501 and the Cys160 peaks were swapped.
[c] The values in parentheses represent the top peaks in the maps and not the anomalous peak heights of the sulfur atoms.

4.3.4 Sub-structure determination

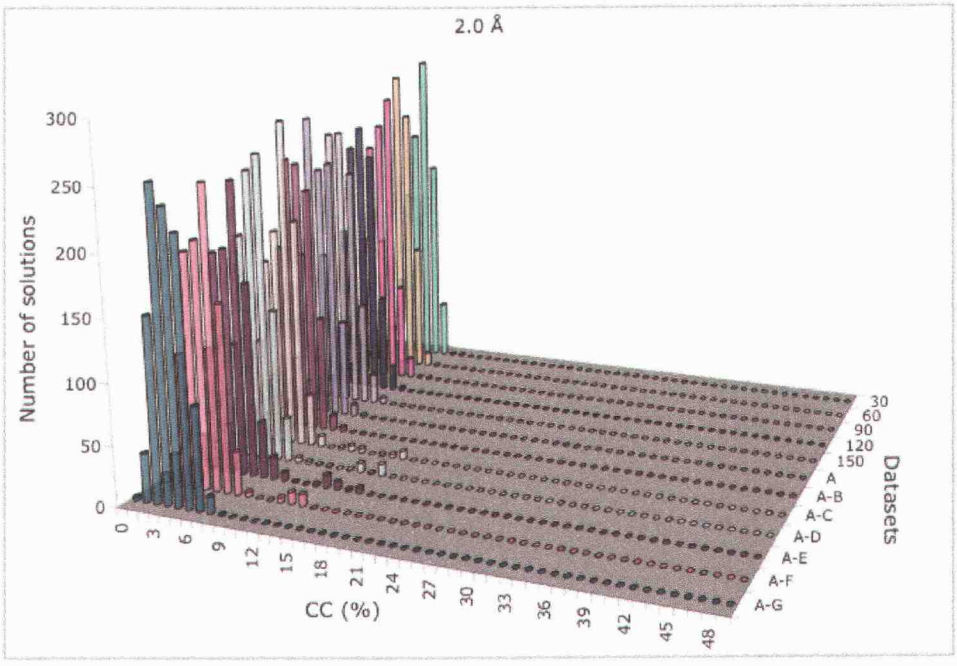
A systematic study of the effects of redundancy, crystal decay, and resolution on sub-structure determination was carried out for datasets "30" through A-G at resolution cutoffs 2.0, 2.5, 3.0 and 3.5 Å. A complete picture of the results of each 1000-trial run is given by histograms of correlation coefficients (Figure 4.4). The results are also summarized in terms of the correlation coefficients and number of correct sulfur positions found (out of three) for the best trial and the number of trials significantly above the noise level (Table 4.5). At all four resolutions, the indicators show that the ability to find the correct sites is maximal for datasets A-C or A-D (or A-E at 2.0 Å resolution) and then steadily decreases both toward lower or higher redundancy. Interestingly, as was seen for the anomalous difference map peak heights, dataset A-G seems roughly similar in phasing power to dataset A. The best datasets (A-C and A-D) led to correct solutions at all resolutions, but for any given dataset the success rate strongly increased with decreasing resolution, for example from 3% at 2.0 Å to 53% at 3.5 Å resolution for dataset A-D (Figure 4.4; Table 4.5). Also, the searches carried out at 2.0 Å resolution only yielded a distinct population of correct solutions for datasets A-C through A-F, whereas those carried out using the 3.5 Å resolution cutoff yielded distinctly correct solutions for a broader set of datasets "90" through A-G (Table 4.5). Finally, the separation of the correlation coefficient of the 'correct' trials from the 'incorrect' trials is increasingly pronounced at the lower resolutions (Figure 4.4).

To separately assess the effects of crystal decay on sub-structure solution, with no influence of redundancy, sub-structure solution attempts were made using datasets A through G, each of which is based on 180° of data. We only tested the behavior using 3.0 and 3.5 Å resolution cutoffs, since we expected solution clarity to decrease with decay, and only these resolutions gave sufficiently strong success rates for dataset A that there would be room for monitoring a decrease in quality. The strong trend at both 3.0 and 3.5 Å resolution is a steady decrease in the number of correct trials with increased decay to the point that for dataset G even though a correct solution was found, its correlation coefficient was not very distinct from those of the non-solutions (Figure 4.5 and Table 4.6).

Figure 4.4. **Sub-structure solution results for datasets with increasing redundancies.** Histograms are shown of the correlation coefficients for the 1000 sub-structure solution trials carried for each of datasets "30" through A-G at four different resolutions – 2.0 Å (a), 2.5 Å (b), 3.0 Å (c), and 3.5 Å (d).

Figure 4.4 (continued)

(a)



(b)

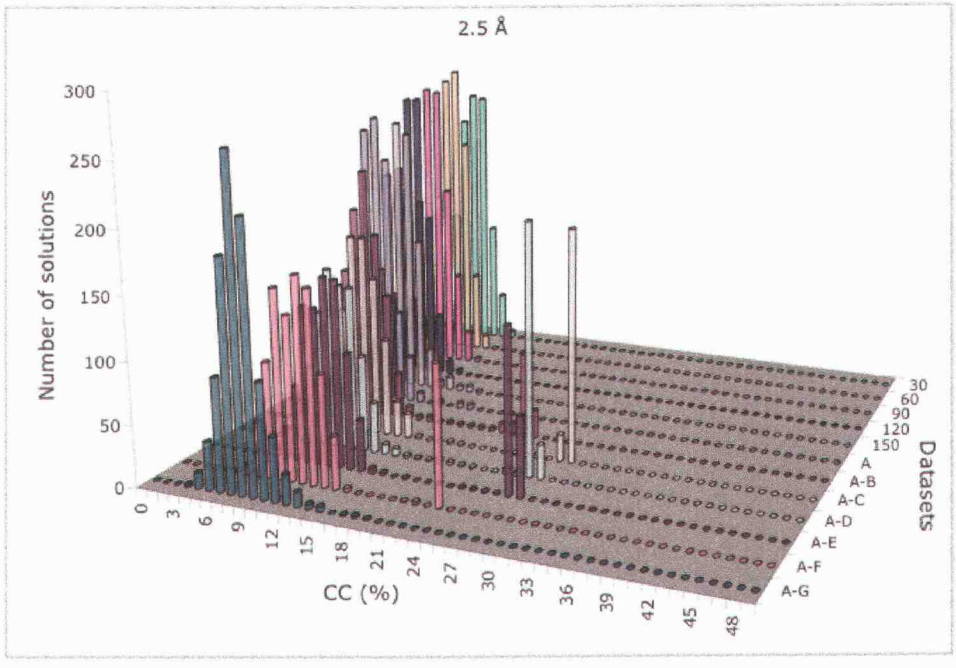
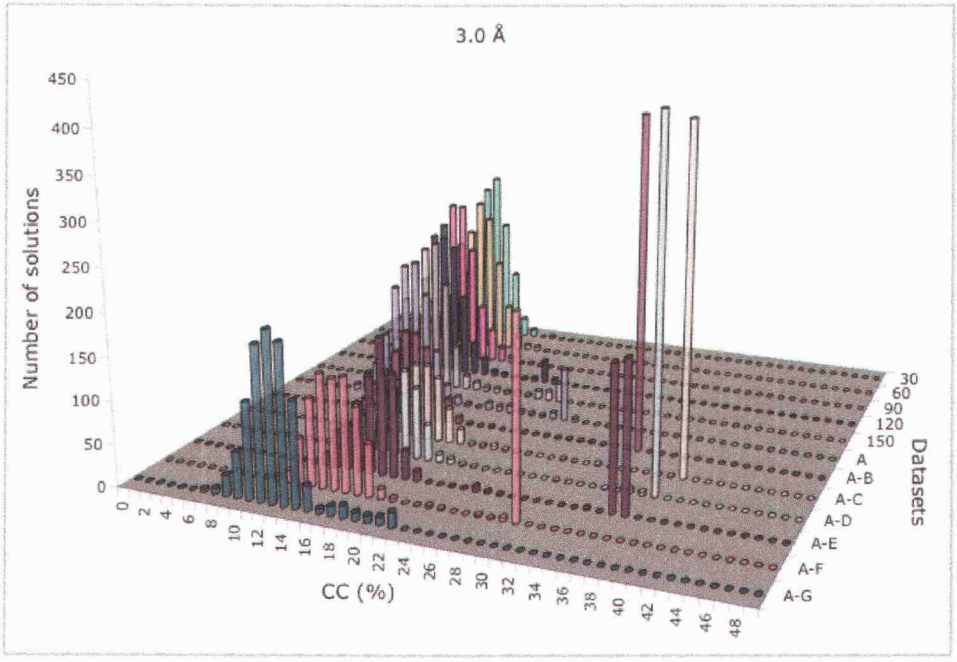


Figure 4.4 (continued)

(c)



(d)

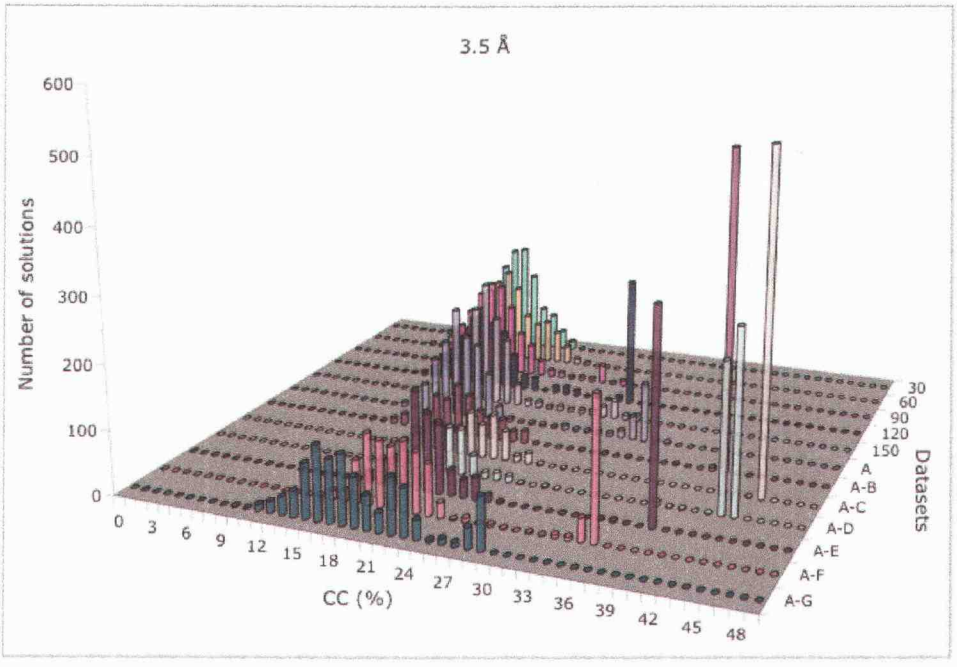
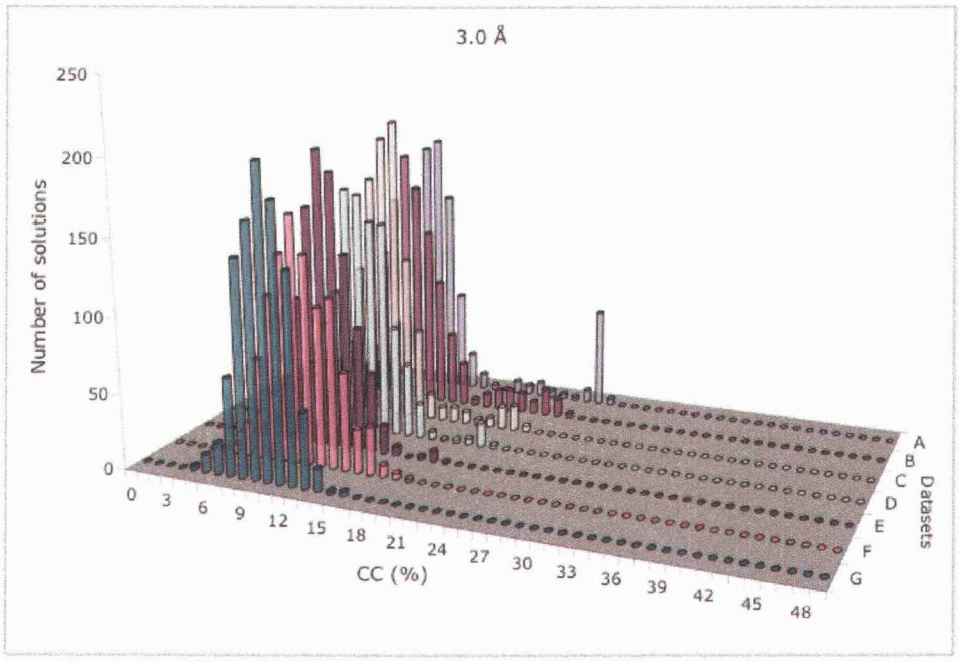


Table 4.5 **Sub-structure solution for increasing redundancy datasets** [a]

	"30"	"60"	"90"	"120"	"150"	A	A-B	A-C	A-D	A-E	A-F	A-G
2. 0 Å	0 (9.1) 0	0 (8.3) 0	0 (9.2) 0	0 (8.8) 1	0 (8.4) 1	0 (8.5) 2	0 (8.7) 3	12 (16.4) 3	20 (16.9) 3	27 (16.9) 3	24 (14.0) 3	0 (8.4) 2
2.5 Å	0 (11.8) 0	0 (11.1) 0	0 (12.0) 3	0 (11.9) 3	22 (15.3) 3	9 (20.1) 3	121 (24.4) 3	214 (29.5) 3	236 (28.7) 3	203 (28.7) 3	115 (23.9) 3	4 (19.1) 3
3.0 Å	0 (17.4) 0	0 (18.4) 0	3 (19.6) 3	44 (23.5) 3	48 (24.5) 3	80 (27.7) 3	411 (34.7) 3	408 (39.7) 3	435 (38.6) 3	348 (37.6) 3	239 (30.6) 3	74 (22.5) 3
3.5 Å	0 (20.4) 0	0 (22.7) 2	52 (27.3) 3	218 (30.6) 3	79 (31.8) 3	165 (33.4) 3	502 (41.4) 3	523 (45.4) 3	525 (44.9) 3	332 (42.8) 3	267 (36.7) 3	120 (29.9) 3

[a] The top line indicates the approximate number of 'correct' solutions out of a 1000 search trials with the correlation coefficient of all reflections of the best solution (as output by SHELXD) in parentheses. The second line of each entry is the number of correct sulfur positions (out of the three) located by SHELXD (see Materials and Methods).

(a)



(b)

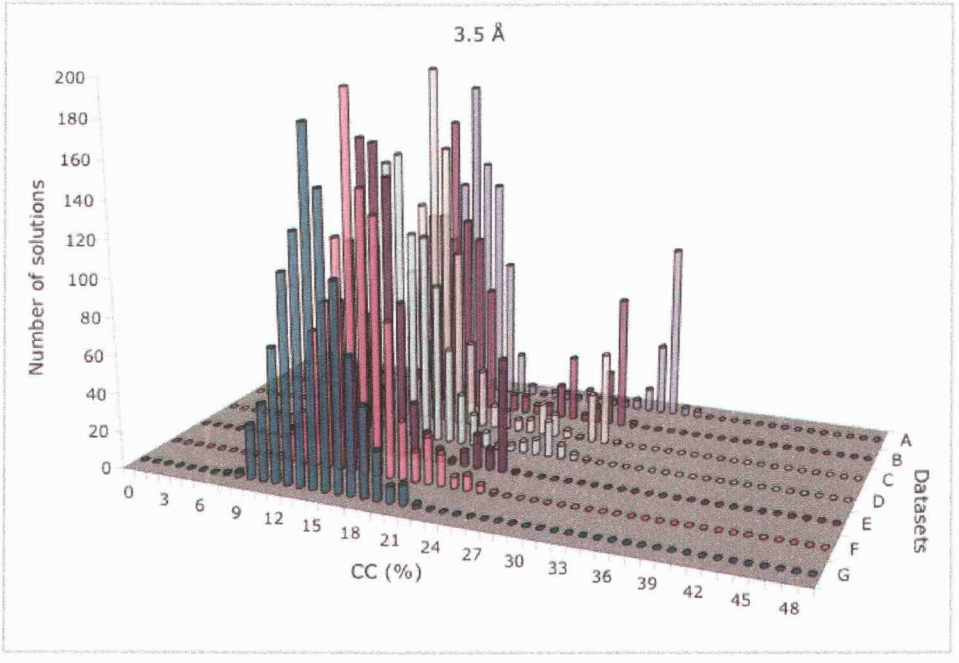


Figure 4.5. **Sub-structure solution for 180° datasets.** Histograms for the correlation coefficients for the 1000 sub-structure solution trials carried out for each of the datasets A through G at 3.0 (a) and 3.5 Å (b) resolutions are shown.

Table 4.6. **Sub-structure solution for the 180° datasets** [a]

	A	B	C	D	E	F	G
3.0 Å	80 (27.7/ 16.7)	41 (25.2/ 15.1)	42 (23.6/ 10.9)	25 (22.2/ 12.5)	10 (20.1/ 9.6)	0 (19.5/ 11.7)	0 (16.5/ 9.4)
3.5 Å	165 (33.4/ 20.8)	110 (31.6/ 15.2)	77 (31.0/ 14.0)	49 (30.2/ 19.7)	100 (26.0/ 12.4)	19 (26.3/ 15.2)	13 (22.8/ 10.7)

[a] Three (of three) correct sulfur positions are located using the datasets at both resolutions. The top line indicates the approximate number of 'correct' solutions out of a 1000 search trials with the correlation coefficient of all and weak reflections of the best solution (as output by SHELXD) in parentheses.

Faced with the result that lower resolution cutoffs yielded better success, we used dataset A-D, the best performing dataset at all resolutions thus far tested, to attempt sub-structure solution at five additional resolution cutoffs – 4.0, 4.5, 5.0, 5.5 and 6.0 Å. Surprisingly, all the way up to 6.0 Å, clear solutions were obtained at all resolutions with ever increasing correlation coefficients (Figure 4.6). The maximal success rate was 63% using the 4.0 Å resolution cutoff.

For the synchrotron data, attempts to solve the sulfur sub-structure were unsuccessful at all the resolutions tried (2.0, 2.5, 3.0 and 3.5 Å). These results are consistent with the very weak anomalous signal present in the synchrotron data (Figure 4.3b).

4.3.5 **Phase determination**

Phases were calculated for each dataset and resolution combination using the sulfur sub-structure determined for that combination. Phase quality was evaluated by calculating the differences between the experimental phases – both before and after density modification – and the final refined 1.65 Å calculated phases (see Materials and Methods). Figure 4.7a shows the set of phase differences calculated for dataset A-D using three sulfur positions at various resolutions. At all resolutions between 2.0 and 4.5 Å, phasing succeeded with comparable accuracy. While phase determination below 4.5 Å worked well, the amount of data is not sufficient for successful density modification. These results are representative, as all datasets at all resolutions led to similar quality phase sets that would easily be accurate enough to guide modeling. This even extends to datasets that did not allow sub-structure solution; for instance, when given the correct sulfur sites even dataset "30" yielded phases at 3.0 Å resolution equivalent to those seen in Figure 4.7a ($\Delta\phi_{pre\text{-}DM}$; $\Delta\phi_{post\text{-}DM}$ = 66.3°; 51.2°). In order to find out the phasing power of the "unsuccessful" sub-structure results, phases were calculated using the incomplete sulfur models found using certain datasets at 2.0 Å resolution (Figure 4.7b). The phases calculated using 2 and 3 sulfur sites are of highly similar quality and the automated model building program ARP/wARP was able to build ca. 90% of the model. Phases calculated with only one site were much les accurate, and experimental maps calculated from them might be

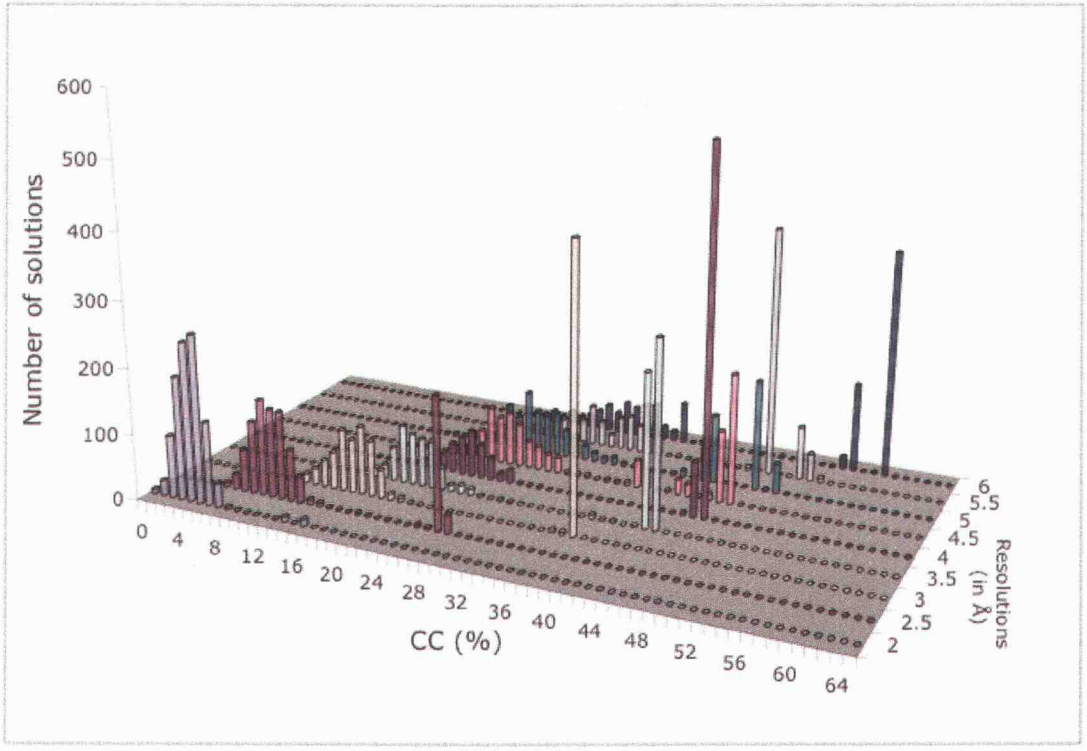
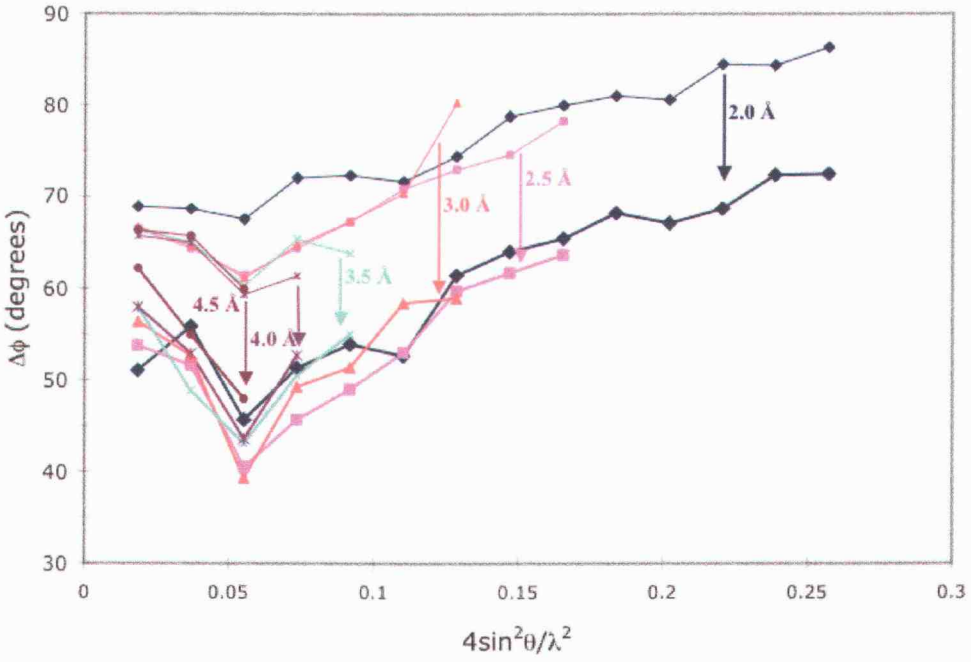


Figure 4.6. **Dependence of sub-structure solution success on resolution.** Histograms for the correlation coefficients for the 1000 sub-structure solutions are shown for dataset A-D using nine resolution cutoffs – 2.0 Å to 6.0 Å at 0.5 Å intervals.

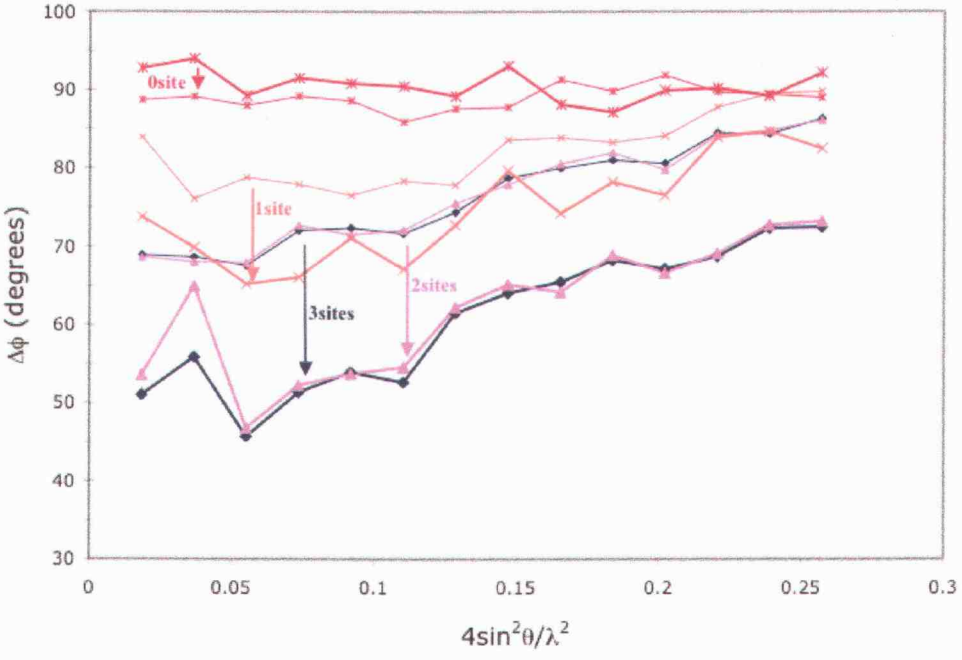
Figure 4.7. **Quality of experimental phases.** Phase differences between experimental phases – before (thin lines) and after (thick lines) density modification – and final refined phases are plotted as a function of resolution. (a) Experimental phases were calculated using three sulfur positions and dataset A-D at six different resolutions – 2.0 Å (blue), 2.5 Å (violet), 3.0 Å (orange), 3.5 Å (cyan), 4.0 Å (purple) and 4.5 Å (red). These differences are representative of the quality of phases calculated using different resolution cutoffs. (b) Experimental phases calculated at 2.0 Å resolution using dataset A-D (3sites, blue), dataset A (2 sites, violet), dataset "150" (1 site, orange), dataset "90" (0 sites, red) are shown. They are representative of phase errors from any dataset using 3, 2, 1 or 0 sulfur site for phasing.

Figure 4.7 (continued)

(a)



(b)



interpretable with difficulty but ARP/wARP failed to build a model. The phases calculated using zero correct sites, as expected, are random.

4.4 Discussion

This study was undertaken to understand the effects of redundancy, crystal decay, and resolution limits on the ability of a given dataset to lead to sub-structure solution and subsequent phasing by S-SAD. We do not know why the anomalous signal in the synchrotron data is so weak (even weaker than the 30-image LAB dataset) and we expected that better data could be collected at a synchrotron. Nevertheless, this case indicates that data collected using a Cu-Kα lab source can be better than synchrotron data for S-SAD.

Our analyses of the ToxA data reinforce a number of well-known principles – that sub-structure solution is more difficult than phasing, that data quality is of great importance, and that increased multiplicity can improve data quality immensely. Furthermore, our analyses allow us to draw two additional conclusions of practical importance. First, that for data collected from a single crystal, as multiplicity increases, the data improve and then get worse so that there is an optimal amount of data to include for successful sub-structure solution. Second, that sub-structure solution success is a sensitive function of resolution and that there is an optimal resolution cutoff for successful sub-structure solution that will be structure and data quality dependent. Finally, the results from this study are used to present a strategy for structure solution using in-house sulfur SAD data.

4.4.1 Higher multiplicity does not necessarily mean better data

It has been pointed out that crystal decay can limit data quality, but in published S-SAD structure determinations, reduced-data quality and phasing power always increases monotonically with increasing redundancy [18]. Here, this is not the case.

The behavior of scaling B-factors (Figure 4.1a), scaling R factors (Table 4.2), *R*-factors calculated between datasets B to G *versus* A (Figure 4.1c), the anomalous signal for datasets A through G (Figure 4.2a and Table 4.4), and sub-structure solution

success using these datasets (Figure 4.5 and Table 4.6), all show that crystal decay is a process that continuously decreases data quality with increased exposure.

Despite the fact that decay begins with the first image, the properties of the datasets with increasing redundancies ("30" through "A-G") indicates that for a certain period (in this case through dataset A-D) adding more data improves overall data quality, as seen in the I/ σ (Figure 4.2a), $R_{mrgd\text{-}F}$ (Figure 4.2b), the anomalous signal (Figure 4.3b and Table 4.4) and sub-structure solution success (Figure 4.4 and Table 4.5). Then, however, there is a change with the merging of datasets E, F and G. While the data quality statistical indicators continue to increase or stay the same, the difference maps (Table 4.4) or sub-structure solution success rate (Figure 4.4) plateaus and then drops dramatically.

This disparate behavior between the data quality statistical measures and the actual data quality of the anomalous signal suggest that the standard statistics are not sensitive indicators of the anomalous signal quality. Among the statistics that can be calculated from the data alone, it seems the anomalous correlation coefficient is the most informative: it shows dramatic increases from dataset "30" through A-C, matching well the true level of anomalous signal. Even this statistic though does not accurately capture the deleterious nature of including datasets E, F and G. These results emphasize the important point that more data are not always better but that this may not be obvious from the statistics.

4.4.2 Sub-structure solution success is highly dependent on resolution cutoff chosen

Our analyses have shown that for datasets with increasing redundancies, the anomalous correlation coefficients are consistently higher in lower resolution bins compared to the higher resolution bins (Figure 4.3). The greater anomalous signal in the lower resolution bins is directly reflected in the ability to locate the sulfur atoms using lower resolution cutoffs. For datasets "90" to A-G, the correlation coefficients, and the number of correct solutions increase with decreasing resolution for each dataset (Figure 4.4 and Table 4.5). Further, the distinction in the correlation coefficient values between the correct and incorrect solutions is much more pronounced using the 3.5 Å cutoff compared to using the 2.0 Å cutoff (Figure 4.4). Our data analyses,

carried out after the 1260 images were collected, shows that anomalous data from only 60 images were sufficient to locate sulfur positions when the search is done using the low resolution of 3.5 Å (Table 4.5).

In fact, when the data are truncated to unprecedented lower resolutions (up to 6.0 Å), the sulfur positions are still located well, and even with a higher correlation coefficient (Figure 4.6). Not surprisingly, the correlation coefficients of the solutions mirror the anomalous correlation coefficients by becoming even higher as resolution gets lower (Figures 4.3b and 4.6). However, the success rate does not simply increase with decreasing resolution, but is maximal using the 4.0 Å cutoff and decreases with increasing or decreasing resolution within dataset A-D (Figure 4.6). This clearly shows that while lower resolution cutoffs are better for sub-structure solution, there exists an optimal resolution cutoff at which the likelihood of structure solution is maximized.

These results contradict a widespread expectation, recently stated by Nagen *et al.* [31] that for data with low theoretical anomalous signal successful sub-structure solution using S-SAD "requires high or even atomic resolution data". This expectation may be a carry over from the legendary 1.5 Å resolution S-SAD phasing of crambin [1], and that one generally needs very high resolution data for direct methods to work [7,17]. The power of low resolution data for S-SAD is actually not a new observation, but was pointed out first by Liu *et al.* [19] who solved their sub-structure at 4.5 Å resolution and pointed out that for solving a sub-structure of isolated S-atom scatterers, 4.5 Å does correspond to atomic resolution. Also, Schneider *et al.* [15,25,32] promoted the concept of only using resolution ranges having anomalous correlation coefficients greater that 30%, because including higher resolution data that has poor anomalous signal will only add noise and thus hinder sub-structure solution. Finally, Dauter *et al.*[32] were surprised to note from their study that, "the solution of the partial structure of anomalous scatterers can be obtained at high and as well as low resolution, although at low resolution the chance of success is higher". Our work emphasizes the point that not only might lower resolution sub-structure solution attempts succeed, it could generally be better. The only real downside of lower resolution sub-structure solution

is the possible inability to resolve the two sulfurs of a disulfide at resolutions lower than about 2.5 Å.

The result of better sub-structure solutions using low resolution cutoffs raises the intriguing question of whether previous studies that reported the failure or limited success of S-SAD [14,33] would have had better success if lower resolution cutoffs were used. In the work of Lemke *et al.* [14] on *E. coli* argininosuccinate synthetase (EAS; a 51 kDa protein with 19 S-atoms and 0.86% theoretical anomalous signal) sub-structure solution barely succeeded (a success rate of 0.17%) using a 2.0 Å resolution cutoff. Based on extrapolation from our analyses, it is quite possible the success rate would have been much higher with a lower resolution cutoff. In the work of Uson *et al.* [33] on R11, sub-structure solution attempts failed using a resolution cutoff of 2.5 Å, and no other cutoffs were tried. However, the 30% anomalous correlation coefficient cutoff cutoff for that dataset was at 3.3 Å resolution, making it quite plausible that the data were sufficient to drive structure solution if a lower resolution cutoff had been used.

The results shown here underline the importance of the idea that higher resolution is not always better for sub-structure solution using S-SAD data.

4.4.3 Strategy for in-house sulfur SAD structure solution

The results of our analyses allow us to propose a concrete strategy for data collection, assessment of the anomalous signal and sub-structure solution for S-SAD using the in-house Cu-Kα X-ray source. The approach would be to start collecting data on a single crystal, monitoring data quality using statistics that are robust indicators of reduced data quality (I/ σ and $R_{mrgd\text{-}F}$). Further, the anomalous signal should be monitored using the anomalous correlation coefficients. Data collection should be continued as long as these parameters improve with increasing redundancy. When these parameters reach a plateau, data collection on a fresh crystal should be started in order to avoid decreasing the data quality. This strategy is an important adjustment from the strategy proposed by Debreczeni *et al.* [18], where it was suggested "simply to continue to collect data, increasing the redundancy, until the structure can be solved".

Since the anomalous correlation coefficient is one of the most useful indicators for determining the amount of anomalous signal (Figure 4.3), it should used to

monitor the increase or decrease in the anomalous signal, which in turn can be used to guide the crucial decision of resolution cutoffs for sub-structure solution. For the original ToxA sub-structure solution, the data were truncated at a resolution (3.0 Å) where the anomalous correlation was ca. 20%. This is slightly lower than the range of 25-30% recommended by Schneider *et al.* [25]. While this range can be used as a guide for determining resolution cutoffs for initial sub-structure solution attempts, it should not be used as a rule. Lower resolutions cutoffs, all the way up to 6.0 Å, should be tried for locating sulfur positions; then, since sub-structure solution is more difficult than phasing, data truncated to low resolutions for locating the sulfur atoms can be extended to higher resolutions for phasing and density modification.

Given the slow nature of in-house data collection, it makes sense to attempt structure solution even while data collection is in progress. In hindsight, the ToxA structure could be solved using only data from 60 images, but we did not try until much later. We suspect merging data from fresh isomorphous crystals should allow data quality to be improved even for crystals that decay rapidly [30].

The analyses presented here underscores suggestions made by others that LAB S-SAD is a powerful technique [2,19,32,34] and provides evidence that the mindset of requiring high resolution data has led to an underappreciation of just how powerful it is. Taking into account the natural presence of sulfur atoms in proteins [7,31], it should find wide applicability in high-throughput structure determination.

4.5 Acknowledgements

We would like to thank Dr. Rick Faber for valuable help in data collection and useful discussions. The help extended by the staff at the ALS (beamline 8.2.1) is also greatly appreciated. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory. This work was supported by a National Science Foundation grant (MCB-0488665) to Drs. Lynda M. Ciuffetti and P.A.K and in part by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number 2001-35319-10017 to L.M.C. This work was made possible in part by

the Proteins and Nucleic Acids Core facility of the Environmental Health Sciences Center at Oregon State University (NIEHS grant P30 ES00210).

4.6 **References**

1. Hendrickson, W. A. & Teeter, M. M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* 290, 107-113.

2. Dauter, Z. (2002). New approaches to high-throughput phasing. *Curr. Opin. Struct. Biol.* 12, 674-8.

3. Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). The C1 subunit of alpha-crustacyanin: the de novo phasing of the crystal structure of a 40 kDa homodimeric protein using the anomalous scattering from S atoms combined with direct methods. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1230-7.

4. Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). Structure of a functional IGF2R fragment determined from the anomalous scattering of sulfur. *EMBO J.* 21, 1054-62.

5. Li, S., Finley, J., Liu, Z. J., Qiu, S. H., Chen, H., Luan, C. H., Carson, M., Tsao, J., Johnson, D., Lin, G., Zhao, J., Thomas, W., Nagy, L. A., Sha, B., DeLucas, L. J., Wang, B. C. & Luo, M. (2002). Crystal structure of the cytoskeleton-associated protein glycine-rich (CAP-Gly) domain. *J. Biol. Chem.* 277, 48596-601.

6. Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). De novo phasing of two crystal forms of tryparedoxin II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 58, 21-8.

7. Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). Phasing on anomalous signal of sulfurs: what is the limit? *Acta Crystallogr. D Biol. Crystallogr.* 59, 1020-7.

8. Lartigue, A., Gruez, A., Briand, L., Blon, F., Bezirard, V., Walsh, M., Pernollet, J. C., Tegoni, M. & Cambillau, C. (2004). Sulfur single-wavelength anomalous diffraction crystal structure of a pheromone-binding protein from the honeybee Apis mellifera L. *J. Biol. Chem.* 279, 4459-64.

9. Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). Determination of a novel structure by a combination of long-wavelength sulfur phasing and radiation-damage-induced phasing. *Acta Crystallogr. D Biol. Crystallogr.* 60, 686-95.

10. Weiss, M. S., Sicker, T., Djinovic-Carugo, K. & Hilgenfeld, R. (2001). On the routine use of soft X-rays in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 57, 689-95.

11. Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). Soft X-rays, high redundancy, and proper scaling: a new procedure for automated protein structure determination via SAS. *Structure (Camb)* 9, 771-7.

12. Kwiatkowski, W., Noel, J. P. & Choe, S. (2000). Use of Cr Ka radiation to enhance the signal from anomalous scatterers including sulfur. *J. Appl. Cryst.* 33, 876-881.

13. Yang, C. & Pflugrath, J. W. (2001). Applications of anomalous scattering from S atoms for improved phasing of protein diffraction data collected at Cu Kalpha wavelength. *Acta Crystallogr. D Biol. Crystallogr.* 57, 1480-90.

14. Lemke, C. T., Smith, G. D. & Howell, P. L. (2002). S-SAD, Se-SAD and S/Se-SIRAS using Cu Kalpha radiation: why wait for synchrotron time? *Acta Crystallogr. D Biol. Crystallogr.* 58, 2096-101.

15. Debreczeni, J. E., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case. *Acta Crystallogr. D Biol. Crystallogr.* 59, 393-5.

16. Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. (2004). Solving the structure of the bubble protein using the anomalous sulfur signal from single-crystal in-house Cu Kalpha diffraction data only. *Acta Crystallogr. D Biol. Crystallogr.* 60, 250-5.

17. Debreczeni, J. E., Girmann, B., Zeeck, A., Kratzner, R. & Sheldrick, G. M. (2003). Structure of viscotoxin A3: disulfide location from weak SAD data. *Acta Crystallogr. D Biol. Crystallogr.* 59, 2125-32.

18. Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). In-house measurement of the sulfur anomalous signal and its use for phasing. *Acta Crystallogr. D Biol. Crystallogr.* 59, 688-96.

19. Liu, Z. J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B. C. (2000). Structure of the Ca2+-regulated photoprotein obelin at 1.7 A resolution determined directly from its sulfur substructure. *Protein Sci.* 9, 2085-93.

20. Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). Away from the edge: SAD phasing from the sulfur anomalous signal measured in-house with chromium radiation. *Acta Crystallogr. D Biol. Crystallogr.* 59, 1943-57.

21. Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* 115, 90-112.

22. Sarma, G. N., Manning, V. A., Ciuffetti, L. M. & Karplus, P. A. (2005). Crystal structure of an RGD-containing host-selective toxin: *Pyrenophora tritici repentis* Ptr ToxA. *Plant Cell* in press.

23. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. J. & Sweet, R. M., eds.), Vol. 276, pp. 307-26.

24. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458-63.

25. Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1772-9.

26. Otwinowski, Z. (1991). Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds.), pp. 80-6. Warrington: Daresbury Laboratory.

27. Cowtan, K. D. & Zhang, K. Y. (1999). Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* 72, 245-70.

28. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760-3.

29. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54 (Pt 5), 905-21.

30. Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4, 269-75.

31. Nagem, R. A. P., Ambrosio, A. L. B., Rojas, A. L., Navarro, M. V. A. S., Golubev, A. M., Garratt, R. C. & Polikarpov, I. (2005). Getting the most out of X-ray home sources. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1022-1030.

32. Dauter, Z. & Nagem, R. A. P. (2002). Direct way to anomalous scatterers. *Z. Kristallogr.* 217, 694-702.

33. Uson, I., Schmidt, B., von Bulow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). Locating the anomalous scatterer substructures in halide and sulfur phasing. *Acta Crystallogr. D Biol. Crystallogr.* 59, 57-66.

34. Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol.* 289, 83-92.

Appendices

Appendix 1

Crystal Structure of a Novel *Plasmodium falciparum* 1-Cys Peroxiredoxin

Ganapathy N. Sarma, Christine Nickel, Stefan Rahlfs, Marina Fischer, Katja Becker and P. Andrew Karplus

Published in *Journal of Molecular Biology* (2005) **346**, 1021-1034

A1.1 Summary

Plasmodium falciparum, the causative agent of malaria, is sensitive to oxidative stress and therefore the family of antioxidant enzymes, peroxiredoxins (Prxs) represent a target for antimalarial drug design. We present here the 1.8 Å resolution crystal structure of *P. falciparum* antioxidant protein, *Pf*AOP, a Prx that in terms of sequence groups with mammalian PrxV. The structure is compared to all eleven known Prx structures to gain maximal insight into its properties. We describe the common Prx fold and show that the dimeric *Pf*AOP can be mechanistically categorized as a 1-Cys Prx. In the active site the peroxidatic Cys is over-oxidized to cysteine sulfonic acid, making this the first Prx structure seen in that state. Now with structures of Prxs in Cys-sulfenic, -sulfinic and -sulfonic acid oxidation states known, the structural steps involved in peroxide binding and over-oxidation are suggested. We also describe that *Pf*AOP has an α-aneurism (a one residue insertion), a feature that appears characteristic of the PrxV-like group. In terms of crystallographic methodology, we enhance the information content of the model by identifying bound water sites based on peak electron densities, and we use that information to infer that the oxidized active site has suboptimal interactions that may influence catalysis. The dimerization interface of *Pf*AOP is representative of an interface that is widespread among Prxs, and has sequence dependent variation in geometry. The interface differences and the structural features (like the α-aneurism) may be used as markers to better classify Prxs and study the evolution of Prxs.

A1.2 Introduction

Plasmodium falciparum causes the most virulent form of malaria. This tropical disease represents a growing threat to human health [1,2]. With an increasing number of antimalarials to which the parasite is now resistant, it has become imperative to develop new drugs directed against novel targets [3-5]. During its life stages in humans, the parasite is challenged by reactive oxygen (ROS) and nitrogen species (RNS) that are produced by hemoglobin digestion and the host immune system [6]. For this reason, the redox systems of *P. falciparum* are likely to be crucial for its pathogenicity, and

we have been studying them in search of drug targets. NADPH-dependent systems are the glutathione system containing glutathione reductase, glutathione (GSH), and glutaredoxin (Grx) [7] and the thioredoxin system containing thioredoxin reductase (TrxR) and thioredoxin (Trx) [8]. For detoxifying peroxides, *P. falciparum* has neither catalase nor glutathione peroxidase (Gpx), therefore peroxiredoxins (Prxs) are thought to be the major defense line against ROS and RNS [8,9].

Prxs form a group of ubiquitously distributed peroxidases that do not harbor any metal or prosthetic group [10,11]. Members of this group can reduce various peroxides including hydrogen peroxide (H_2O_2), *t*-butyl hydroperoxide (*t*-BOOH), the aromatic cumene hydroperoxide, peroxy forms of phosphatidylcholine and linoleic acid [12], and in some cases can also reduce RNS such as peroxynitrite [13,14]. In organisms containing additional peroxide reducing enzymes, the relative importance of antioxidant activities of Prxs, catalase and glutathione peroxidases (GPx) is unknown. Interestingly evidence is accumulating that in many eukaryotes Prxs act as modulators of signaling pathways that appear to use hydrogen peroxide as a secondary messenger [12,15-18]. The mammalian Prxs have been named PrxI through PrxVI, and other Prxs go by a variety of names such as thiol peroxidase (Tpx), tryparedoxin peroxidase, and AhpC. The named groups do not always match the levels of sequence similarity [15].

As has been recently reviewed by Wood *et al.* [19], all Prxs carry out their peroxidase function with a conserved Cys that is present in the first turn of an α-helix and sits at the base of the active site pocket. This Cys, known as the peroxidatic cysteine (C_P), attacks the peroxide substrate and is oxidized to cysteine sulfenic acid (C_P–SOH) in the first step of the peroxidase reaction [20,21]. The second step of the reaction is thought to require the active site unfold locally so it is accessible for attack by a free thiol (or thiolate) to release H_2O and form a disulfide [12,22]. This idea is supported by two Prx crystal structures that each reveal that the reduced protein is present both in fully folded (active site intact) and locally unfolded conformations [23,24]. The presence or absence of the Cys (C_R) that resolves the cysteine sulfenic acid is the basis of the present classification of Prxs into two groups: 2-Cys and 1-Cys Prxs [11,19]. In 2-Cys Prxs, C_R is present on the same chain (atypical 2-Cys) or on the second chain

(typical 2-Cys) of a functional dimer. The disulfide is then reduced by known disulfide oxidoreductases such as thioredoxin, AhpF or tryparedoxin [25-27]. In 1-Cys Prxs no resolving Cys is present and the active site sulfenic acid is directly recycled via redox active proteins such as Grx and Trx [23,28]. In the case of the *H. influenzae* hybrid 1-Cys Prx, the resolving Grx domain is present on the same chain as the Prx domain [23].

Crystal structures of eleven Prxs have been elucidated so far in a variety of redox and oligomeric states, and the Prxs have been seen to adopt two different kinds of dimers. The first structure of a Prx – human PrxVI [29] – was observed as a dimer with a β-sheet based interface, which we call the B-type interface (for β-sheet). Later structures of Prxs were observed to be doughnut-shaped decamers formed by a pentamer of dimers with the B-type interface. Recent studies of *H. influenzae* hybrid PrxV [23] and *E. coli* Tpx [30] show it to be a distinct type of dimer that has a loop-based interface. We call this the A-type interface (for alternate). Interestingly, the decamers are built up using both these interfaces, and they dissociate into the B-type dimers during the catalytic cycle [22].

In *P. falciparum*, four Prxs have so far been described: two 2-Cys Prx (TPx1 and TPx2) [31-34], a classic 1-Cys Prx [33,35], and a GPx-like thioredoxin peroxidase (TPxGl) [36]. A TPx1 knock out leads to parasites viable but more sensitive to ROS and RNS [37] whereas the disruption of the complete Trx system by knocking out the central TrxR is lethal for the erythrocytic stages of *P. falciparum* [38]. TPx1 and TPxGl are clearly linked to the thioredoxin system whereas the *in vivo* reducing partners of TPx2 and the 1-Cys Prx remain to be established.

Screening the *P. falciparum* genome, we identified a novel Prx that has high sequence similarity (> 30%) to a poplar Prx [28] and *H. influenzae* hybrid PrxV [23]. According to Hofmann *et al.* [15], this would place these Prxs in the PrxV-like branch of the family. This branch is named after human PrxV [39] and contains both 1-Cys and 2-Cys Prxs. The recombinantly produced protein has activity with *t*-BOOH, and is efficiently reduced by either Trx or Grx (Nickel *et al.* manuscript in preparation). Without yet knowing its physiological role, we designated it the "antioxidant protein", *Pf*AOP. The *Pf*AOP gene is preceded by an apparent apicoplast targeting sequence –

the apicoplast being an organelle similar to a chloroplast (but non-photosynthetic) that is unique to *Apicomplexa*. Because differences between *Pf*AOP and its human counterparts might be exploited for the development of novel anti-parasitic agents, we have undertaken structural studies of *Pf*AOP. Here we report the x-ray structure of recombinant *Pf*AOP.

A1.3 Results and Discussion

A1.3.1 Overall structure

The structure of *Pf*AOP was determined by molecular replacement, leading to a model with a final R/ R_{free} of 18.7%/ 21.8% at 1.80 Å resolution (Table A1.1). The large majority of the main chain is well–ordered with strong, clear electron density (Figure A1.1) but the termini and some loops are rather mobile. Residues not modeled due to weak electron density are the N-terminal His tag of both molecules, molecule A residues 22–29, 181 and 182 and molecule B residues 1, 2, 181 and 182. Based on a Luzzati analysis [40,41], the estimated coordinate accuracy of well-ordered parts of the protein is ca. 0.2 Å. This is consistent with the level of agreement between the two monomers in the asymmetric unit (overall r.m.s.d of 0.3 Å for 169 Cα atoms). An interesting difference between the two monomers involves their mobilities, with molecule B being better ordered for residues 15–35 and molecule A being better ordered for the active site helix α2 (residues 55–75) and for residues 170–175. In all cases, crystal-packing interactions are present that stabilize the better-ordered structure, implying that the less ordered example is the more representative of the level of mobility present in solution.

The structure of *Pf*AOP shows a strongly interacting dimer made up of the two 21 kDa monomers related by a non-crystallographic two-fold symmetry axis (Figure A1.2). Dynamic light scattering experiments (apparent M_r = 50 kDa) and analytical gel filtration (apparent M_r =37 kDa) reveal that *Pf*AOP is also present as a dimer in solution (data not shown), suggesting that the crystallographic dimer seen for *Pf*AOP is relevant *in vivo*. Also supporting this conclusion is the observation that this dimer forms via the same interface (A-type) that was seen for *H. influenzae* hybrid PrxV [23],

Table A1.1 **Data collection and refinement statistics**

	Nat
Data	
Resolution limits (Å)	100 – 1.80 (1.86 – 1.80)
Unique observations	33775
Multiplicity	14.3 (9.7)
Completeness (%)	99.9 (99.5)
Average I/σ	15.6 (3.5)
R_{meas} [a] (%)	10.4 (58.6)
$R_{mrgd\text{-}F}$ [a] (%)	6.8 (39.4)
Refinement	
Number of amino acid residues	350
Number of solvent atoms	277
Total number of atoms	3146
Average B (Å^2) of protein atoms	26
Average B (Å^2) of solvent atoms	39
R_{cryst} (%)	18.7
R_{free} (%)	21.8
R.m.s.d bond lengths (Å)	0.005
R.m.s.d bond angles (degrees)	1.3

Numbers in parentheses correspond to values in the highest resolution bin.
[a] R_{meas} is the multiplicity weighted merging R-factor and $R_{mrgd\text{-}F}$ is an indicator of the quality of reduced data [55].

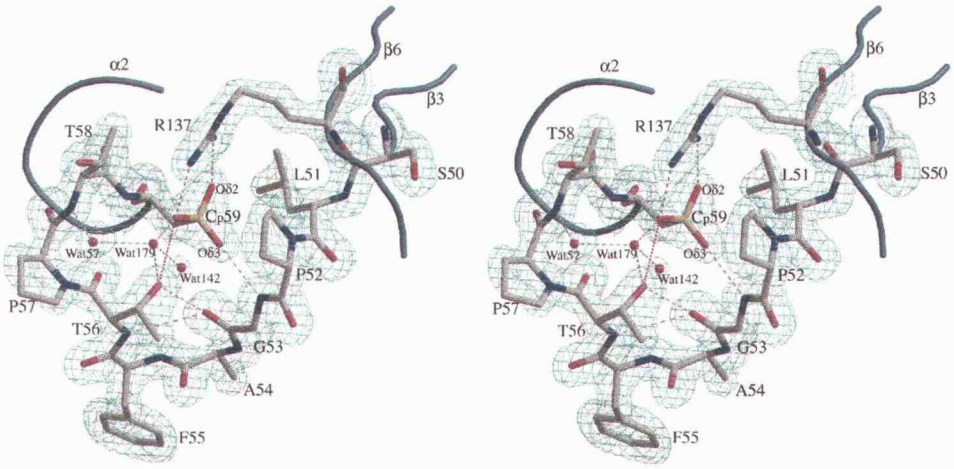


Figure A1.1 **Electron density map quality and active site structure.** A stereoview of the $2F_o–F_c$ electron density map contoured at 1.5 ρ_{rms} shows clear density for the cysteine sulfonic acid (C_P59) of monomer A. Atoms C_P59 – Oδ2 and – Oδ3 are identified, and Oδ1, the third C_P oxygen, is not labeled due to space. Three ordered water sites, the residues, and the secondary structures contributing to the active site are also labeled. Electron density for Wat179 appears at a lower contour level of 1.3 ρ_{rms}. Hydrogen bonds are indicated by broken gray lines and the close approach of Thr –Oγ to C_P59 –Oδ1 by red dotted lines. The view is similar to that of the upper monomer in Figure A1.2. The figure was prepared using Bobscript [65] and Raster3D [66].

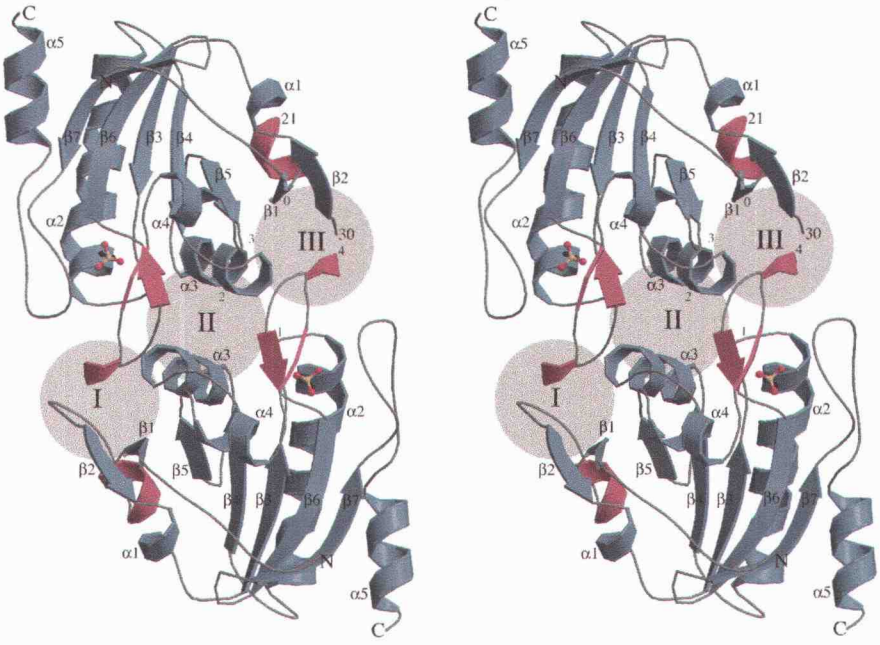


Figure A1.2 **The structure of the *Pf*AOP homodimer.** A stereo ribbon diagram of *Pf*AOP is shown looking down the 2-fold axis. The core secondary structure elements common to all Prxs (β1–β7 and α1–α5) are colored blue. The helix α1 may be an α- or a 3_{10}-helix in various Prxs. Additional *Pf*AOP β-strands and helices (α- and 3_{10}-) are colored red. Residues 21 and 30 bordering the mobile unmodeled region in the upper monomer backbone (monomer A) are labeled. The active site cysteines, seen as a sulfonic acid, are represented by ball-and-stick models. The family-oriented numbering scheme for secondary structures, focusing on the core elements and not on the additional secondary structural elements updates the scheme based on three structures that was presented by Alphey *et al.* [24]. The three clusters of interactions (I, II and III) at the dimer interface are indicated by gray shaded circles and the interacting regions, 0 through 4, are labeled with small green numbers. Note, clusters I & III are in front and cluster II is in the back. Thus the tilts, as discussed in the text (see Results and Discussion) and shown in Figure A1.7a, tilt the upper chain forward or backward compared to *Pf*AOP. The figure was prepared using Molscript [67] and Raster3D [66].

human PrxV [39,42], and *S. pneumoniae* (PDB code: 1PSQ), *E. coli* [30] and *H. influenzae* (PDB code: 1Q98) thiol peroxidases. The dimer interface buries ca. 910 Å^2 of surface area per monomer and is described further below.

With twelve peroxiredoxin structures (including *Pf*AOP) now available, we briefly describe the tertiary structure in a way that emphasizes the elements common to all Prxs (Figure A1.2). Essentially, the Prx fold consists of a central β-sheet of seven strands (β1 through β7) and five flanking helices (α1 through α5). Each individual Prx structure varies in the lengths of the common secondary structure elements and has additional secondary elements present as insertions. In addition to the Prx core elements, *Pf*AOP has a short helix inserted between β1 and β2, and two anti-parallel β-strands and a 3_{10}-helix inserted between α4 and β6.

The PrxV-like group of Prxs has both 2-Cys and 1-Cys representatives, so the presence of two cysteines in the *Pf*AOP sequence (Cys59 and Cys85) raised the question of its class. The structure of *Pf*AOP reveals that Cys59 is indeed the peroxidatic cysteine (C_P; present as cysteine sulfonic acid), but Cys85 is buried as part of the well-defined, core β-sheet where it is poorly suited to be the resolving cysteine. Also supporting the conclusion that *Pf*AOP is a 1-Cys Prx is that the known 2-Cys Prx, human PrxV (PDB code: 1HD2) [39] and a known 1-Cys Prx from poplar [43] both have Cys residues equivalent to Cys85 that are not involved in the catalytic mechanism.

A1.3.2 **Solvent structure**

Preferred hydration sites [44], also routinely called ordered water molecules, are an important part of protein structures and at 1.8 Å resolution these are reliably observed [45,46]. These sites are locations in the crystal that are favored for water binding. For enhanced information content [47], we have numbered the observed water sites from 1 to 277 according to their densities in the final $2F_o–F_c$ map (with water 1 having the highest density), and classified each water by its environment either as buried, crevice or surface (see Materials and Methods). 132 of these ordered waters (2 × 66) are at equivalent sites in both monomers (within 1.0 Å) [48,49], suggesting these can be considered a reproducible part of the protein structure. Interestingly these conserved waters are not all the highest density sites, but are spread evenly over the

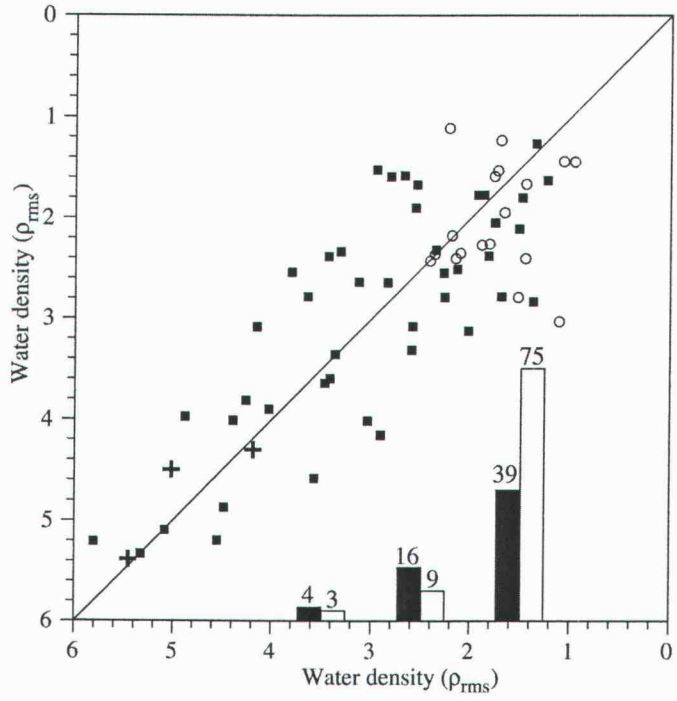


Figure A1.3 **Preferred hydration sites in the two monomers of *Pf*AOP.** The correlation of electron densities between 132 (2 × 66) equivalent water sites in the two monomers is plotted along the two axes. Waters are identified by their environment (see Materials and Methods) as buried (plusses), crevice (closed boxes) and surface (open circles). The diagonal line represents the theoretically perfect correlation between electron densities. The electron density information for 145 non-equivalent water sites are plotted as histograms with the crevice waters represented by closed bars and the surface waters by open bars. No non-equivalent waters are buried. For clarity, the few (four crevice and three surface) non-equivalent solvent molecules that have densities slightly lower than 1 ρ_{rms} have been grouped with the waters between 1 and 2 ρ_{rms}. Despite the water selection criteria, some water sites ended in densities slightly lower than 1 ρ_{rms}, but were retained because their reliability is equivalent to that of those with densities slightly greater than 1ρ_{rms}.

whole range of densities (Figure A1.3). Nevertheless, the high correlation (0.73) of the two sets of water densities supports the expectation that at this resolution, peak electron density strength is a fairly well defined physical property that reflects how well localized each water site is. Indeed, the densities of the conserved water sites are distinguished by environment, with all of the buried waters having very strong densities ($\geq 4\ \rho_{rms}$) and the crevice waters having stronger densities than the surface ones. This trend makes sense as buried waters make the most interactions with protein atoms, followed by crevice sites and then surface sites. Conversely, the non-conserved waters tend to have weaker densities (e.g. none have $\rho > 4\ \rho_{rms}$), with surface waters mainly located at the crystal lattice contacts (where they would not be expected to accurately reflect solution hydration properties), and crevice ones mainly associated with the weakly defined parts of the protein. The non-conserved waters also generally show higher densities for the crevice versus the surface waters. The three exceptionally well-ordered, non-conserved surface waters (densities $> 3\ \rho_{rms}$), form bridging hydrogen bonds at crystal contacts, and thus are not physiologically relevant despite their high density. In the active site and dimerization interface sections (see below), our discussion illustrates the value of having solvent sites numbered according to peak electron density.

A1.3.3 **Comparison of Prx structures**

In order to get the most insight from the *Pf*AOP structure, the eleven known Prx structures have been overlaid onto *Pf*AOP. The r.m.s Cα deviations per monomer range from about 0.8 to 2.9 Å and the sequence identities range from 13 to 72% (Table A1.2). Table A1.2 notes the type of oligomerization interfaces seen in the crystal structures, and for Prxs showing the type A interface, the angular differences in the dimer interaction, ranging from 4° to 42°, are given. Table A1.2 also indicates the Prx type of each structure based on the current classification [11,19]. Finally, Table A1.2 indicates the oxidation state of the active site cysteine for each structure, showing that the series of known structures cover all possibilities with 4 SH, 1 SOH, 1 SO_2H, 1 SO_3H, and 4 SS representatives. Because the variations between the structures are largely a result of sequence changes, we also show a structure-based sequence alignment for all the structures (Figure A1.4).

Table A1.2 **Comparison of various Prx structures** [a]

PDB code	Res.	Oligomer types	Prx. Type	Oxidation state		*Pf*AOP	*Hi*HyPrxV	PrxV	*Sp*Tpx	*Ec*TpxI	*Hi*Tpx	AhpC (oxd)	AhpC (red)	*Cf*TrpPx	PrxII	PrxI	PrxVI
This study	1.8 Å	A	1-Cys	SO_3^{2-}	***Pf*AOP**	--	1.8 (157) 36%	1.6(152) 28%	2.4 (142) 15%	2.7 (164) 14%	2.6 (142) 15%	2.5 (145) 13%	2.3 (146) 14%	2.5 (147) 18%	2.5 (146) 14%	2.5 (146) 16%	2.5 (144) 19%
1NM3	2.8 Å	A	1-Cys	SH	***Hi*HyPrxV**	4°	--	1.7(153) 33%	2.2 (139) 16%	2.8 (140) 18%	2.6 (142) 15%	2.2 (144) 19%	2.1 (140) 20%	2.3 (142) 14%	2.2 (141) 23%	2.4 (142) 22%	2.4 (140) 24%
1HD2 [b]	1.5 Å	A	Atypical 2-Cys	SH	**PrxV**	20°	22°	--	2.3 (145) 23%	2.4 (145) 17%	2.3 (145) 17%	2.2 (143) 24%	1.9 (142) 25%	2.0 (144) 22%	2.0 (144) 24%	2.0 (143) 22%	2.6 (148) 18%
1PSQ	2.3 Å	A	Atypical 2-Cys	SH	***Sp*Tpx**	32°	32°	37°	--	2.0 (155) 45%	1.8 (156) 44%	2.1 (145) 20%	1.9 (146) 20%	2.1 (148) 25%	2.0 (148) 24%	2.2 (146) 21%	2.2 (144) 19%
1QXH	2.2 Å	A	Atypical 2-Cys	SS	***Ec*TpxI**	30°	30°	38°	8°	--	1.0 (164) 61%	2.4 (139) 18%	2.3 (137) 17%	2.5 (142) 25%	2.5 (141) 24%	2.7 (141) 25%	2.9 (137) 18%
1Q98	1.9 Å	A	Atypical 2-Cys	SS	***Hi*Tpx**	27°	26°	38°	13°	9°	--	2.3 (142) 18%	2.2 (139) 17%	2.3 (143) 19%	2.4 (142) 18%	2.6 (143) 18%	2.7 (140) 17%
1KYG	2.5 Å	AB	Typical 2-Cys	SS	**AhpC (oxd)**	23°	21°	39°	23°	17°	12°	--	0.9 (163) 99%	1.5 (161) 32%	1.5 (162) 35%	1.6 (162) 33%	2.2 (158) 23%
1N8J	2.2 Å	AB	Typical 2-Cys	(SH) [c]	**AhpC (red)**	23°	20°	38°	23°	18°	12°	0°	--	1.4 (161) 32%	1.8 (185) 37%	1.7 (164) 30%	2.3 (179) 23%
1E2Y	3.2 Å	AB	Typical 2-Cys	SH	***Cf*TrpPx**	26°	23°	41°	23°	17°	11°	3°	4°	--	0.8 (166) 57%	1.3 (166) 57%	2.1 (157) 30%
1QMV	1.7 Å	AB	Typical 2-Cys	SO_2^-	**PrxII**	26°	23°	42°	27°	21°	15°	5°	5°	4°	--	1.3 (170) 72%	2.2 (179) 30%
1QQ2	2.6 Å	(A)B [d]	Typical 2-Cys	SS	**PrxI**	--	--	--	--	--	--	--	--	--	--	--	2.1 (157) 30%
1PRX	2.0 Å	B	1-Cys	SOH	**PrxVI**	--	--	--	--	--	--	--	--	--	--	--	--

Table A1.2. (continued)

[a] The right-hand portion of the table shows statistics for each pair of Prx. The numbers below the diagonal are the differences in interface angles for dimers having the A-type interface. The decameric Prxs have multiple, crystallographically distinct A-type interfaces; since the variations in the interface angles are small (< 4°), we report a representative value. Above the diagonal, the top line in each entry gives the Cα r.m.s deviations with the number of atoms used for the calculation in parentheses; the second line of each entry number is the percent sequence identity. The boxes represent the grouping of Prxs with similar interface angles, lower r.m.s.ds, and higher sequence identities.

[b] A crystal structure of human PrxV containing a disulfide linking two PrxV dimers (PDB code: 1OC3) has been recently published [42]. Biochemical data indicated the mechanism involves an intramolecular disulfide bond, so the disulfide-linked tetramer was unexpected. A model was proposed for the catalytically relevant intramolecular disulfide bonded structure (see Figure 1.4c, 4d of Evrard *et al.* [42]). We do not include 1OC3 here, because the crystal structure likely represents a crystallization artefact, and we did not have access to the proposed model of the catalytically relevant intramolecular disulfide form.

[c] The C46S mutant (active-site peroxidatic cysteine) of AhpC mimics the reduced state.

[d] Crystal structure shows only the B-type interface, but the protein is thought to function as an AB oligomer.

Figure A1.4 **Structure-based sequence alignment of selected Prxs.** The sequences of Prxs for which structures have been determined are aligned. Dots identify every tenth residue of *Pf*AOP with every other dot numbered. Residues involved in α-helices (light blue), 3_{10}-helices (dark blue) and β-strands (green) are indicated by coils (α- and 3_{10}-helices) and arrows (β-strands) above the sequences. The Prx core secondary structural elements are labeled accordingly. Regions involved in the type A and B oligomeric interfaces (see Results and Discussion) are also indicated by 'A' or 'B' below the sequences. For the A-type dimer interface, regions 0 through 4 are also specified. Four residues absolutely conserved in all Prxs are colored red, and the resolving Cys of the 2-Cys Prxs are printed in white on black.

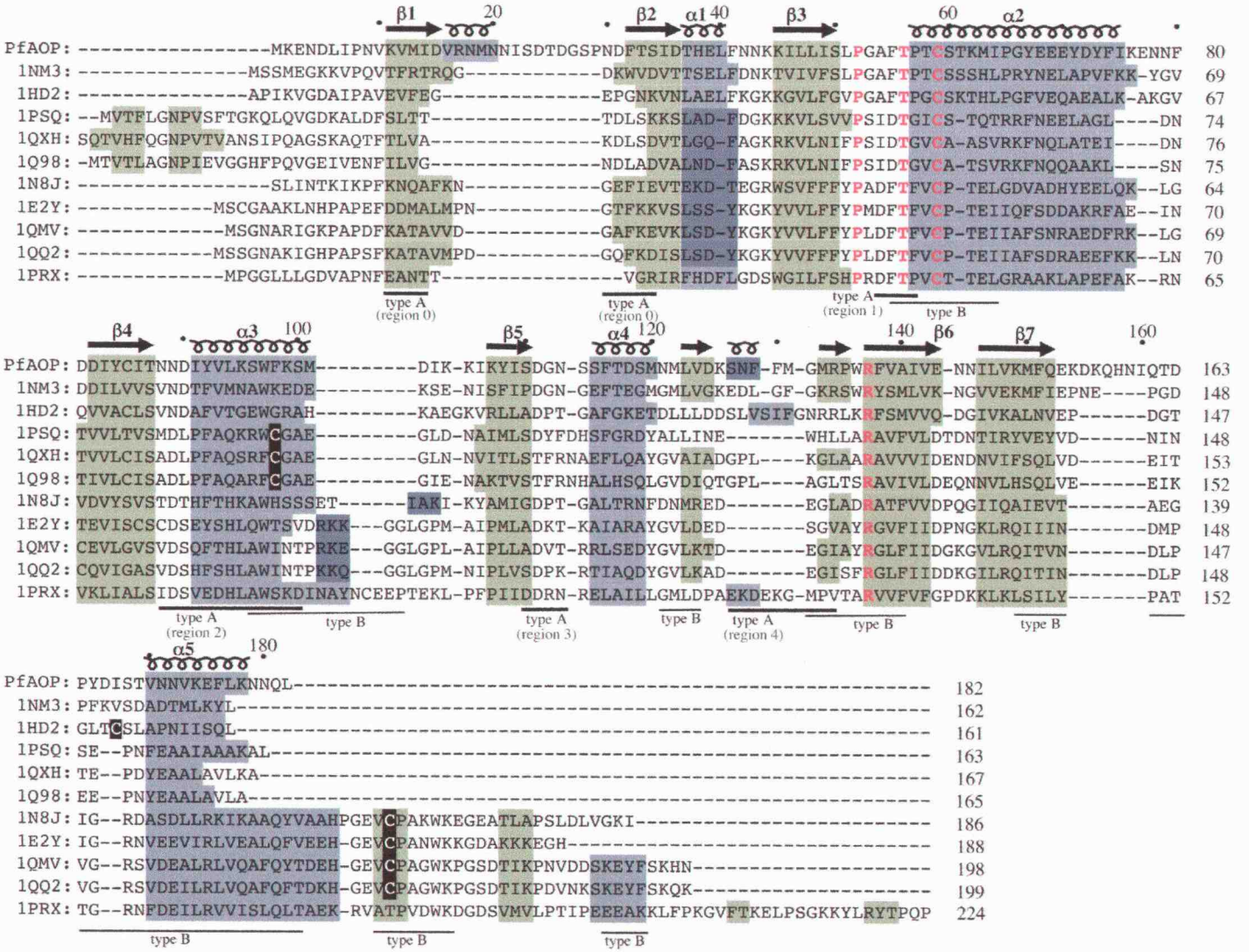

β1 20 β2 α1 40 β3 60 α2
PfAOP: ----------MKENDLIPNVKVMIDVRNMNNISDTDGSPNDFTSIDTHELFNNKKILLISLPGAFTPTCSTKMIPGYEEEYDYFIKENNF 80
1NM3: ----------MSSMEGKKVPQVTFRTRQG----------DKWVDVTTSELFDNKTVIVFSLPGAFTPTCSSSHLPRYNELAPVFKK-YGV 69
1HD2: ----------APIKVGDAIPAVEVFEG----------EPGNKVNLAELFKGKKGVLFGVPGAFTPGCSKTHLPGFVEQAEALK-AKGV 67
1PSQ: --MVTFLGNPVSFTGKQLQVGDKALDFSLTT----------TDLSKKSLAD-FDGKKKVLSVVPSIDTGICS-TQTRRFNEELAGL----DN 74
1QXH: SQTVHFQGNPVTVANSIPQAGSKAQTFTLVA----------KDLSDVTLGQ-FAGKRKVLNIFPSIDTGVCA-ASVRKFNQLATEI----DN 76
1Q98: -MTVTLAGNPIEVGGHFPQVGEIVENFILVG----------NDLADVALND-FASKRKVLNIFPSIDTGVCA-TSVRKFNQQAAKL----SN 75
1N8J: ----------SLINTKIKPFKNQAFKN----------GEFIEVTEKD-TEGRWSVFFFYPADFTFVCP-TELGDVADHYEELQK--LG 64
1E2Y: ----------MSCGAAKLNHPAPEFDDMALMPN----------GTFKKVSLSS-YKGKYVVLFFYPMDFTFVCP-TEIIQFSDDAKRFAE--IN 70
1QMV: ----------MSGNARIGKPAPDFKATAVVD----------GAFKEVKLSD-YKGKYVVLFFYPLDFTFVCP-TEIIAFSNRAEDFRK--LG 69
1QQ2: ----------MSSGNAKIGHPAPSFKATAVMPD----------GQFKDISLSD-YKGKYVVFFFYPLDFTFVCP-TEIIAFSDRAEEFKK--LN 70
1PRX: ----------MPGGLLLGDVAPNFEANTT----------VGRIRFHDFLGDSWGILFSHPRDFTPVCT-TELGRAAKLAPEFAK--RN 65
type A (region 0)
type A (region 0)
type A (region 1)
type B
β4 α3 100 β5 α4 120 140 β6 β7 160
PfAOP: DDIYCITNNDIYVLKSWFKSM---------DIK-KIKYISDGN-SSFTDSMNMLVDKSNF-FM-GMRPWRFVAIVE-NNILVKMFQEKDKQHNIQTD 163
1NM3: DDILVVSVNDTFVMNAWKEDE---------KSE-NISFIPDGN-GEFTEGMGMLVGKEDL-GF-GKRSWRYSMLVK-NGVVEKMFIEPNE----PGD 148
1HD2: QVVACLSVNDAFVTGEWGRAH---------KAEGKVRLLADPT-GAFGKETDLLLDDSLVSIFGNRRLKRFSMVVQ-DGIVKALNVEP------DGT 147
1PSQ: TVVLTVSMDLPFAQKRWCGAE---------GLD-NAIMLSDYFDHSFGRDYALLINE-------WHLLARAVFVLDTDNTIRYVEYVD------NIN 148
1QXH: TVVLCISADLPFAQSRFCGAE---------GLN-NVITLSTFRNAEFLQAYGVAIADGPL----KGLAARAVVVIDENDNVIFSQLVD------EIT 153
1Q98: TIVLCISADLPFAQARFCGAE---------GIE-NAKTVSTFRNHALHSQLGVDIQTGPL----AGLTSRAVIVLDEQNNVLHSQLVE------EIK 152
1N8J: VDVYSVSTDTHFTHKAWHSSSET------IAKI-KYAMIGDPT-GALTRNFDNMRED-------EGLADRATFVVDPQGIIQAIEVT-------AEG 139
1E2Y: TEVISCSCDSEYSHLQWTSVDRKK---GGLGPM-AIPMLADKT-KAIARAYGVLDED-------SGVAYRGVFIIDPNGKLRQIIIN-------DMP 148
1QMV: CEVLGVSVDSQFTHLAWINTPRKE---GGLGPL-AIPLLADVT-RRLSEDYGVLKTD-------EGIAYRGLFIIDGKGVLRQITVN-------DLP 147
1QQ2: CQVIGASVDSHFSHLAWINTPKKQ---GGLGPM-NIPLVSDPK-RTIAQDYGVLKAD-------EGISFRGLFIIDDKGILRQITIN-------DLP 148
1PRX: VKLIALSIDSVEDHLAWSKDINAYNCEEPTEKL-PFPIIDDRN-RELAILLGMLDPAEKDEKG-MPVTARVVFVFGPDKKLKLSILY-------PAT 152
type A (region 2)
type B
type A (region 3)
type B
type A (region 4)
type B
type B
α5 180
PfAOP: PYDISTVNNVKEFLKNNQL---------- 182
1NM3: PFKVSDADTMLKYL---------- 162
1HD2: GLTCSLAPNIISQL---------- 161
1PSQ: SE--PNFEAAIAAAKAL---------- 163
1QXH: TE--PDYEAALAVLKA---------- 167
1Q98: EE--PNYEAALAVLA---------- 165
1N8J: IG--RDASDLLRKIKAAQYVAAHPGEVCPAKWKEGEATLAPSLDLVGKI---------- 186
1E2Y: IG--RNVEEVIRLVEALQFVEEH-GEVCPANWKKGDAKKKEGH---------- 188
1QMV: VG--RSVDEALRLVQAFQYTDEH-GEVCPAGWKPGSDTIKPNVDDSKEYFSKHN---------- 198
1QQ2: VG--RSVDEILRLVQAFQFTDKH-GEVCPAGWKPGSDTIKPDVNKSKEYFSKQK---------- 199
1PRX: TG--RNFDEILRVVISLQLTAEK-RVATPVDWKDGDSVMVLPTIPEEEAKKLFPKGVFTKELPSGKKYLRYTPQP 224
type B
type B
type B

Figure A1.4

In the following two sections, we describe interesting features of the active site and at the dimerization interface of *Pf*AOP and use the overlaid set of structures to gain insight into structure-function relations of *Pf*AOP and peroxiredoxins in general.

A1.3.4 **Active site**

The peroxidatic active site of *Pf*AOP (see Figure A1.1) consists of residues preceding and in the first turn of helix α2 (the C_P–loop), and strands β3 and β6, and it includes four residues – Pro52, Thr56, Cys59 and Arg137 – that are absolutely conserved in all the known Prxs [19]. It has been thought that Thr56 positions the Cys facilitating an unidentified catalytic base to abstract a proton. The negatively charged Cys is then stabilized by the positively charged Arg and poised for a reaction with a peroxide. The Pro could play the structural role of shielding the reactive cysteine sulfenic acid intermediate from further oxidation by peroxides [19]. For *Pf*AOP, the active sites of the two monomers are very similar, and since they are separated by ca. 24 Å with no apparent interactions, each active site appears to act independently during catalysis.

Two features of the *Pf*AOP active site merit special mention, as they have not been described before. The first is the environment of the cysteine sulfonic acid (C_P59). To facilitate comparisons, we have named the three oxygens of the Cys-sulfonic acid so that the oxygen occupying the position similar to the Cys–sulfenic acid oxygen in Prx VI [29] is Oδ1, the oxygen that occupies the position of the second oxygen of Cys–sulfinic acid in PrxII [50] is Oδ2, and the remaining oxygen is Oδ3. All three oxygens make hydrogen bonds: C_P59–Oδ1 and –Oδ2 align well with Arg137–NH1 and –NH2 (Figure A1.1). C_P59–Oδ1 also forms a hydrogen bond with Wat179 and makes a close approach to Thr56–Oγ (2.9 Å); and C_P59–Oδ3 is stabilized by a hydrogen bond from the peptide backbone Gly53–NH (3.0 Å). A comparison of the structures shows that in order to accommodate the extra oxygen on the cysteine of *Pf*AOP the main chain containing Gly53 is shifted out slightly. In all the other Prx structures, this amide group is buried forming a long hydrogen bond (ranging from 3.4 to 3.8 Å) with the Cys sulfur atom.

Comparing the structure of *Pf*AOP C_P59 with the structures of PrxII and PrxVI sheds light into the possible steps that occur during over-oxidation. The peroxide substrate binds to the active site pocket with one of its oxygens present near the position of Oδ1. The Cys is then oxidized to the –SOH intermediate. For each further oxidation step, it is necessary to accommodate another incoming peroxide molecule. For the formation of $-SO_2H$, this would be accomplished by the oxygen on the cysteine rotating to and occupying the position of Oδ2, and for the formation of $-SO_3H$, the two oxygens would rotate further such that they occupy the positions of Oδ2 and Oδ3 thereby leaving the position of Oδ1 open for the binding of the peroxide. This structural insight raises the likelihood that susceptibility of any given Prx to over-oxidation to the sulfinic or sulfonic states will depend not just on the thermodynamics of active site unfolding as discussed by Wood *et al.* [12], but also on the ease with which the structure can accommodate oxygens in the Oδ2 and/ or Oδ3 positions.

In terms of the normal catalytic cycle, one additional insight is that the surroundings of C_P59–Oδ1 indicate the presence of sub-optimal interactions (see Figure A1.1). First, the Thr56–Oγ to C_P59–Oδ1 close approach (2.8 Å) is not a hydrogen bond but a repulsive interaction because the hydrogen on Thr56 will be taken in the geometrically favorable hydrogen bond with Gly53–O (2.7 Å). Second, Wat179 is a relatively weak water site, indicating that it has a low occupancy or high mobility (or both). While an equivalent solvent molecule for Wat179 is not modeled in monomer B, there is weak (2.5 ρ_{rms}) difference density at the equivalent position in molecule B consistent with the presence of a similarly low occupancy water site that due to the higher disorder of this part of molecule B, did not meet the criteria for inclusion in the model. Wat179 is held in a hydrogen-bonding network involving Thr56–Oγ, Wat57, Wat142 and C_P59–Oδ1 with very short distances of only 2.5 Å from both Wat57 and Wat142, suggesting it is squeezed in a space not quite large enough for it. Finally, the structure as seen leaves the buried C_P59–NH with no good H-bonding partner. Taken together, the suboptimal H-bonding network around C_P59–Oδ1 (the low occupancy of Wat179, the close contact Thr56–Oγ – C_P59–Oδ1 and the unfulfilled H-bonding of C_P59–NH) suggests that oxidation to –SOH

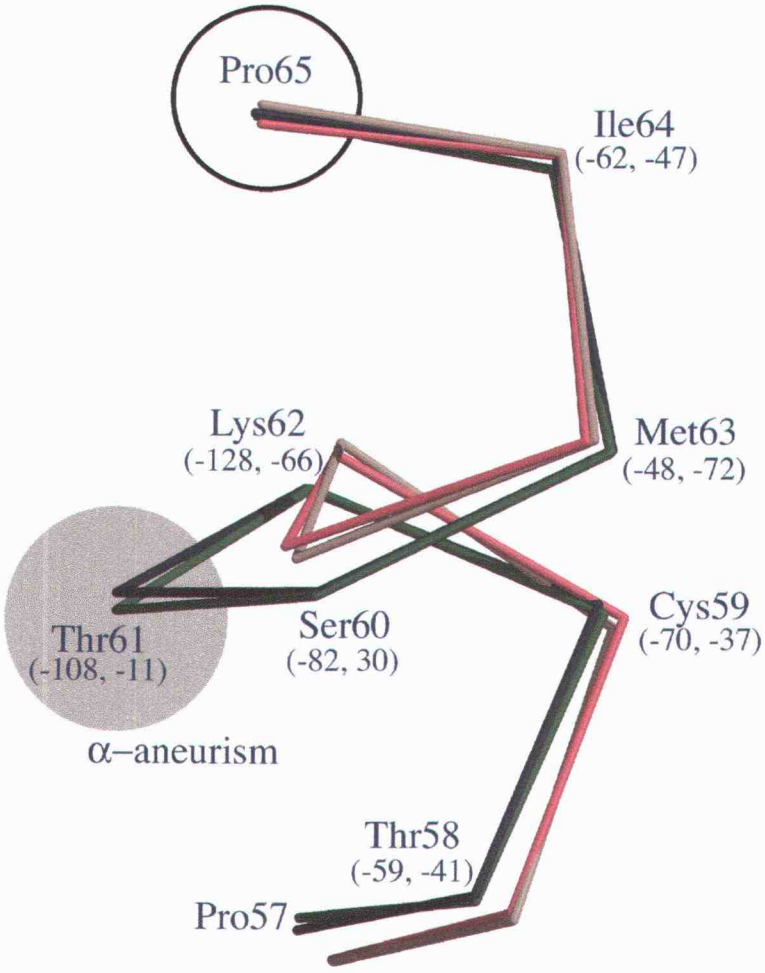


Figure A1.5 **An α-aneurism near the active site of *Pf*AOP.** The Cα atoms of the α2-helix of *Pf*AOP (blue), human PrxV (green), AhpC (red), and PrxII (light brown) are overlaid to illustrate the α-aneurism in *Pf*AOP and human PrxV. A shaded and a transparent circle indicate the position of the α-aneurism and the Pro residue conserved in PrxV-like Prxs respectively. The *Pf*AOP atoms are labeled and their ϕ, ψ angles are noted in parentheses. The ϕ, ψ angles of human PrxV are all within 20° of those of *Pf*AOP. The hydrogen bonding patterns and the torsion angles in *Pf*AOP and human PrxV are very similar to those described for the prototypical α-aneurism [51]. The figure was prepared using Molscript [65] and Raster3D [66].

destabilizes the active site. This could be a feature that stimulates local unfolding of the –SOH oxidized active site which would enhance the rate of disulfide formation (resolution) and decrease the rate of over-oxidation. The higher mobility of the active site C_P–loop in monomer B is consistent with (in the absence of crystal packing interactions) this loop fluctuating between locally unfolded and fully folded conformations.

The second special feature of the *Pf*AOP active site is that the helical turn following C_P59 contains a single residue insertion (Figure A1.5). This so-called α-aneurism [51] is also present in the *H. influenzae* hybrid PrxV and human PrxV (Figure A1.4), but was not explicitly described in those papers. Remarkably, this disruption in the α-helix H-bonding pattern does not significantly change the active site geometry. Also, the presence of Pro65, exactly one α-helical turn past the aneurism (Figure A1.5), does not disrupt the helical geometry, and we suggest its conservation in PrxV-like Prxs [52] is related to the presence of the aneurism. We do not have evidence as to how the insertion affects function, but it may influence the local unfolding required for catalysis. The insertion along with the conserved Pro, thus far seen only in PrxV-like Prxs, could provide a useful marker for helping trace the evolution of Prxs.

In terms of drug design, the active site comparisons are not very encouraging. The overlaid structures show that the active site residues and geometry are highly conserved in all Prxs. For these reasons, we suspect that a strongly specific active site-directed inhibitor would be difficult to design.

A1.3.5 **A-type dimerization interface**

To provide a context for the following discussion, we will describe the A-type dimer interface of *Pf*AOP. Following the nomenclature developed by Wood *et al.* [22] for the decamer building interface of AhpC, there are five regions that are involved in the A-type dimerization interface: region 1 (residues 54–57), region 2 (residues 88–100), region 3 (residues 110–113), region 4 (residues 127–134), and an N-terminal region (residues 11–14 and 34) we term here as region 0 (Wood *et al.* [22] did not name this region because in AhpC it only included a single residue).

The dimer interface can be divided into three main clusters of interactions shown as I, II and III in Figure A1.2. The symmetry-related clusters of I and III are present at the periphery of the interface and are formed by residues from regions 0 and 3 on one chain interacting with residues from region 4 on the other chain (Figure A1.6a). Phe130 from one monomer and Asn113 from the other appear to be key residues in these clusters.

Cluster II makes up the core of the dimerization interface and is formed mostly by residues from regions 1 and 2 from both chains coming together. In addition, Met131 and Arg134 from region 4 are involved. Since the cluster surrounds the non-crystallographic two-fold axis, identical residues from the two monomers are involved (Figure A1.6b). Similar to cluster I, stabilization includes both hydrophobic and hydrogen-bonded interactions, with key hydrophobic residues being Phe55, Tyr92, and Val93, and key hydrogen-bonded residues being Asn89, Asp90, and Arg134. The water molecules involved –Wat1, Wat2, Wat3, Wat5, Wat7 and Wat8 – are among the best-defined solvent sites in the structure indicating that the interface is quite well ordered.

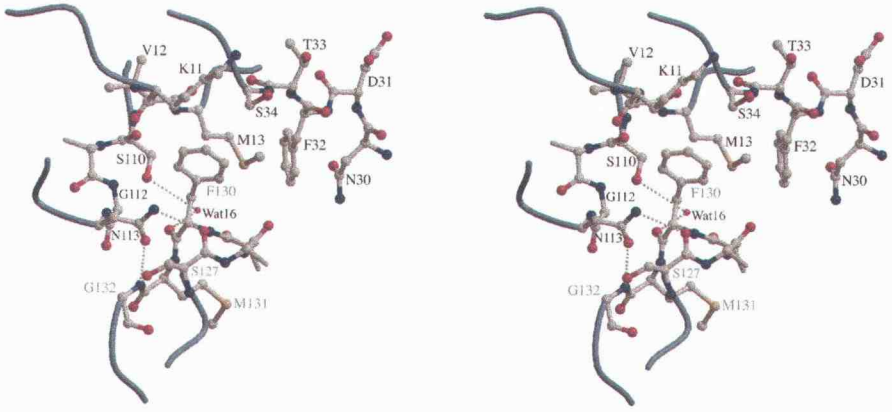
For the nine other structures having an A-type interface (Table A1.2), the same sequence regions are involved, but the interfacial angle varies by up to 42°. A comparison of the structures and the sequences provides an explanation for the differences. Based on both the interface angles and on the sequence identities, the ten Prxs showing A-type interfaces can be divided into three groups. The PrxV-like Prxs form one group with hybrid PrxV and *Pf*AOP sharing nearly identical dimerization interfaces, and human PrxV being tilted by ca. 20° compared to them (Figure A1.7a). The tilt is due to a 2-residue insertion in human PrxV region 4 (clusters I/ III) compared to *Pf*AOP that effectively pushes the interaction region 3 away.

The second group includes the four Prxs that form both type A and B interfaces (i.e., the decameric Prxs). These proteins all have a five-residue deletion in region 4 compared to *Pf*AOP, and this leads to a ca. 20° tilt in a direction opposite to that of human PrxV (Figure A1.7a). This explains how they are ca. 20° different from *Pf*AOP and ca. 40° different than human PrxV.

Figure A1.6 **The dimerization interface of *Pf*AOP.** Stereoviews of a) cluster III (2-fold related to cluster I) and b) cluster II showing the interactions at the dimerization interface. In panel a, the residues for the different monomers are indicated by black and green labels. In panel b, only residues from one monomer are labeled. For clarity Wat8 is not labeled, but the equivalent water across the pseudo 2-fold axis, Wat7, has been. Dotted gray lines indicate hydrogen bonds. Met131 is shown in both, clusters III and II as it constitutes a small overlap between the clusters. Residues with buried surface area $\geq$ 30 Å^2 are (in parentheses are buried surface areas in Å^2): Pro57 (30), Ser96 (30), Met131 (30), Arg134 (30), Gly112 (35), Met13 (40), Val93 (45), Asp90 (65), Phe55 (85), Tyr92 (90), Asn113 (90) and Phe130 (175). The views are similar to that of Figure A1.2. The figure was prepared using Molscript [65] and Raster3D [66].

Figure A1.6

(a)



(b)

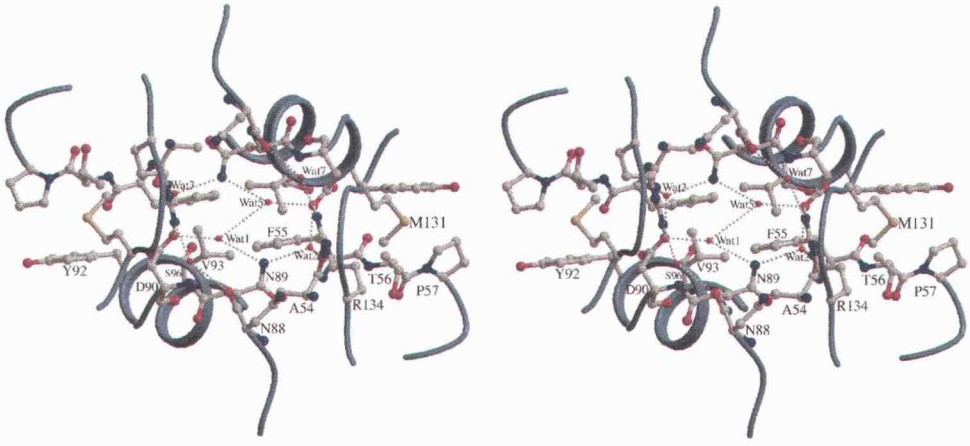
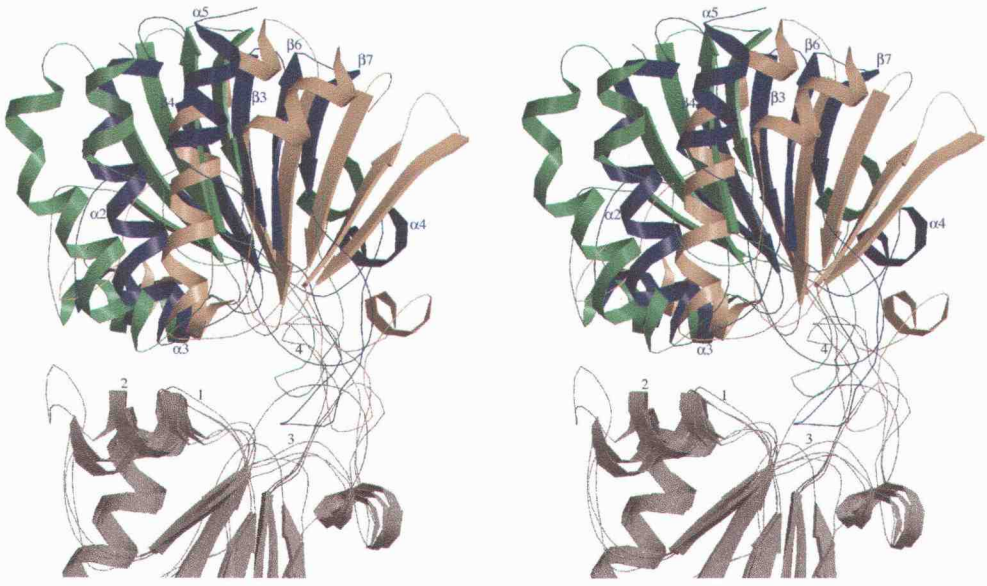


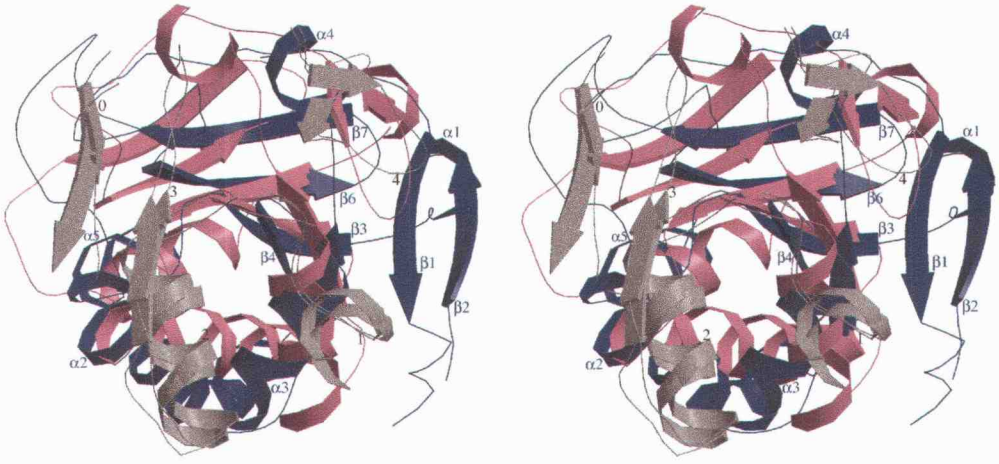
Figure A1.7 **Variations in the A-type dimerization interface.** Stereo ribbon diagrams illustrating the a) tilt and b) twist at the dimerization interface. The reference subunit on which the overlay is based (see Materials and Methods) is colored gray (and only partially shown), and the orientations of the second subunit of the dimers (in color) are compared. In both panels, the *Pf*AOP structure is blue. Most core secondary structural elements are labeled in the colored *Pf*AOP chain, as are regions 0 through 4 in the gray chain. In panel a, human PrxV (green) and PrxII (yellow) are seen to tilt ca. 20° relative to *Pf*AOP in opposite directions around a similar axis. To preserve clarity, the Prx core elements of β1, β2 and α1 are not shown. In panel b, *S. pneumoniae* Tpx (red) is twisted relative to *Pf*AOP ca. 30° counter clockwise. For reference, to generate these views, the dimer in Figure A1.2 would be rotated clockwise to orient the dimer interface horizontally. Then, panel a is equivalent to viewing this molecule from the left side (the dimer axis is horizontal), and panel b is equivalent to viewing it from below (the dimer axis is vertical).

Figure A1.7

(a)



(b)



The final group consists of the three bacterial thiol peroxidase (Tpx) structures. They are not tilted, but are twisted by ca. 30° compared to *Pf*AOP. While in the other cases, an insertion/ deletion in clusters I/ III led to a tilt, for this group it is a dramatic change in the core of cluster II that leads to the twist. Choi *et al.* [30] observed that in *E.coli* Tpx, an Arg-Asp salt bridge forms an important interaction at the dimer interface. Indeed, the sequence alignment (Figure A1.4) shows that the Asp in thiol peroxidases substitutes the key interface residue Phe55 (see Figure A1.6b) in *Pf*AOP. This change is complemented by the substitution of Ser96 of *Pf*AOP by the Arg and the resulting Arg-Asp salt bridge forms a highly restructured (and twisted) core of the interface (Figure A1.7b). At the same time, the packing of clusters I/ III minimizes changes in the tilting.

Interestingly, even though the B-type dimer interface was observed first, Table A1.2 shows that the A-type interface is much more widespread, associated with all Prxs but human PrxVI. Also, it has been suggested that due to the juxtaposition of this interface with the peroxidatic active site, that oligomerization using this interface may be linked with catalytic activity [22]. For these reasons, we hypothesize that the A-type dimerization interface preceded the B-type interface in evolution. (If true, the "A" could stand for "ancestral" rather than "alternate").

With regards to drug design, if the A-type dimer formation is linked with activity then it may be possible to develop inhibitors directed against dimerization such as was done for human and *P. falciparum* glutathione reductase [53,54]. While the core of the interface is structurally similar to other Prxs, at the periphery of the interface residues 30 and 31 make dimerization interactions and are part of a loop that is unique to *Pf*AOP. Though these interactions are not extensive, the uniqueness of the loop makes it a potential target for structure-based drug design.

A final point of interest we have derived from these comparisons is a sequence feature associated with the Prxs that form the type B dimer (Figure A1.4). The clear correlation is that proteins forming B-type dimers have longer chains and more specifically a longer helix α5. This helix is a major player in the B-type interface, as it and its symmetry mate pack together with each other and with the interfacial β-strands to build the major hydrophobic core at the interface. Using this longer helix as a

marker for predicting dimerization mode could help understand functionality and, along with the α-aneurism could help reconstruct the evolutionary history of this ubiquitous and physiologically important family of enzymes.

A1.4 Materials and Methods

A1.4.1 Amplification, cloning, overexpression and purification

By screening the PlasmoDB (www.plasmodb.org) database an open reading frame with homologous to Prxs was identified and named "antioxidant protein" (GenBank accession number **AY306209**). A PCR product corresponding to the sequence coding for the mature protein (without predicted targeting sequence) was generated with the primers OpfAOPhH: 5′ GCGCAAGCTTTTATAACTGATTATTTTTTAAAAACTCTTTTAC 3′ and OpfAOPvB: 5′ CGCGGGATCCAAAGAAAATGATCTTATTCCTAACG 3′. A blood stage cDNA library was used as a template. The conditions of the amplification reaction were 95 °C, 3 min; 95 °C, 30 s; 50 °C, 30 s; 72 °C, 1 min, 35 cycles; 72 °C, 2 min. The 546 bp product was cloned into the expression vector pQE30 via *Bam*HI and *Hin*dIII restriction sites. Clones were controlled by sequencing of both strands. The plasmid containing the correct fragment was transformed into *E. coli* M15 cells and overexpressed. Using the N-terminal hexahistidyl tag provided by the pQE vector, recombinant AOP was purified to homogeneity via Ni-NTA-agarose. Further details of purification are to be published elsewhere with the biochemical characterization of *Pf*AOP.

A1.4.2 Crystallization

Crystals of purified recombinant *Pf*AOP were grown using the hanging-drop vapor-diffusion method. Prior to crystallization, the protein was dialyzed against 1 mM EDTA, 30 mM Tris, pH 8.0. Rod-like crystals growing to a final size of ca. 0.25 x 0.05 x 0.05 mm^3 were obtained by mixing the protein (14 mg/ml) with an equal volume of the reservoir solution consisting of 20% (v/v) iso-propanol, 20% (w/v) PEG 4000, 100 mM sodium citrate, pH 5.6 at 4 °C. Crystals were stored in the reservoir solution, and the reservoir solution plus 5% glycerol, added to compensate for the rapid evaporation of iso-propanol, served as a cryo-protectant.

A1.4.3 **Data collection**

For data collection (at –170 °C), crystals were flash frozen in loops by dipping into liquid nitrogen. The crystals belong to the space group $P2_12_12_1$ with a = 42.5 Å, b = 79.8 Å, c = 108.5 Å; they have two *Pf*AOP molecules in the asymmetric unit and a solvent content of 47%.

Four native data sets were collected in all. First, a 2.7 Å data set (Nat1) was collected in our laboratory (Raxis IV detector and Cu-K_α radiation from a Rigaku rotating anode generator running at 50 kV, 100 mA; $\Delta\phi = 1°$, 140 15 min images). Second a partial data set (Nat2) extending to 2.4 Å was collected using the in-house source ($\Delta\phi = 1°$, 55 25 min images). This data collection was interrupted so that the crystal could be saved to collect better data at a synchrotron source. Third, the same crystal was used for a complete data set (Nat3), useful to ~1.85 Å resolution, at beamline 8.2.2 of the Advanced Light Source (ALS, Lawrence Berkeley National Laboratory; $\lambda = 1.0$ Å, $\Delta\phi = 1.5°$, 120 4 s images). Fourth, a data set on a different crystal (Nat4), also useful to ~1.85 Å, was collected at beamline 8.2.1 of the ALS (high resolution pass: $\lambda = 1.0$ Å, $\Delta\phi = 0.8°$, 115 15 s images; and low resolution pass: $\lambda = 1.0$ Å, $\Delta\phi = 1.0°$, 115 5 s images). Knowing that increasing redundancy improves data quality, Nat2, Nat3 and Nat4 were merged to obtain the data used for refinements (Nat). The final statistics, especially $R_{rmgd\text{-}F}$, an indicator of the quality of the reduced data [55], showed the merged data set was usable out to 1.80 Å resolution (Table A1.1). All data sets were processed using the HKL suite of programs [56].

A1.4.4 **Structure determination**

The three dimensional structure of the *Pf*AOP was solved by molecular replacement using the original Nat1 data and a direct Patterson search as implemented in the program CNS [57]. A database search revealed the Prx domain of *Haemophilus influenzae* hybrid PrxV (PDB code: 1NM3) [23] as the known structure with the highest sequence similarity (~36%). For molecular replacement, the non-Gly residues were changed to Ala and the resulting structure (one chain) was used as a search model against data from 15–4 Å resolution. The top 200 rotation function solutions were subjected to a translation search that yielded a clear solution (for the first molecule a

correlation coefficient of 0.221 versus 0.180 for the next best solution, and for the second molecule a correlation coefficient of 0.387 versus 0.294 for the next best solution). With 10% of the reflections set aside for cross-validation, automated rigid-body, simulated annealing using torsion angles [57-59], and individual B-factor refinement yielded an R and R_{free} of 0.40 and 0.48 respectively. We then switched to the combined Nat data set (the lower resolution cross-validation reflections were maintained and extended) and continued automated refinement at 1.80 Å resolution to yield an $R = 0.32$ and $R_{free} = 0.42$. The resulting electron density maps allowed us to build almost all the side chains, multiple conformations, visible insertions and deletions using the molecular graphics program O [60]. Water molecules were placed by the water pick utility in CNS [57] using the following criteria: (1) a peak of ≥ 3 ρ_{rms} in the F_o–F_c map and a peak of ≥ 1 ρ_{rms} in the $2F_o$–F_c map, and (2) a distance of ≥ 2.5 Å and ≤ 3.5 Å to nearby hydrogen bond donor or acceptor.

Alternate side chain conformations were observed for molecule A residues Met13, Met63, Ser96 and Ser127 and molecule B residues Met13 and Ser34. To obtain the occupancies of each pair of alternate conformations, a set of test refinements with various occupancies were carried out and the one that yielded similar B-factors for the alternate conformations was retained. Nine residues with weak side-chain density were modeled as alanines: A21, A178-A180, and B3, B4, B178-B180. Further rounds of positional and B-factor refinement, coupled with manual rebuilding, yielded final R and R_{free} values of 18.7% and 21.8% respectively. No non-crystallographic symmetry restraints were used during refinement. A Ramachandran plot, using the definition of Kleywegt and Jones [61], has 97% of the residues in the core region and 3% in the non-core region. Data collection and refinement statistics are presented in Table A1.1.

A1.4.5 Structural comparisons and analyses

Secondary structure assignments were made using DSSP [62]. Structure-based sequence alignments were done using SEQUOIA [63] with a distance cutoff of 3.0 Å.

For assigning the waters to the three groups – buried, crevice and surface – the molecular surface and cavity characterization program, VOIDOO [64] was used. First, after excluding waters, a molecular surface of the protein was calculated using a probe

radius of 1.0 Å. Waters that were present in fully internal cavities defined in this manner were classified as buried waters. In order to distinguish crevice versus surface waters, another molecular surface was calculated with a probe radius of 3.0 Å. All non-buried waters enclosed at a distance greater than 1 Å of the surface calculated using the 3 Å-radius probe were classified as crevice waters. Accordingly, all waters close to (within 1.0 Å) or outside the surface defined by the 3 Å-radius probe were designated as surface waters.

In order to calculate the interfacial angles, dimers of Prxs were rotated onto one another, using a matrix calculated based on only one of the monomers. The angle between the non-overlaid monomers was classified as the change in interfacial angle. Visual inspection was used to differentiate between the kind of deviations – tilt and twist.

A1.4.6 **Dynamic light scattering and gel filtration**

Dynamic light scattering data for *Pf*AOP (2 mg/ml in 1 mM EDTA, 30 mM Tris, pH 8.0) were obtained at 4 °C using the DynaPro-MSTC40 instrument (Protein Solutions). For measurement, the sample was filtered through a Whatman 0.02 μm filter disk and manually injected into a 12 μL quartz cuvette and the scattered photons measured at a fixed, preset angle of 90° with an incident laser beam of 830 nm wavelength. Data analyses were carried out using the software Dynamics, provided with the instrument. The monomodal curve assuming a single macromolecular species was used to estimate the molecular weight.

*Pf*AOP dimerization in solution (4 mg/ml in 300 mM NaCl, 50 mM Na-phosphate, pH 8.0) was studied by analytical gel filtration chromatography employing a ÄKTA FPLC system (Amersham-Pharmacia) and a Superdex-75 HiLoad 16/ 60 size exclusion column (120 mL bed volume). BSA (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa), and RNase A (13.7 kDa) were taken as molecular weight standards. Blue dextran (~2 kDa) and water indicated the exclusion volume (V_0) and the total volume (V_t) of the columns respectively. The apparent M_r was evaluated by plotting log (M_r) *vs* K_{av}, where $K_{av} = (V_e - V_0)/ (V_t - V_0)$ and V_e is the observed elution volume.

A1.4.7 ***Coordinates***

The coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org/pdb/) as entry 1XIY.

A1.4.8 **Acknowledgements**

We would like to thank Dr. Anthony Addlagatta for collecting data (Nat3) at the Advanced Light Source (ALS). We would also like to thank Drs. Zac Wood and Rick Faber for helpful discussions. The invaluable help rendered by the staff at the ALS is also appreciated. This work was supported by a sub-contract to P.A.K of NIH grant GM-50389 to Dr. Leslie B. Poole and by the Deutsche Forschungsgemeinschaft (SFB 535 project A12) to K.B. The authors wish to acknowledge the Proteins and Nucleic Acids Core facility of the Environmental Health Sciences Center at Oregon State University (NIEHS grant P30 ES00210). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U. S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

A1.5 **References**

1. Greenwood, B. & Mutabingwa, T. (2002). Malaria in 2002. *Nature* 415, 670-2.

2. Sachs, J. & Malaney, P. (2002). The economic and social burden of malaria. *Nature* 415, 680-5.

3. Olliaro, P. (2001). Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* 89, 207-19.

4. Hyde, J. E. (2002). Mechanisms of resistance of *Plasmodium falciparum* to antimalarial drugs. *Microbes Infect.* 4, 165-74.

5. Ridley, R. G. (2002). Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 415, 686-93.

6. Becker, K., Tilley, L., Vennerstrom, J. L., Roberts, D., Rogerson, S. & Ginsburg, H. (2004). Oxidative stress in malaria parasite-infected erythrocytes: host-parasite interactions. *Int. J. Parasitol.* 34, 163-89.

7. Becker, K., Rahlfs, S., Nickel, C. & Schirmer, R. H. (2003). Glutathione - functions and metabolism in the malarial parasite *Plasmodium falciparum*. *Biol. Chem.* 384, 551-66.

8. Rahlfs, S., Schirmer, R. H. & Becker, K. (2002). The thioredoxin system of *Plasmodium falciparum* and other parasites. *Cell. Mol. Life Sci.* 59, 1024-41.

9. Henkle-Duhrsen, K. & Kampkotter, A. (2001). Antioxidant enzyme families in parasitic nematodes. *Mol. Biochem. Parasitol.* 114, 129-42.

10. Kim, K., Kim, I. H., Lee, K. Y., Rhee, S. G. & Stadtman, E. R. (1988). The isolation and purification of a specific "protector" protein which inhibits enzyme inactivation by a thiol/Fe(III)/O2 mixed-function oxidation system. *J. Biol. Chem.* 263, 4704-11.

11. Chae, H. Z. & Rhee, S. G. (1994). A thiol-specific antioxidant and sequence homology to various proteins of unknown function. *Biofactors* 4, 177-80.

12. Wood, Z. A., Poole, L. B. & Karplus, P. A. (2003). Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* 300, 650-3.

13. Bryk, R., Griffin, P. & Nathan, C. (2000). Peroxynitrite reductase activity of bacterial peroxiredoxins. *Nature* 407, 211-5.

14. Flohé, L., Budde, H. & Hofmann, B. (2003). Peroxiredoxins in antioxidant defense and redox regulation. *Biofactors* 19, 3-10.

15. Hofmann, B., Hecht, H. J. & Flohé, L. (2002). Peroxiredoxins. *Biol. Chem.* 383, 347-64.

16. Jin, D.-Y. & Jeang, K.-T. (2000). Peroxiredoxins in cell signaling and HIV infection. In *Antioxidant and redox regulation of genes* (Sen, C. K., Sies, H. & Baeuerle, P. A., eds.), pp. 381-407. Academic Press.

17. Forman, H. J. & Cadenas, E. (1997). *Oxidative stress and signal transduction*, Chapman & Hall.

18. Fujii, J. & Ikeda, Y. (2002). Advances in our understanding of peroxiredoxin, a multifunctional, mammalian redox protein. *Redox Rep.* 7, 123-30.

19. Wood, Z. A., Schroder, E., Robin Harris, J. & Poole, L. B. (2003). Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem. Sci.* 28, 32-40.

20. Ellis, H. R. & Poole, L. B. (1997). Novel application of 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole to identify cysteine sulfenic acid in the AhpC component of alkyl hydroperoxide reductase. *Biochemistry* 36, 15013-8.

21. Ellis, H. R. & Poole, L. B. (1997). Roles for the two cysteine residues of AhpC in catalysis of peroxide reduction by alkyl hydroperoxide reductase from *Salmonella typhimurium*. *Biochemistry* 36, 13349-56.

22. Wood, Z. A., Poole, L. B., Hantgan, R. R. & Karplus, P. A. (2002). Dimers to doughnuts: redox-sensitive oligomerization of 2-cysteine peroxiredoxins. *Biochemistry* 41, 5493-504.

23. Kim, S. J., Woo, J. R., Hwang, Y. S., Jeong, D. G., Shin, D. H., Kim, K. & Ryu, S. E. (2003). The tetrameric structure of *Haemophilus influenza* hybrid Prx5 reveals interactions between electron donor and acceptor proteins. *J. Biol. Chem.* 278, 10790-8.

24. Alphey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. (2000). The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2-Cys-peroxiredoxins. *J. Mol. Biol.* 300, 903-16.

25. Bryk, R., Lima, C. D., Erdjument-Bromage, H., Tempst, P. & Nathan, C. (2002). Metabolic enzymes of mycobacteria linked to antioxidant defense by a thioredoxin-like protein. *Science* 295, 1073-7.

26. Poole, L. B., Reynolds, C. M., Wood, Z. A., Karplus, P. A., Ellis, H. R. & Li Calzi, M. (2000). AhpF and other NADH:peroxiredoxin oxidoreductases, homologues of low Mr thioredoxin reductase. *Eur. J. Biochem.* 267, 6126-33.

27. Seo, M. S., Kang, S. W., Kim, K., Baines, I. C., Lee, T. H. & Rhee, S. G. (2000). Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate. *J. Biol. Chem.* 275, 20346-54.

28. Rouhier, N., Gelhaye, E., Sautiere, P. E., Brun, A., Laurent, P., Tagu, D., Gerard, J., de Faÿ, E., Meyer, Y. & Jacquot, J. P. (2001). Isolation and characterization of a new peroxiredoxin from poplar sieve tubes that uses either glutaredoxin or thioredoxin as a proton donor. *Plant Physiol.* 127, 1299-309.

29. Choi, H. J., Kang, S. W., Yang, C. H., Rhee, S. G. & Ryu, S. E. (1998). Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. *Nat. Struct. Biol.* 5, 400-6.

30. Choi, J., Choi, S., Cha, M. K., Kim, I. H. & Shin, W. (2003). Crystal structure of *Escherichia coli* thiol peroxidase in the oxidized state: insights into intramolecular disulfide formation and substrate binding in atypical 2-Cys peroxiredoxins. *J. Biol. Chem.* 278, 49478-86.

31. Rahlfs, S. & Becker, K. (2001). Thioredoxin peroxidases of the malarial parasite *Plasmodium falciparum. Eur. J. Biochem.* 268, 1404-9.

32. Kawazu, S., Komaki, K., Tsuji, N., Kawai, S., Ikenoue, N., Hatabu, T., Ishikawa, H., Matsumoto, Y., Himeno, K. & Kano, S. (2001). Molecular characterization of a 2-Cys peroxiredoxin from the human malaria parasite *Plasmodium falciparum. Mol. Biochem. Parasitol.* 116, 73-9.

33. Krnajski, Z., Walter, R. D. & Muller, S. (2001). Isolation and functional analysis of two thioredoxin peroxidases (peroxiredoxins) from *Plasmodium falciparum. Mol. Biochem. Parasitol.* 113, 303-8.

34. Akerman, S. E. & Müller, S. (2003). 2-Cys peroxiredoxin PfTrx-Px1 is involved in the antioxidant defence of *Plasmodium falciparum. Mol. Biochem. Parasitol.* 130, 75-81.

35. Kawazu, S., Tsuji, N., Hatabu, T., Kawai, S., Matsumoto, Y. & Kano, S. (2000). Molecular cloning and characterization of a peroxiredoxin from the human malaria parasite *Plasmodium falciparum. Mol. Biochem. Parasitol.* 109, 165-9.

36. Sztajer, H., Gamain, B., Aumann, K. D., Slomianny, C., Becker, K., Brigelius-Flohé, R. & Flohé, L. (2001). The putative glutathione peroxidase gene of *Plasmodium falciparum* codes for a thioredoxin peroxidase. *J. Biol. Chem.* 276, 7397-403.

37. Komaki-Yasuda, K., Kawazu, S. & Kano, S. (2003). Disruption of the *Plasmodium falciparum* 2-Cys peroxiredoxin gene renders parasites hypersensitive to reactive oxygen and nitrogen species. *FEBS Lett.* 547, 140-4.

38. Krnajski, Z., Gilberger, T. W., Walter, R. D., Cowman, A. F. & Müller, S. (2002). Thioredoxin reductase is essential for the survival of *Plasmodium falciparum* erythrocytic stages. *J. Biol. Chem.* 277, 25970-5.

39. Declercq, J. P., Evrard, C., Clippe, A., Stricht, D. V., Bernard, A. & Knoops, B. (2001). Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å resolution. *J. Mol. Biol.* 311, 751-9.

40. Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* 5, 802-810.

41. Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure* 4, 897-904.

42. Evrard, C., Capron, A., Marchand, C., Clippe, A., Wattiez, R., Soumillion, P., Knoops, B. & Declercq, J. P. (2004). Crystal structure of a dimeric oxidized form of human peroxiredoxin 5. *J. Mol. Biol.* 337, 1079-90.

43. Rouhier, N., Gelhaye, E. & Jacquot, J. P. (2002). Glutaredoxin-dependent peroxiredoxin from poplar: protein-protein interaction and catalytic mechanism. *J. Biol. Chem.* 277, 13609-14.

44. Karplus, P. A. & Faerman, C. (1994). Ordered water in macromolecular structure. *Current Opinion in Structural Biology* 4, 770-776.

45. Daopin, S., Davies, D. R., Schlungegger, M. P. & Grutter, M. G. (1994). Comparison of two crystal structures of TGF-b2: the accuracy of refined protein structures. *Acta Crystallogr. D Biol. Crystallogr.* 50, 85-92.

46. Malin, R., Zielenkiewicz, P. & Saenger, W. (1991). Structurally conserved water molecules in ribonuclease T1. *J. Biol. Chem.* 266, 4848-52.

47. Karplus, P. A. & Schulz, G. E. (1987). Refined structure of glutathione reductase at 1.54 Å resolution. *J. Mol. Biol.* 195, 701-29.

48. Karplus, P. A. & Schulz, G. E. (1989). Substrate binding and catalysis by glutathione reductase as derived from refined enzyme: substrate crystal structures at 2 Å resolution. *J. Mol. Biol.* 210, 163-80.

49. Greenblatt, H. M., Ryan, C. A. & James, M. N. (1989). Structure of the complex of Streptomyces griseus proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. *J. Mol. Biol.* 205, 201-28.

50. Schröder, E., Littlechild, J. A., Lebedev, A. A., Errington, N., Vagin, A. A. & Isupov, M. N. (2000). Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure Fold. Des.* 8, 605-15.

51. Keefe, L. J., Sondek, J., Shortle, D. & Lattman, E. E. (1993). The alpha aneurism: a structural motif revealed in an insertion mutant of staphylococcal nuclease. *Proc. Natl. Acad. Sci. U S A* 90, 3275-9.

52. Trivelli, X., Krimm, I., Ebel, C., Verdoucq, L., Prouzet-Mauleon, V., Chartier, Y., Tsan, P., Lauquin, G., Meyer, Y. & Lancelin, J. M. (2003). Characterization of the yeast peroxiredoxin Ahp1 in its reduced active and overoxidized inactive forms using NMR. *Biochemistry* 42, 14139-49.

53. Nordhoff, A., Tziatzios, C., van den Broek, J. A., Schott, M. K., Kalbitzer, H. R., Becker, K., Schubert, D. & Schirmer, R. H. (1997). Denaturation and

reactivation of dimeric human glutathione reductase--an assay for folding inhibitors. *Eur. J. Biochem.* 245, 273-82.

54. Sarma, G. N., Savvides, S. N., Becker, K., Schirmer, M., Schirmer, R. H. & Karplus, P. A. (2003). Glutathione reductase of the malarial parasite *Plasmodium falciparum*: crystal structure and inhibitor development. *J. Mol. Biol.* 328, 893-907.

55. Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269-75.

56. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. J. & Sweet, R. M., eds.), Vol. 276, pp. 307-26.

57. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54 (Pt 5), 905-21.

58. Rice, L. M. & Brunger, A. T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* 19, 277-90.

59. Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci. U S A* 94, 5018-23.

60. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47 (Pt 2), 110-9.

61. Kleywegt, G. J. & Jones, T. A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure* 4, 1395-400.

62. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-637.

63. Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B. & Tainer, J. A. (1999). Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J. Mol. Biol.* 288, 427-39.

64. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* 50, 178-185.

65. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* 55 (Pt 4), 938-40.

66. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* 277, 505-524.

67. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystall.* 24, 946-950.

A1.6 **Footnotes**

Abbreviations used: Prx, peroxiredoxin; *Pf*AOP, *Plasmodium falciparum* antioxidant protein; ρ_{rms}, root-mean-square electron density of map, often reported as σ.

Appendix 2

Glutathione Reductase of the Malarial Parasite *Plasmodium falciparum*: Crystal Structure and Inhibitor Development

Ganapathy N. Sarma, Savvas N. Savvides, Katja Becker, Markus Schirmer, R. Heiner Schirmer and P. Andrew Karplus

Published in *Journal of Molecular Biology* (2003) **328**, 893-907

A2.1 Summary

The malarial parasite *Plasmodium falciparum* is known to be sensitive to oxidative stress, and thus the antioxidant enzyme glutathione reductase (GR; NADPH + GSSG + H^+ $\leftrightarrows$ $NADP^+$ + 2 GSH) has become an attractive drug target for antimalarial drug development. Here we report the 2.6 Å resolution crystal structure of *P. falciparum* GR. The homodimeric flavoenzyme is compared to the related human GR with focus on structural aspects relevant for drug design. The most pronounced differences between the two enzymes concern the shape and electrostatics of a large (450 $Å^3$) cavity at the dimer interface. This cavity binds numerous non-competitive inhibitors and is a target for selective drug design. A 34-residue insertion specific for the GRs of malarial parasites shows no density, implying that it is disordered. The precise location of this insertion along the sequence allows us to explain the deleterious effects of a mutant in this region and suggests new functional studies. To complement the structural comparisons, we report the relative susceptibility of human and plasmodial GRs to a series of tricyclic inhibitors as well as to peptides designed to interfere with protein folding and dimerization. Enzyme-kinetic studies on GRs from chloroquine-resistant and chloroquine-sensitive parasite strains were performed and indicate that the structure reported here represents GR of *P. falciparum* strains in general and thus is a highly relevant target for drug development.

A2.2 Introduction

Tropical malaria represents an increasing threat to human health and welfare[1,2]. The disease is caused by the multiplication of the protozoan parasite *Plasmodium falciparum* in human erythrocytes [3]. The emerging resistance of *Plasmodium* to chloroquine and other antimalarials underlines the need for the development of new chemotherapeutic agents with different modes of action [4-6].

During the erythrocytic stages of its life cycle, the parasite is exposed to oxidative stress produced by activated macrophages of the host and also by toxic heme and other decomposition products of hemoglobin. We are pursuing enhanced oxidative stress as an attractive avenue for drug development, because a number of lines of

evidence suggest that this can effectively inhibit parasite growth (reviewed in Schirmer *et. al.*[7]). Indeed the toxic effects of chloroquine and other quinoline antimalarials that inhibit β-hematin formation [8-12] may be partly due to increased levels of O_2-activating heme, and peroxide antimalarials such as artemisinin are believed to be activated by heme resulting in the formation of alkylating free radicals[13].

The glutathione system plays a key role in the defense of malarial parasites against oxidative stress and in the development of drug resistance. This is supported by the fact that chloroquine resistant parasites exhibit increased glutathione concentrations [14-16]. The tripeptide GSH detoxifies reactive oxygen species both directly and via glutathione peroxidase and glutathione *S*-transferase catalyzed reactions. High levels of reduced glutathione are maintained by *de novo* synthesis from the constituent amino acids [17] and by the enzyme glutathione reductase (GR), a ubiquitous FAD-containing protein. GR is a homodimer that catalyzes the reduction of glutathione disulfide to GSH using NADPH as the source of reducing equivalents [18-20]:

$$GSSG + NADPH + H^+ \leftrightarrows NADP^+ + 2\ GSH$$

P. falciparum GR (*Pf*GR) has been characterized biochemically and kinetically both in its native authentic form isolated from malarial parasites [21] and in its recombinant form [22-25]. As systems depending on adequate GR activity, glutaredoxin and glutathione *S*-transferase have been described in *P. falciparum* [18, 26]. As an intraerythrocytic parasite, *P. falciparum* depends on an intact host cell milieu. Evidence for this is a lack of either erythrocyte FAD, the prosthetic group of GR, or of its substrate NADPH – as observed in glucose-6-phosphate dehydrogenase deficiency – leads to significant protection against malaria [27-33]. Red blood cells rendered severely GR-deficient by the drug carmustine were able to fulfil their physiological functions – although for a shortened lifetime – but they did not serve as host cells for *P. falciparum* [34-36]. Thus both *Pf*GR and hGR are potential targets for the development of antimalarial drugs[23, 35, 37-39]. We have recently shown that *Pf*GR may be the target for methylene blue, an antimalarial which efficiently inhibits *Pf*GR and not hGR at therapeutically used concentrations [23]. Indeed, GR is the only protein that has been identified as a target of methylene blue in the parasite. Furthermore, on the basis of a

GR inhibitor a novel strategy for overcoming chloroquine resistance in malarial parasites has recently been developed [37, 40]. Quinoline-based alcohols with known antimalarial activity were combined with a GR inhibitor, represented by a menadione derivative, via a metabolically labile ester bond to give double-headed prodrugs. The quinoline moiety served also for directing the prodrug into the parasite. The activity of the compounds was proven in cell culture and in a mouse model [37]. Also, the drug combination of methylene blue-chloroquine (BlueCQ) is currently being tested in Nouna, Burkina Faso, as a drug against malaria complicated by methemoglobinemic anemia [41] (Coulibaly, B. and Schirmer, R. H, unpublished data).

Human GR has been very well studied in terms of enzymology and crystal structure [42, 43] (Figure A2.1). The binding of substrates, substrate analogs and inhibitors including different physiologic nitric oxide containing compounds [44-46] have been elucidated in detail [20, 42, 47]. Steps have already been taken to investigate the inhibitory properties of various drugs on hGR [48-50]. Among these inhibitors are isoalloxazine [50, 51], menadione [48] and xanthene [49] derivatives. All of these compounds bind in a cavity that is located at the dimer interface (Figure A2.1), and inhibit hGR non-competitively, the K_i values being in the lower micromolar range. Indirect evidence suggests that this cavity is also the binding site of methylene blue in *Pf*GR[22,24,52]. Methylene blue and isoalloxazine derivatives are also considered as constituents of double-headed drugs; the other head is designed to bind at the GSSG site, and the two heads are connected by a linker that fits in the channel connecting the intersubunit cavity with the GSSG site [46] (Figure A2.1). In general, phenothiazines and isoalloxazine derivatives are of interest as lead compounds for antimalarial drugs because they show low toxicity and are inexpensive.

There are three major features distinguishing *Pf*GR from hGR that are thought to be of relevance for selective inhibitor design [22, 25]. The first is an insertion of 34 residues within the central domain (residues 314-347) of *Pf*GR that is known to be highly antigenic [22], but has an unknown effect on function. The second regards the amino acid residues lining the wall of the cavity at the dimer interface, where only 9 out of 21 residues are conserved in *Pf*GR. The third feature is the pair of helices (H11/H11′) at the core of the dimer interface [43] (Figure A2.1).

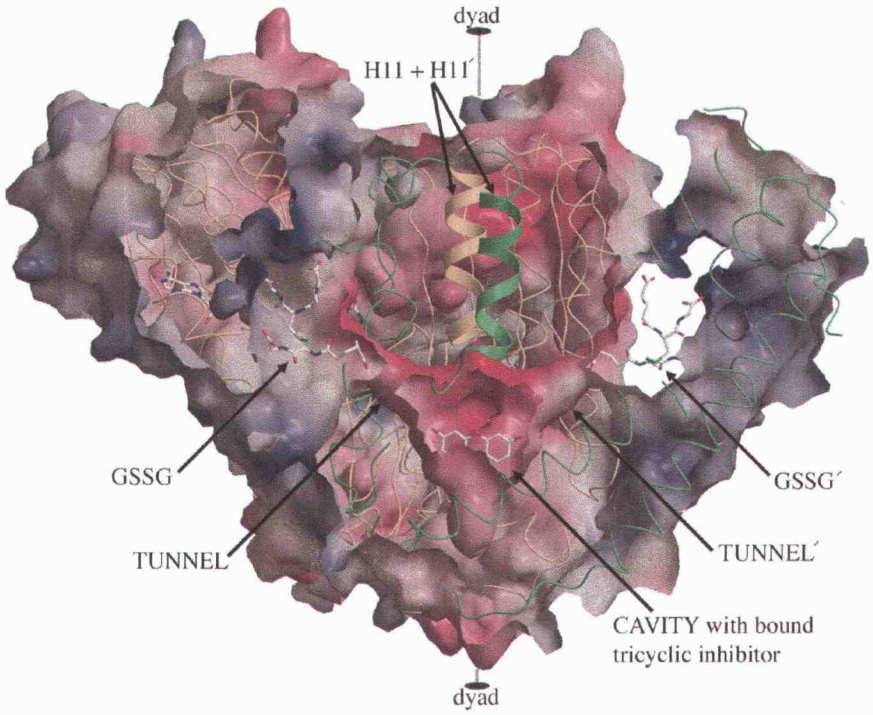


Figure A2.1 **The dimer cavity and helices of hGR.** The solvent accessible surface of hGR is shown with the front portion clipped away to reveal the cavity at the dimer interface. The surface is colored by electrostatic potential from dark blue (+4.65 kT) to dark red (-5.33 kT). A frame of reference is provided by a trace of the Cα backbone (yellow and green for each monomer). The helices at the dimer interface whose interaction is targeted by dimerization inhibitors are highlighted by representation as ribbons. Stick models (atom coloring) of the substrate GSSG [48] and the non-competitive inhibitor 6-hydroxy-3-oxo-3H-xanthene-9-propionic acid (XAN) [49] are shown to indicate the GSSG and the non-competitive inhibitor binding positions. Note the tunnels connecting the cavity to the two GSSG binding sites.

These helices are regarded as a dimerization and folding center of GR [53-55] and it has been shown that synthetic peptides can bind to these helices to interfere with the dimerization of hGR[38,53,54].

A valuable stepping stone for the further development of rationally designed drugs against *Pf*GR would be the three-dimensional structure of the enzyme. In this paper, we describe the three-dimensional structure of *P. falciparum* glutathione reductase and analyze those features that have been noted as potential targets for the design of selective *Pf*GR inhibitors. In addition we describe novel inhibitor design work related to *Pf*GR as a drug target.

A2.3 Results and Discussion

A2.3.1 Structure solution and model validation

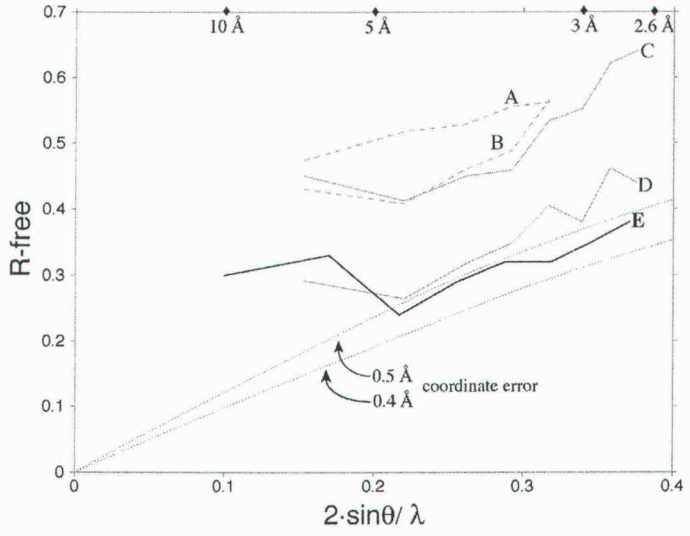
*Pf*GR crystallizes with a monomer in the asymmetric unit, and with the obligatory homodimer of *Pf*GR generated by a crystallographic 2-fold axis. The three-dimensional structure was solved by molecular replacement using hGR as the search model and refinement has led to the final model, at a nominal resolution of 2.6 Å, with an R of 25.1% and R_{free} of 30.2% and reasonable geometry. Based on a Luzzati analysis [56, 57], we estimate the coordinate accuracy of well-ordered parts of the protein to be near 0.5 Å (Figure A2.2a).

The best fit to the diffraction data was obtained using the translation, libration, screw-rotation (TLS) refinement [58] as implemented in REFMAC [59]. This refinement dropped both the R and R_{free} by over 1%, but led to a reported average B-factor of near 80 $Å^2$ (from CNS refinement the average B is 65 $Å^2$). This value is much higher than is often associated with 2.6 Å resolution structures, leading to some concern about whether the refined structure is a truly independent valid structure. Figure A2.2 presents four lines of evidence that the *Pf*GR structure presented here is indeed a reliable medium resolution structure. The first line of evidence is that the R_{free} improves in all resolution ranges during refinement and ends up at reasonable values (Figure A2.2a). Even in the highest resolution bin, the R_{free} is below 40% indicating that the data at 2.6 Å do contain real information that is accounted for by the model.

Figure A2.2 **Validation of the *Pf*GR structure.** (a) Variation in R-free as a function of resolution during refinement. R-factors calculated against the 3.0 Å and 2.6 Å data are represented by dashed and solid lines respectively. The final refined model is represented by the bold line. Curves are shown for models after rigid body refinement (A), at the stage where refinement switched from the 3.0 Å (B) to the 2.6 Å (C) data, at the end of the CNS refinement (D), and after TLS refinement (E). The dotted lines are Luzzati curves for 0.4 Å and 0.5 Å coordinate error. The R_{free} curve for the final refined model (E) closely approximates the Luzzati line for 0.5 Å coordinate error. (b) Quality of unbiased $2F_O$-F_C electron density. An annealed omit map (done according to the protocol in Hodel *et. al.* [82]) for residues Tyr424 to Glu432 (dark grey) is shown along with the final model for the omitted residues (dark grey) and adjacent residues that had not been omitted (light grey). The contour level is 1.3 ρ_{rms}. All *Pf*GR residues are labeled except Ser422 which is mostly hidden. This is not the best ordered part of the structure but is shown because it is an insertion in *Pf*GR relative to hGR (Figure A2.3) and so cannot be influenced by search model bias. The distinct hGR chain path is shown as a Cα trace with the end residues labeled in parentheses. (c) Ramachandran plot of the main-chain torsion angles of the final model. 93% of the non-glycine residues (plusses) lie in the core region (grey shaded area) and 7% (asterisks) lie in the non-core region. Glycine residues are represented by open squares. The significant scatter around the core regions is appropriate for a structure at this resolution, as we have refrained from adjusting the model for the sole purpose of making the Ramachandran plot look better. (d) Survey of the average B-factors of 675 protein structures determined at between 2.5 and 3.0 Å resolution and deposited in the Protein Data Bank between January 2000 and October 2002. Open bars are for the 467 structures noted as being determined using synchrotron radiation and the solid bars are for the remaining 208 structures determined by non-synchrotron sources. The averages of the distributions are 51 $Å^2$ and 34 $Å^2$ respectively.

Figure A2.2 (continued)

(a)



(b)

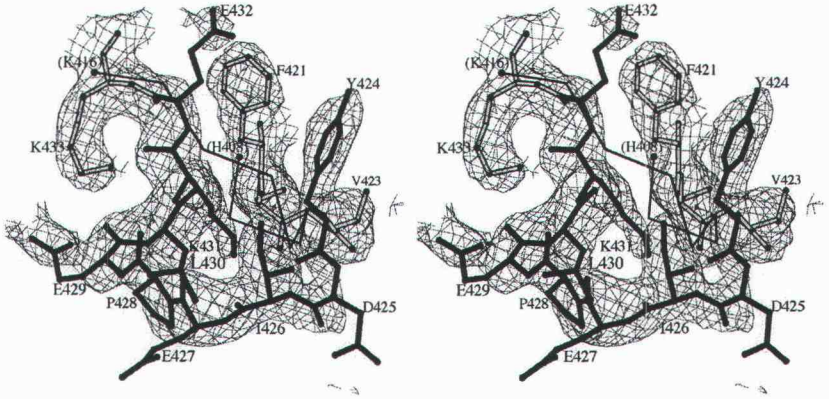
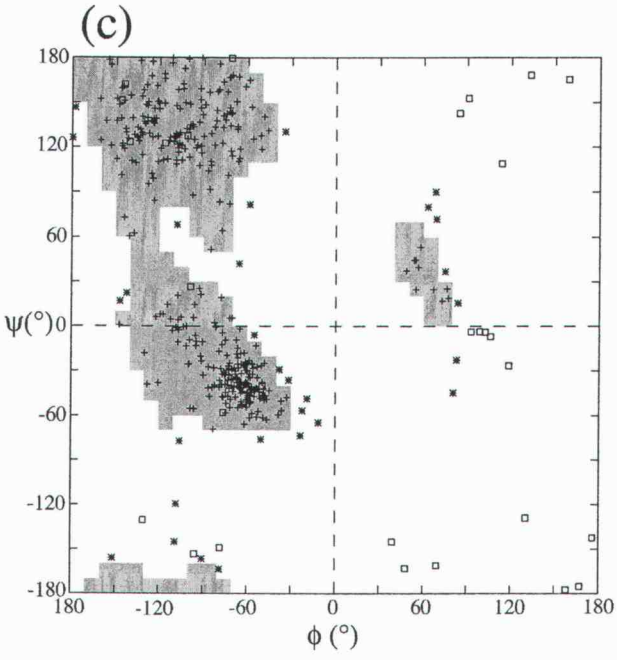
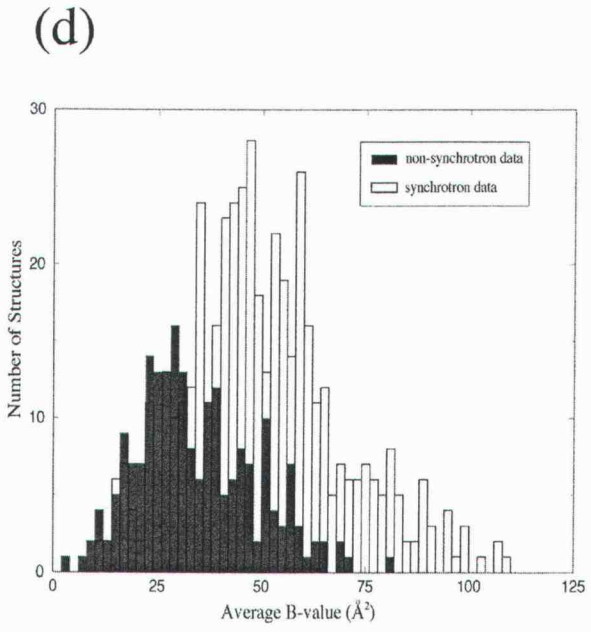


Figure A2.2 (continued)





A second line of evidence is the quality of the electron density seen in an annealed omit map for a region of structure that is unique to *Pf*GR (Figure A2.2b). While we know of no generally accepted method for quantifying the effective resolution of an electron density map, the density shows some definition of carbonyl bumps and, for ordered side chains (Tyr423, Ile426 and Lys431), it shows a level of detail appropriate for a map calculated in the 2.6 – 3.0 Å resolution range. A third line of evidence for the validity of the structure is a well-behaved Ramachandran plot (Figure A2.2c), which with about 7% of residues in non-core regions (as defined in Kleywegt and Jones [60]), is typical for structures in the 2.5 – 3.0 Å resolution range. Finally, as seen in Figure A2.2d, a survey of known structures shows that an average B-factor near 65 to 80 $Å^2$ is actually not a gross outlier for modern structures determined with synchrotron radiation at this level of resolution. The wide spread observed for average B-factors and the clear difference between the average B-factors of structures determined using synchrotron data versus laboratory data can both be understood by realizing that the average atomic spatial disorder (modeled by the B-factor) is an underlying feature of a given crystal form, whereas the reported resolution limit of a data set is highly dependent on arbitrary operator choices such as what resolution data are targeted, and on many experimental variables such as crystal size, beam intensity, detector sensitivity, and exposure time. Bluntly put, crystals diffracting to 2.5 – 3.0 Å using a synchrotron will tend to be of lower quality (i.e. less ordered) than those diffracting to 2.5 – 3.0 Å using a laboratory source.

A2.3.2 **Overall structure**

*Pf*GR shares 45% sequence identity with both hGR and *Ec*GR, and the overall architectures are very similar, including the conservation of two *cis*-peptide bonds which occur before Pro388 and Pro485. Structural comparisons between monomers of *Pf*GR, hGR and *Ec*GR yield rmsds of 1.7 Å (430 Cα atoms) for *Pf*GR with hGR, 1.6Å (425 Cα atoms) for *Pf*GR with *Ec*GR and 1.1 Å (446 Cα atoms) for hGR with *Ec*GR. Structure-based sequence alignment (Figure A2.3) and overlay (Figure A2.4) of *Pf*GR and hGR show that the structures are highly similar with large differences only occurring near insertions and deletions, and that all insertions and deletions are at the

surface of the proteins. The structure-based positioning of the insertions and deletions differs our previously published alignment [22, 25] and replaces it. None of these small insertions/deletions appears to have an important functional role in catalysis.

With the structure of *Pf*GR in hand, we can now compare it with the structure of hGR to discover unique features of *Pf*GR that could be targeted for inhibitor design. In the following four sections the four proposed regions of interest are described. These are three of the *Pf*GR insertions, the cavity at the dimer interface, the helices H11 and H11' that are at the core of the dimer interface and, of course, the active site.

A2.3.3 ***Pf*GR insertions relative to hGR**

There are three insertions in *Pf*GR with respect to hGR that have been proposed to be of importance [24]. As seen in the structure-based sequence alignment (Figure A2.3), these are residues 123-134 in the FAD binding domain, residues 314-347 in the central domain, and residues 496-499 at the C-terminus. While this structure reveals where the insertions exist, most of the residues in the three insertions do not have significant density and they have not been modeled. This implies that they are disordered relative to the main part of the protein, and provides no further indication for whether they have any influence on the enzyme function.

The second insertion (314-347) merits further comment because of its size, because database searches indicate that it is unique to all *Plasmodia* GRs, and because its role has been studied by deletion mutagenesis [25]. The deletion mutant of *Pf*GR lacked residues 317 to 349 and was found to form a dimeric FAD-free apoenzyme [25]. It was inferred that the insertion is important for FAD-binding. The structure allows us to more fully interpret the result. It is clear from the crystal structure that the residues through Val315 and residues Tyr348 and beyond are at least partly buried making favorable interactions to stabilize the folded protein. It also shows that residues Asp311 and Leu352, just before and just after the insertion, make intimate interactions with FAD, so that any disruption of their positions could be expected to disrupt FAD binding. Finally it shows that atom 316C and atom 350N, which must be covalently connected in the deletion mutant, are ca. 12 Å apart and cannot approach each other without seriously disturbing the integrity of the FAD binding domain.

Figure A2.3 **Structure-based sequence alignment of *Pf*GR with hGR.** The sequence of *Pf*GR is aligned with that of hGR based on a structural comparison. Dots identify every tenth residue with every 20^{th} residue numbered. Black lines delineate residues involved in secondary structures in each sequence. α-helices (gray cylinders), β-strands (gray arrows) and 3_{10}-helices (open cylinders) are indicated. The sequences are closely spaced for residues that are structurally equivalent (see methods) and they are farther apart for residues that are not spatially equivalent due to insertions or conformational differences. Identical residues are shaded. Residues lining the cavity at the dimer interface are marked with asterisks. Residues not modeled due to lack of density are denoted by lower case letters. In some cases, minor differences in DSSP-defined secondary structures between *Pf*GR and hGR are not shown, because by visual inspection the conformations were not significantly different; these regions were assigned based on the higher resolution hGR structure.

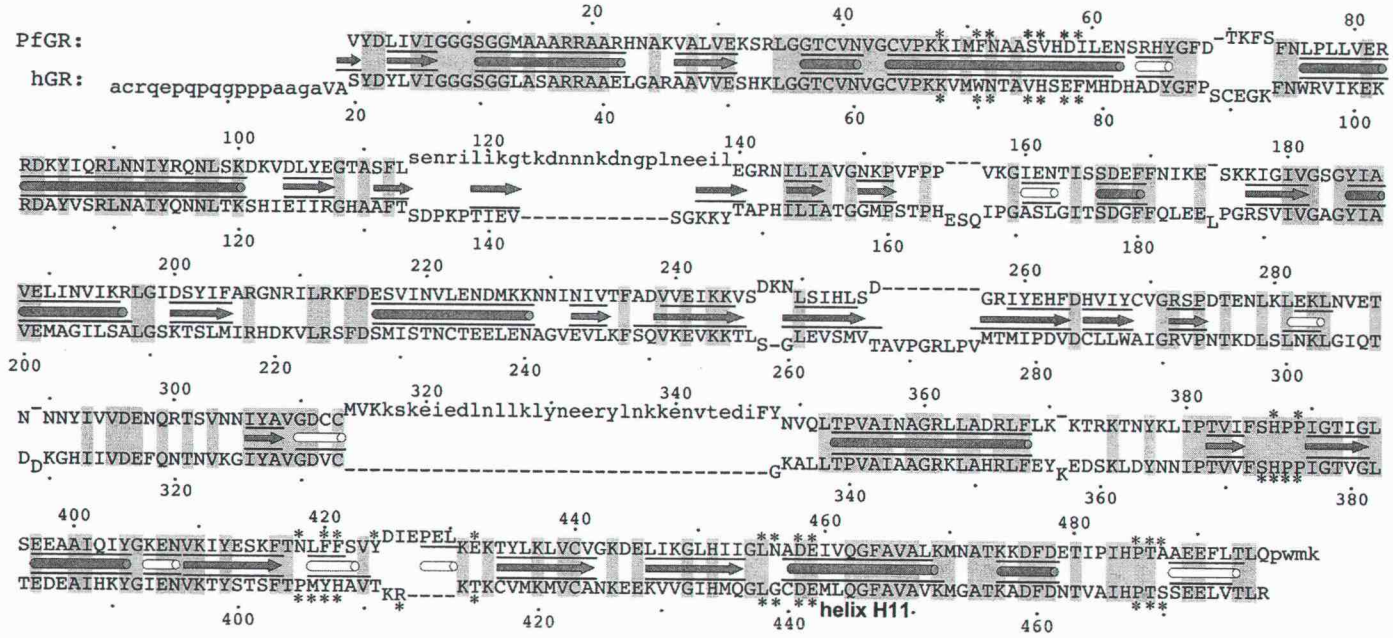


Figure A2.3

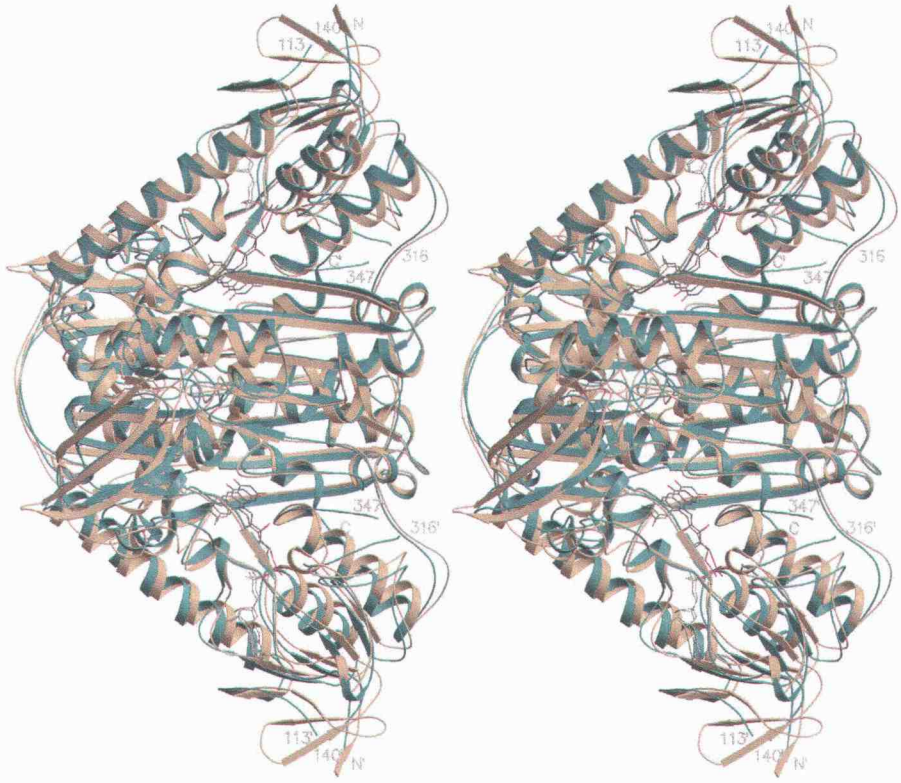


Figure A2.4 . **Comparison of *Pf*GR and hGR structures.** Stereo view of the superimposed backbone ribbons of *Pf*GR (blue) and hGR (yellow). The breaks in the *Pf*GR backbone, occurring where residues were not modeled due to lack of density (see Figure A2.3), are labeled.

Thus, the loss of FAD binding by this mutant is due to disruption of residues that are not part of the insertion and does not imply that the insertion itself is important for FAD binding. On the other hand, the lack of density in itself does not prove that the insertion is unimportant for FAD binding; this can only be tested by the creation of a deletion mutant that does not perturb the integrity of the FAD binding domain. Inspecting the structure, we see that Val315 and Tyr348 are opposite each other in an anti-parallel β-sheet. Thus, we suggest this insertion could be changed to a simple β-hairpin by replacing all residues between Lys316 and Phe347 with a Gly-Gly linker.

A2.3.4 **Cavity at the dimer interface**

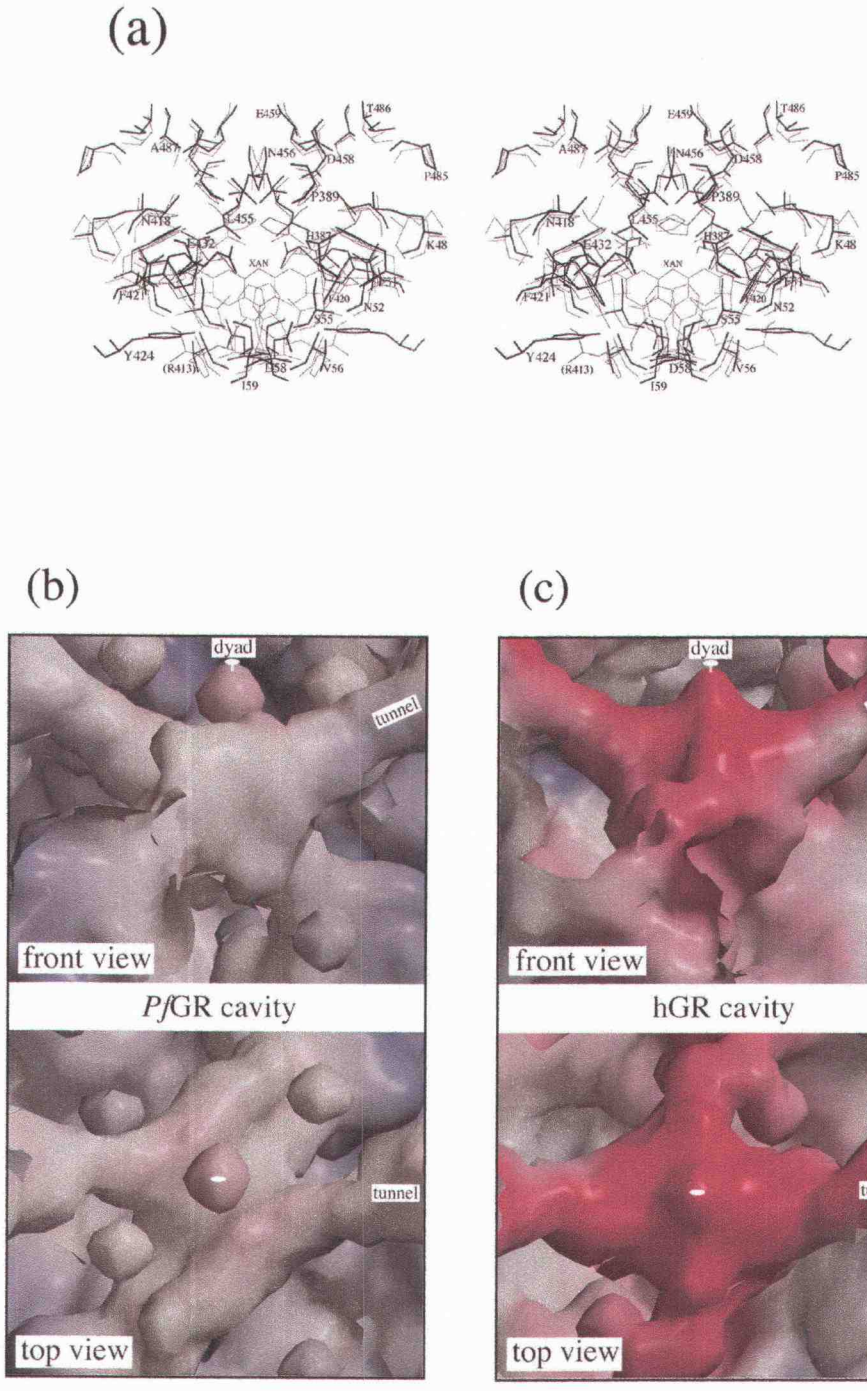
A comparison of the *Pf*GR and hGR cavities is shown in Figure A2.5. The *Pf*GR cavity is lined by 21 residues and has a volume of ca. 450 $Å^3$. The corresponding values for hGR are 24 residues and ca. 460 $Å^3$. Despite the similar volumes of the *Pf*GR and hGR cavities, their shapes and chemical characteristics are quite different (Figures 5b and 5c). Most notably, the upper part of the *Pf*GR cavity is much smaller compared to the hGR cavity. The increased boxiness of the lower part of the cavity in *Pf*GR compared to the cavity in hGR compensates for this. The charges on the cavity walls are also very different, with the hGR cavity being highly negatively charged, and the *Pf*GR cavity being rather neutral. Especially noteworthy for inhibitor design, is that residue Phe78, which with its symmetry mate Phe78′ directly 'sandwiches' the inhibitors in hGR, is missing (see Figure A2.5a). The major differences in cavity shape and charge imply that binding modes of *Pf*GR inhibitors cannot be reliably inferred from the hGR complexes.

A2.3.5 **Helices at the interface**

Another proposed target for inhibitor design against *Pf*GR is the pair of parallel running helices at the core of the dimerization interface [54, 55], namely residues 456-471 and 456′-471′ (Figure A2.1). Six of these 16 residues differ between *Pf*GR and hGR (Figure A2.3) but the *Pf*GR structure reveals that the distorted conformation of these helices is identical in the two proteins. In fact, all six residues that are different point away from the dimer interface, so that we would not expect these sequence differences to lead to selectivity in inhibitor action.

Figure A2.5 **Comparison of the cavities at the dimer interface of *Pf*GR and hGR.** (a) Stereo view of the residues that line the cavities. *Pf*GR residues (thick lines) have been numbered, with labels on only one of the symmetry mates for clarity. For hGR (thin lines; pdb entry 3GRS) only Arg413 is labeled with the number in parentheses. A bound tricyclic inhibitor (XAN, taken from pdb entry 1XAN) is shown for reference. (b) Closeup views of the molecular surface of the *Pf*GR cavity. The surfaces are colored as in Figure A2.1. The views are perpendicular to the dimer axis (as in Figure A2.1) and along the dimer axis (looking down from the top of Figure A2.1). (c) the corresponding views of the hGR cavity. Key residue replacements leading to the differences can be seen in panel (a). The upper part of the cavity becomes smaller in *Pf*GR because Gly439 (hGR) is replaced by Asn456 (*Pf*GR) and a different side chain rotamer is adopted by Leu455 (*Pf*GR) compared to Leu438 (hGR). In contrast, the lower part of the cavity is much larger in *Pf*GR, residues with smaller side chains replacing ones with larger side chains: Ser55, Val56, Asp58, and Ile59 (*Pf*GR) replace Val74, His75, Glu77, and Phe78 (hGR), respectively. Similarly the main replacements that appear to alter the electrostatic properties are Asp58, Asn62, His65 and Arg196 (*Pf*GR) with Glu77, Asp81, Asp84 and Ala208 (hGR). The change from Asp58 (*Pf*GR) to Glu77 (hGR) makes a difference in the electrostatic property of the cavity by virtue of the fact that Glu77 points towards the cavity whereas Asp58 points away from it.

Figure A2.5


(a)
E459
T486
A487
N456
D458
P485
P389
N418
L455
E432
H387
K48
XAN
F421
Y420
N52
S55
Y424
(R413)
D58
V56
I59
(b)
dyad
tunnel
front view
PfGR cavity
tunnel
top view
(c)
dyad
tunnel
front view
hGR cavity
tunnel
top view

A2.3.6 **Active site**

The active site of GR includes the redox active Cys39/Cys44 pair and the residues involved in FAD, NADPH, and GSSG binding. All of these have been structurally well characterized for hGR [43, 47, 61]. In these regions all but 7 residues are identical between hGR and *Pf*GR: 18 of 19 residues at the GSSG binding site, 9 of 10 residues in the FAD binding site, and 14 of 19 residues at the NADPH binding site. At the GSSG binding site, the only difference is Leu419′ in *Pf*GR replacing Met406′ in hGR. The amide nitrogen of this residue makes contact with GSSG and on visual inspection no difference in the positions of the main chain can be seen that would suggest a difference in binding of the substrate. At the FAD binding site in *Pf*GR Lys32 replaces Ser51 of hGR. The interaction, however, involves the main chain nitrogen atom and no difference is seen in the conformation of the main chain which makes a difference in FAD-binding unlikely. One notable change at the NADPH binding site is Gly207 in *Pf*GR replacing His219 in hGR. His219 is peripherally involved in the binding of the 2′-phosphate group of NADPH; it does not make any direct hydrogen bonds but it is close enough for electrostatic attraction. The change from His219 to a Gly207 could lead to weaker binding of NADPH in *Pf*GR and might explain why the preference for NADPH over NADH is less in *Pf*GR than in hGR [21]. The other changes in the NADPH binding site do not appear to be important.

A2.3.7 **Non-crystallographic studies on inhibitor development**

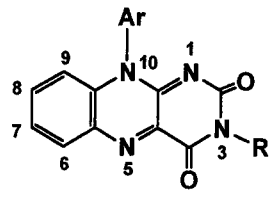
A2.3.7.1 Tricyclic heteroaromatic compounds as non-competitive GR inhibitors.

The crystallographic results contribute directly to the design and development of tricyclic compounds as antimalarials. For instance both methylene blue and arilloxazines form dimers in physiological solutions with dissociation constants of 170 μM [62] and 150 μM [63] respectively. Spectroscopic and enzyme inhibition studies could not eliminate the possibility that the dimers are inhibitory species, but the structure rules them out. It shows that the shape of the cavity at the intersubunit interface cannot accommodate a dimer, consisting of two stacked and rotated monomers that would require a box-like space of 13 Å × 5 Å × 8 Å[64].

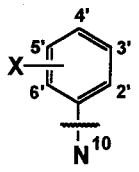
As a step towards understanding structure-activity relationship in the absence of the structure of a complex (no suitable crystals of a *Pf*GR-methylene blue or *Pf*GR-arilloxazine complex have yet been grown), we have determined the K_i values of a series of arilloxazine derivatives. In the presence of saturating substrate concentrations, these compounds were found to act as purely non-competitive inhibitors of both hGR and *Pf*GR (Table A2.1).

One striking trend is that the arilloxazines generally (all but compound 2e) bind more tightly to hGR than to *Pf*GR. The most obvious structural explanation for this difference is the presence of Phe78/Phe78′ in hGR (see above and Figure A2.5a), and mutagenesis experiments are underway to test this hypothesis. At a more detailed level, isoalloxazines with bulky substituents at position 10 bind more readily to hGR than to *Pf*GR (Table A2.1). Even naphthyl and anthryl derivatives (compounds 19 and 20) are good inhibitors of the human enzyme. In soluble arilloxazines the presence of a trifluoromethyl group in position 3' and 5' (compound 4c in Table A2.1) appears to be most unfavorable. Also, comparing the results for compounds 6 and 6a indicates that a positive charge in the side chain is not favorable for binding in the cavity: Finally, we note that 8-azido and 8-fluoro derivatives of compound 2a inhibit as effectively as the parent compound (see 12 and 14, Table A2.1). These derivatives may be helpful for chemical labeling experiments [50] that could guide efforts to predict the binding modes of the rigid arilloxazines by molecular modeling [17]. In photoaffinity labelling experiments using compound 12, more than 50% of the colored label ended up the in the tryptic peptides p(49-84) and p(410-430) of *Pf*GR, with none of the other peptides containing more than 5% of the label (Kirsch, P. and Schirmer, R. H, unpublished data). Both labelled peptides contain cavity-lining residues (Figures 3 and 5).

Table A2.1. **Arilloxazines as non-competitive inhibitors of *Pf*GR and hGR**[a].



if Ar = phenyl :



Isoalloxazine Derivative	Substituents at N3	Substituent at N10	K_i for *Pf*GR (μM)	K_i for hGR (μM)
2e	H	pentafluorophenyl	1.9	2.5
2a	H	4'-chlorophenyl	5.4	0.6
4b	(ethoxycarbonyl)methyl	3',5'-dichlorophenyl	13	6.6
4c	(ethoxycarbonyl)methyl	3'-trifluoromethylphenyl	> 1000	9.3
5b	Carboxymethyl	3',5'-dichlorophenyl	14	7.2
6	Methyl	methylpyridinium-4'-yl	> 1000	>200
6a	Methyl-	4'-pyridyl	15	2.5
12	3-methyl (and 8-azido)[b]	4'-chlorophenyl	6.2	1.4
14	3–methyl (and 8-fluoro)[b]	4'-chlorophenyl	5.9	0.4
19	Methyl	1'-naphthyl	11	0.5
20	Methyl	1'-anthryl	>200	2.5

[a] The inhibition parameters determined in previous studies for human GR [50,51,63] have a competitive component. This was eliminated here by using saturating substrate concentrations. If the substituent in position 10 is a methyl group or a dimethylaminopropyl group the K_i-values are above 50 μM,[83] which indicates that an aromatic substituent in position 10 is favorable. In a previous crystallographic study [50] compound 5b has been shown to be bound in the intersubunit cavity of hGR.
[b] compounds 12 and 14 also have a substituent at C8.

A2.3.7.2 Helix 11-based dimerization inhibitors.

According to our crystallographic results, peptides which are known to interact with helix 439-454 of hGR are expected to interact equally well with helix 456-471 of *Pf*GR. When testing this hypothesis for P11, the peptide analog of hGR helix 11 (residues 436-459) originally shown to inhibit hGR folding [54], the results were ambiguous because rather high peptide concentrations and long incubation times are necessary. We therefore synthesized and studied a series of shorter peptides.

These peptides had C-terminal Gly454 (equivalent to Asn471 in *Pf*GR) in common and were 11 to 19 amino acids long. Only peptide $K16_{hGR}$ (see methods for nomenclature) strongly inhibited dimerization and folding of both hGR ($EC_{50\%}$ = 120 nM) and *Pf*GR ($EC_{50\%}$ = 80 nM) at submicromolar concentrations. Longer (K17 to K19) or shorter (K11 to K15) peptides were found to be much weaker inhibitors: $K17_{hGR}$ with $EC_{50\%}$ = 16 μM for both enzymes was the second best, and the other K peptides gave <15% folding inhibition when tested at a concentration of 16 μM. Human thioredoxin reductase whose helix 11 sequence differs in 8 positions from hGR was found to be much less affected by $K16_{hGR}$ than the glutathione reductases, the $EC_{50\%}$ value being 12 μM.

These results suggest that peptide $K16_{hGR}$ could serve as a lead structure for probing dimerization and folding of FAD-dependent disulfide reductases. Specificity for an individual enzyme of this superfamily, for instance for *Pf*GR, might be obtainable by systematic mutations of the soluble and structurally stable $K16_{hGR}$ sequence. In addition, the structure of $K16_{hGR}$ recommends itself for developing peptidomimetic agents as folding inhibitors or as dimer-dissociating agents. Such inhibitors would be expected to act in catalytic amounts since the resulting protein monomers are readily degraded, thereby releasing the dissociating agent [53-55]. The concept of denaturing inhibitors is supported by studies on the helix 11 mutants – G446E/ F447P of hGR [54] and G463E/F464P of *Pf*GR [53]. These mutants that form full-length protein subunits but without detectable folding in any domain are extremely sensitive to proteolytic degradation.

A2.3.7.3 Intraspecies structural variation of P. falciparum *GR.*

The hypothesis that structure and kinetic characteristics are not rigidly conserved among GRs from different *P. falciparum* strains and that the drug chloroquine may be the selecting force has questioned whether *Pf*GR is a general target of antimalarial drugs. This hypothesis is based on the observation that the catalytic competence k_{cat}/K_M of recombinant GR of the chloroquine-resistant strain K1 strain is about 5 fold higher than the k_{cat}/K_M value of recombinant His-tagged GR from the chloroquine-sensitive strain 3D7. It should be noted that inhibition analysis using carmustine, methylene blue, and an arilloxazine did not reveal significant differences between the two *Pf*GR variants [25].

Apart from the presence of the initiation methionine and the C terminal His tag, the primary structures differ in 2 positions [25]. Residues 280–282, which form part of an isolated surface loop in the central domain, are 280KLE282 in the K1 enzyme and 280NLG282 in 3D7 GR, and residue Arg334 in K1-GR is changed to Thr334 in 3D7-GR. The latter exchange is located in one of the loops that were too flexible to be modeled (Figures 1 and 3). Thus structural links between alterations of these residues and the activity of GR were not established in the crystallographic study and the difference in the k_{cat}/K_M value between recombinant *Pf*K1-GR and *Pf*3D7-GR remain unexplained. The similarity of the K_M values when comparing the GRs in extracts of the two *P. falciparum* strains (Table A2.2) suggests that the high K_M value of 211 μM observed for recombinant *Pf*3D7GR by Gilberger *et. al.* [25] might be due to effects of the C-terminal His-tag and the initiation methionine residue on GSSG-binding and reduction. Val1 in K1-GR is already an integral part of the FAD domain so that its modification by a methionyl residue might have an effect on catalysis.

The structure-based conclusions were substantiated by further biological studies. From 8 different *P. falciparum* strains – 4 chloroquine-sensitive and 4 chloroquine-resistant strains – synchronized trophozoites were produced in cell culture, separated from host cells and extracted. In these parasite extracts we determined specific GR activity and K_M values for GSSG (Table A2.2). No obvious differences could be detected between chloroquine-sensitive and resistant strains, with the mean GR activity being 25 mU/mg in sensitive and 24 mU/mg in resistant strains; the mean K_M values were 66

μM and 77 μM, respectively. Thus, the attractive hypothesis that chloroquine-resistant parasites have a GR with lower K_M values for GSSG – which would strengthen their GSH-dependent and chloroquine-susceptible heme detoxification system [10, 25] was not substantiated. Rather, GR activity levels are likely to adapt to cellular demands by increasing the enzyme concentration [21, 65].

It is most likely that the structure determined here represents GR of *P. falciparum* strains in general and that it is suitable as a basis for future development of inhibitors as potential antimalarial drugs. Such compounds will also be valuable for studying the role of the enzyme in parasite metabolism, drug action and multiple drug resistance [10, 37].

A2.3.8 **Outlook**

The known GR inhibitors – with the notable exceptions of methylene blue [23] and compound 2e (Table A2.1) – are more effective as inhibitors of human GR than of *Pf*GR. Following the dogma of current antimicrobial drug development – *hit the pathogen but spare the functions of the human host* – future inhibitor development must exploit the structural differences between human GR and *Pf*GR to obtain more *Pf*GR-specific inhibitors. However, at the same time, compelling arguments suggest that we should not neglect human erythrocyte GR as a potential target of antimalarial drugs [35, 39]. These arguments include clinical and epidemiological observations as well as pharmacological studies showing that low cytosolic levels of GR activity – which are usually restricted to erythrocytes – do not significantly interfere with erythrocyte function but they offer protection from malarial parasites [7, 20, 27, 35]. Another argument in favor of choosing a host cell enzyme as a target is the *a priori* prevention of drug resistance [7, 39]. In conclusion, on a short term basis the structure-based development of *Pf*GR-specific inhibitors will have high priority; on a long term basis, however, compounds inactivating both *Pf*GR and mammalian GRs [19, 55] might become equally important for the chemotherapy of malaria.

Table A2.2 **Characteristics of *Pf*GR in various strains of *P. falciparum* [a].**

	Chloroquine-sensitive				Chloroquine-resistant			
Strain	HB3	S106	D10	3D7	Dd2	K1	7G8	FCR3
GR activity [mU/mg]	35	23	9	33	22	47	13	13
K_M for GSSG [μM]	76	73	45	71	74	88	80	65

[a] Glutathione reductase activity and K_M values for GSSG in crude extracts (see methods). GR activity is given in mU/mg total protein in the parasite extract. All values represent means of three determinations.

A2.4 Materials and Methods

A2.4.1 Crystallization and structure determination

Crystals of purified recombinant *Pf*GR expressed in *E. coli* [23] were grown using the hanging-drop vapor diffusion method. *Pf*GR was initially dialyzed against 2 mM EDTA, pH 7.0 for three hours, and the dialysate was clarified by centrifugation. 5 µl of the resulting protein solution (10 mg/ml) was mixed with 5 µl reservoir solution (14% (w/v) PEG-MME 550, 70 mM NaCl, 70 mM bicine, pH 8.7-9.0 to form drops that were equilibrated over 0.8 ml of reservoir solution at 20 °C. Crystals of *Pf*GR grew to their final size (typically 0.05 x 0.1 x 0.15 mm^3) in 1 week and were subsequently harvested and stored in reservoir solution. Native data sets were collected to 3.0 Å resolution (Nat1) at beam-line 14BM-C of the Advanced Photon Source (Argonne National Laboratory, USA) and to 2.6 Å resolution (Nat2) at beam-line 5.0.2 of the Advanced Light Source (Lawrence-Berkeley National Laboratory, USA), respectively. Data were collected at –170 °C from crystals that were frozen in a stream of cold nitrogen after a 2 minute incubation in 14% PEG-MME 550, 25% (v/v) glycerol, 0.1 M bicine, pH 9.0 (Table A2.1). *Pf*GR crystals belong to the tetragonal space group $P4_122$ (a=90.64 Å, b=90.64 Å, c=123.32 Å), with one half-dimer of *Pf*GR in the asymmetric unit and 44% solvent.

The structure of *Pf*GR was determined at 3.0 Å resolution (Nat1, Table A2.3) by molecular replacement using the Patterson correlation strategy in the program CNS v1.0 [66]. The structure of hGR (PDB code 3GRS) [43] was used as the search model after mutating non-conserved residues to alanine, and 10% of the data were set aside for cross validation. The molecular replacement solution was unambiguous, having a correlation coefficient of 0.52, compared to 0.30 for the second best solution. R and R_{free} for this model were 0.490 and 0.498 respectively. Rigid-body refinement and conjugate-gradient minimization dropped the R-factors to R=0.405 and R_{free}=0.468 (Figure A2.2a, curve A), and 2Fo-Fc and Fo-Fc electron density maps validated the solution as they revealed electron density for some missing side-chains and other structural differences between the search model and *Pf*GR. Manual rebuilding was carried out using the program O [67]. The correct side chains and visible insertions and

deletions were built into the model. A round of simulated annealing, positional refinement and B-factor refinement using the 3.0 Å resolution data yielded R=0.366 and R_{free}=0.442 (Figure A2.2a, curve B). The higher quality native ALS data extending to 2.6 Å resolution (Table A2.3) were then used, retaining the R_{free} test set. Several rounds of simulated annealing, positional and B-factor refinement together with manual rebuilding yielded a model with R=0.273 and R_{free}=0.315 (Figure A2.2a, curve D). During these rounds, omit maps and trial refinements were used to determine the best interpretation of regions with weak electron density and exactly where to cut-off the model for the three disordered chain segments. Water molecules were placed with the aid of the water-pick utility in CNS [66] and watpeak utility in the CCP4 package [68] with the following criteria: (1) a $\geq 3\rho_{rms}$ peak in the Fo-Fc map and a $\geq 1\rho_{rms}$ peak in the 2Fo-Fc map and (2) a distance of $\geq$ 2.6 Å and $\leq$ 3.5 Å to nearby hydrogen bond donor or acceptor. Model geometry was monitored using the program PROCHECK [69].

Because CNS refinements indicated significant crystal anisotropy ($B_{11} = B_{22}$ = -9 Å^2, B_{33} = 18 Å^2 from overall anisotropic scaling), we tried to account for this better by using the program REFMAC [59] which incorporates translation, librations, and screw-rotation (TLS) refinement [58]. *Pf*GR was treated as four independent rigid bodies: two parts of the FAD-binding domain, the NADPH binding domain and the interface domain (Table A2.4). Several rounds of manual rebuilding and TLS with positional refinement led to the final model; the R_{free} value dropped significantly to 0.302 and the R factor dropped to 0.251 (Figure A2.2a, curve E). The refinement was terminated when there was not sufficient improvement in the density to guide further model changes. For database submission, the program TLSANL [70] was used to generate a pdb file with the anisotropic mobility parameters and with an "equivalent" isotropic B-factor for each atom. The average B-factor was a surprisingly high 80 Å^2. To assess the validity of these B-factors, CNS was used to carry out a test refinement of this model. The overall anisotropic scaling B-factors were the same as before (see above) and the average atomic B-factors dropped to 65 Å^2 while the R_{free} reported by CNS rose slightly to 0.321. Combining the overall anisotropic scale factors with the average B leads to the insight that the effective average B-factors for the structure in

the x, y, and z directions are 74 $Å^2$ (=65+9), 74 $Å^2$, and 47 $Å^2$ (=65-18), respectively. Thus, the observed average B of the structure is complex and somewhat refinement program dependent, but this does not impact the biological significance of the structure determination. In the final model, regions that were not modeled due to lack of density included residues 114-139, 317-346 and 496-499 and the side chains for residues Glu^{140}, Lys^{316} and Phe^{347} (near the edges of the disordered segments). Data collection and refinement statistics are given in Tables 3 and 4.

A2.4.2 **Structural comparisons and analyses**

Secondary structure assignments were made using DSSP [71]. Structure-based sequence alignments were done using SEQUOIA [72] with a distance cutoff of 3.0 Å. The 1.54 Å resolution structure of hGR (PDB code 3GRS) [43] and the 1.86 Å resolution structure of *E. coli* (PDB code 1GER) [73] were used for structural comparisons. Quantitative cavity characterization were done with the program VOIDOO [74] using a probe radius of 1.2 Å. The probe radius was chosen so that the residues identified as lining the cavity matched those selected by visual inspection. Because the interface cavity is not technically a 'cavity' but a channel with narrow openings on both sides (see Figure A2.1) a few waters blocking these openings were included in the calculations; all other water molecules were removed. Electrostatic surface calculations were done with the program GRASP [75] using a probe radius of 1.4 Å. The van der Waals radii for the surface calculations were values reported as r_b in Table A2.3 of ref. 76. Structure figures were generated using MOLSCRIPT [77], Raster3D [78], GRASP [75] and the OPLOT routine in O [67].

A2.4.3 **Enzymes**

Recombinant hGR [55], recombinant *Pf*GR (strain K1) [22] and authentic human thioredoxin reductase [79] were prepared as previously described. The protein concentrations were determined based on the FAD absorption using a $e_{465\ nm}$ = 11.3 $mM^{-1}\ cm^{-1}$ for the oxidized enzymes.

Table A2.3 **Data collection and refinement statistics** [a]

	Nat2	*Nat1*
Data		
Resolution limits (Å)	30 - 2.6 (2.78 – 2.60)	30 - 3.0 (3.05 – 3.00)
Unique observations	16434	10678
Multiplicity	7.5	8.9
Average I/σ	13.5 (4.1) [b]	18.9 (2.8)
R_{meas} [c] (%)	9.4 (38.8)	9.1 (41.8)
Completeness (%)	98.7 (99.9)	98.3 (90.0)
Refinement		
Reflections with F>0 σ	15604	
Protein + FAD atoms	3511	
Solvent atoms	43	
R_{cryst} (%), R_{free} (%) [d]	25.1 (28.5), 30.2 (37.0)	
Stereochemical ideality		
R.m.s.d bond lengths (Å)	0.019 Å	
R.m.s.d bond angles (degrees)	2.1°	

[a] Numbers in parentheses correspond to values in the highest resolution shell.
[b] Reflecting the diffraction anisotropy, in the highest resolution bin the I/ σ is near 2 for high h and high k reflections and is near 6 for high l reflections.
[c] $R_{meas} = \sum_h (n_h/n_h\text{-}1)^{1/2} \sum_i |I_h - I_{h,i}| / \sum_h \sum_i I_{h,i}$, multiplicity-weighted R_{merge} [84].
[d] $R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$, $R_{free} = R_{cryst}$ for 10% of reflections against which the model was not refined [85].

Table A2.4 **TLS motion parameters** [a].

Domain [b]		TLS tensor values							
FAD part 1	T	0.2256	0.1509	0.4933	-0.0355	-0.0111	-0.0278		
(1-148)	L	3.5001	4.0511	1.0533	2.3815	-0.7895	-1.0087		
	S	0.0139	0.0970	0.0206	0.7606	0.2729	0.1818	-0.4943	0.1075
NADPH	T	0.2412	0.3106	0.3095	-0.2006	0.0100	0.0549		
(149-275)	L	1.6757	3.5382	4.4883	0.1591	1.1554	0.5247		
	S	0.6719	-0.4960	0.4447	0.3272	0.1859	-0.5678	-0.6401	0.1104
FAD part 2	T	0.3106	0.5410	0.6999	-0.2862	0.0487	0.0180		
(276-378)	L	3.4438	7.0780	1.5295	1.8778	-2.5387	-0.3815		
	S	0.3943	-0.1274	0.6415	0.2894	-0.5811	-0.6268	-0.3282	1.1447
Interface	T	0.0240	0.3143	0.3830	0.0525	0.0533	-0.1129		
(379-495)	L	3.2584	0.9094	3.0948	-0.0928	-0.7980	-1.1240		
	S	0.0949	-0.1015	0.3833	-0.3890	-0.3283	-0.1659	0.1251	0.6999

[a] TLS values correspond to T-matrix at T_{11}, T_{22}, T_{33}, T_{12}, T_{13}, T_{23}, L-matrix at L_{11}, L_{22}, L_{33}, L_{12}, L_{13}, L_{23} and S-matrix at S_{22}-S_{11}, S_{11}-S_{33}, S_{12}, S_{13}, S_{21}, S_{23}, S_{31}, S_{32}. The units of T and S are $Å^2$ and the units of L are degrees^2. The details of TLS refinement are described in Winn *et. al.* [58].

[b] Domain boundaries listed are not based on precise calculations, but simple inspection. FAD part 2 has also been called the central domain [85].

A2.4.4 Inhibition studies with 10-arylisoalloxazines (arilloxazines)

These compounds – for which the designation arilloxazines is introduced here - were synthesized as described before [63]. They are designated (*2a to 19*) according to Schönleben-Janas *et. al.* [50] and not to the differing system of Kirsch *et. al.* [63]. As the arilloxazines, above all the 8-azido derivative, are sensitive to light, they were prepared fresh as 1 mM stock solutions in DMSO every day and kept in the dark prior to conducting the assays

The inhibition assays were carried out in GR assay buffer (50 mM potassium phosphate, 200 mM KCl, 1 mM EDTA, pH 6.9 at 25 °C). The assay mixture contained 100 nM NADPH and 5-20 mU *Pf*GR or hGR and various concentrations of inhibitor. After checking for non-specific NADPH oxidation for 2 min, 1 µmol of GSSG was added to start the enzymatic reaction. The decrease in NADPH concentration as a measure of enzyme activity was monitored spectrophotometrically ($\varepsilon_{340\,nm}$=6.22 mM^{-1} cm^{-1}). Residual enzyme activity in the presence of inhibitor was determined relative to controls containing DMSO [50, 52].

A2.4.5 Peptide analogues of helix 11

A set of nine helix 11 analog peptides, 11 to 19 residues long and ending at the position of Gly454 of hGR, were synthesized by Dr. R. Pipkorn, German Cancer Research Center, Heidelberg. They are designated K (for Kontakt) peptides. For instance, $K19_{hGR}$ has the sequence QGLGCDEMLQGFAVAVKMG (see Figure A2.3). The 500 µM stock solutions were prepared in DMSO as most peptides were not sufficiently soluble in water; only for $K16_{hGR}$ and $K19_{hGR}$ was it also possible to make up 500 µM aqueous stock solutions. For testing the peptides as dimerization and folding inhibitors of hGR and *Pf*GR we used the previously described denaturation/renaturation assay [38, 53, 54].

A2.4.6 Analysis of GR activity in *P. falciparum* strains

Intraerythrocytic stages of 8 different *P. falciparum* strains – 4 chloroquine-sensitive strains and 4 chloroquine-resistant strains – were cultured *in vitro* according to Trager and Jensen [80], and synchronised to ring stages by the sorbitol method [81]. After 24 h, parasites in the trophozoite stage were isolated by suspending the red cells

in a 20-fold volume of buffer containing 7 mM K_2HPO_4, 1 mM NaH_2PO_4, 11 mM $NaHCO_3$, 58 mM KCl, 56 mM NaCl, 1 mM $MgCl_2$, 14 mM glucose, 0.02% saponin, pH 7.3, for 10 minutes at 37°C. The pelleted parasites were washed three times, then taken up in 150 µl buffer, and disrupted by three freeze-thaw cycles. After centrifugation, the supernatant was used for further analyses. Protein in the extracts was determined using the BioRad (München, Germany) protein dye assay with bovine serum albumin as a standard.

GR activity was measured with a standard protocol [55]. In this case, 0.5 ml assay mixture containing 100 µM NADPH and 50 µl parasite extract in assay buffer, pH 6.9, was equilibrated at 25°C. The reaction was then started by adding GSSG at 1 mM. For determining K_M values, the concentration of GSSG was systematically varied from 10 µM to 1000 µM.

A2.4.7 **Coordinates**

The coordinates and structure factors have been deposited in the Protein Data Bank (http://www.rcsb.org) as entry 1ONF.

A2.4.8 **Acknowledgements**

The excellent technical assistance of Petra Harwaldt and Irene König is gratefully acknowledged. We would also like to thank Dr. Rick Faber for carrying out the analysis presented in Figure A2.2d and helpful discussions. This work was supported by Deutsche Forschungsgemeinschaft (Sonderforschungsbereiche 535 and 544 "Control of Tropical Infectious Diseases" and grant BE 1540/7-1 (to K. B) and Schi 102/8-1 (to R. H. S), and by National Science Foundation grant MCB-9982727 (to P. A. K).

A2.5 **References**

1. Greenwood, B. & Mutabingwa, T. (2002). Malaria in 2002. *Nature* 415, 670-672.

2. Sachs, J. & Malaney, P. (2002). The economic and social burden of malaria. *Nature* 415, 680-685.

3. Miller, L. H., Baruch, D. I., Marsh, K. & Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature* 415, 673-679.

4. Olliaro, P. (2001). Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* 89, 207-219.

5. Hyde, J. E. (2002). Mechanisms of resistance of *Plasmodium falciparum* to antimalarial drugs. *Microbes Infect.* 4, 165-174.

6. Ridley, R. G. (2002). Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 415, 686-693.

7. Schirmer, R. H., Müller, J. G. & Krauth-Siegel, R. L. (1995). Disulfide-reductase inhibitors as chemotherapeutic agents: The design of drugs for typanosomiasis and malaria. *Angew. Chem. Int. Ed. Engl.* 34, 141-154.

8. Loria, P., Miller, S., Foley, M. & Tilley, L. (1999). Inhibition of the peroxidative degradation of haem as the basis of action of chloroquine and other quinoline antimalarials. *Biochem. J.* 339, 363-370.

9. Foley, M. & Tilley, L. (1998). Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents. *Pharmacol. Ther.* 79, 55-87.

10. Ginsburg, H., Famin, O., Zhang, J. & Krugliak, M. (1998). Inhibition of glutathione-dependent degradation of heme by chloroquine and amodiaquine as a possible basis for their antimalarial mode of action. *Biochem. Pharmacol.* 56, 1305-1313.

11. Egan, T. J., Ross, D. C. & Adams, P. A. (1994). Quinoline anti-malarial drugs inhibit spontaneous formation of beta-haematin (malaria pigment). *FEBS Lett.* 352, 54-57.

12. Zhang, J., Krugliak, M. & Ginsburg, H. (1999). The fate of ferriprotoporphyrin IX in malaria infected erythrocytes in conjunction with the mode of action of antimalarial drugs. *Mol. Biochem. Parasitol.* 99, 129-141.

13. Meshnick, S. R., Taylor, T. E. & Kamchonwongpaisan, S. (1996). Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol. Rev.* 60, 301-315.

14. Dubois, V. L., Platel, D. F., Pauly, G. & Tribouley-Duret, J. (1995). Plasmodium berghei: implication of intracellular glutathione and its related enzyme in chloroquine resistance in vivo. *Exp. Parasitol.* 81, 117-124.

15. Famin, O., Krugliak, M. & Ginsburg, H. (1999). Kinetics of inhibition of glutathione-mediated degradation of ferriprotoporphyrin IX by antimalarial drugs. *Biochem. Pharmacol.* 58, 59-68.

16. Platel, D. F., Mangou, F. & Tribouley-Duret, J. (1999). Role of glutathione in the detoxification of ferriprotoporphyrin IX in chloroquine-resistant *Plasmodium berghei*. *Mol. Biochem. Parasitol.* 98, 215-223.

17. Austin, S. E., Khan, M. O. & Douglas, K. T. (1999). Rational drug design using trypanothione reductase as a target for anti-trypanosomal and anti-leishmanial drug leads. *Drug Des. Discov.* 16, 5-23.

18. Rahlfs, S., Fischer, M. & Becker, K. (2001). *Plasmodium falciparum* possesses a classical glutaredoxin and a second, glutaredoxin-like protein with a PICOT homology domain. *J. Biol. Chem.* 276, 37133-37140.

19. Iozef, R., Becker, K., Böhme, C. C., Schirmer, R. H. & Werner, D. (2000). Assembly and functional expression of murine glutathione reductase cDNA: a sequence missing in expressed sequence tag libraries. *Biochim. Biophys. Acta* 1500, 137-141.

20. Schirmer, R. H., Bauer, H. & Becker, K. (2002). Glutathione reductase. In *Wiley Encyclopedia of Molecular Medicine* (Creighton, T. E., Ed.), pp. 1471-1475, John Wiley & Sons, Inc., New York.

21. Krauth-Siegel, R. L., Müller, J. G., Lottspeich, F. & Schirmer, R. H. (1996). Glutathione reductase and glutamate dehydrogenase of *Plasmodium falciparum*, the causative agent of tropical malaria. *Eur. J. Biochem.* 235, 345-350.

22. Färber, P. M., Becker, K., Müller , S., Schirmer, R. H. & Franklin, R. M. (1996). Molecular cloning and characterization of a putative glutathione reductase gene, the *Pf*GR2 gene, from *Plasmodium falciparum*. *Eur. J. Biochem.* 239, 655-661.

23. Färber, P. M., Arscott, L. D., Williams, C. H., Jr., Becker, K. & Schirmer, R. H. (1998). Recombinant *Plasmodium falciparum* glutathione reductase is inhibited by the antimalarial dye methylene blue. *FEBS Lett.* 422, 311-314.

24. Böhme, C. C., Arscott, L. D., Becker, K., Schirmer, R. H. & Williams, C. H., Jr. (2000). Kinetic characterization of glutathione reductase from the malarial parasite *Plasmodium falciparum*. Comparison with the human enzyme. *J. Biol. Chem.* 275, 37317-37323.

25. Gilberger, T. W., Schirmer, R. H., Walter, R. D. & Müller, S. (2000). Deletion of the parasite-specific insertions and mutation of the catalytic triad in glutathione reductase from chloroquine-sensitive *Plasmodium falciparum* 3D7. *Mol. Biochem. Parasitol.* 107, 169-179.

26. Harwaldt, P., Rahlfs, S. & Becker, K. (2002). Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*: characterization of a potential drug target. *Biol. Chem.* 383, 821-830.

27. Anderson, B. B., Giuberti, M., Perry, G. M., Salsini, G., Casadio, I. & Vullo, C. (1993). Low red blood cell glutathione reductase and pyridoxine phosphate oxidase activities not related to dietary riboflavin: selection by malaria? *Am. J. Clin. Nutr.* 57, 666-672.

28. Atamna, H., Pascarmona, G. & Ginsburg, H. (1994). Hexose-monophosphate shunt activity in intact *Plasmodium falciparum*-infected erythrocytes and in free parasites. *Mol. Biochem. Parasitol.* 67, 79-89.

29. Cappadoro, M., Giribaldi, G., O'Brien, E., Turrini, F., Mannu, F., Ulliers, D., Simula, G., Luzzatto, L. & Arese, P. (1998). Early phagocytosis of glucose-6-phosphate dehydrogenase (G6PD)-deficient erythrocytes parasitized by *Plasmodium falciparum* may explain malaria protection in G6PD deficiency. *Blood* 92, 2527-2534.

30. Thurnham, D. I., Oppenheimer, S. J. & Bull, R. (1983). Riboflavin status and malaria in infants in Papua New Guinea. *Trans. R. Soc. Trop. Med. Hyg.* 77, 423-424.

31. Ruwende, C. & Hill, A. (1998). Glucose-6-phosphate dehydrogenase deficiency and malaria. *J. Mol. Med.* 76, 581-588.

32. Beutler, E. & Luzzatto, L. (1999). Hemolytic anemia. *Semin. Hematol.* 36, 38-47.

33. Ginsburg, H. & Golenser, J. (1999). Redox metabolism in glucose-6-phosphate dehydrogenase deficient erythrocytes and its relation to antimalarial chemotherapy. *Parassitologia* 41, 309-311.

34. Schirmer, R. H., Krauth-Siegel, R. L. & Schulz, G. E. (1989). Glutathione. In *Coenzymes and Cofactors* (Dolphin, D., Poulsen, R. & Avramovic, O., Eds.), Part A, pp. 553-596, John Wiley & Sons, Inc., New York.

35. Zhang, Y. A., Hempelmann, E. & Schirmer, R. H. (1988). Glutathione reductase inhibitors as potential antimalarial drugs. Effects of nitrosoureas on *Plasmodium falciparum* in vitro. *Biochem. Pharmacol.* 37, 855-860.

36. Becker, K. & Schirmer, R. H. (1995). 1,3-Bis(2-chloroethyl)-1-nitrosourea as thiol-carbamoylating agent in biological systems. *Methods Enzymol.* 251, 173-188.

37. Davioud-Charvet, E., Delarue, S., Biot, C., Schwöbel, B., Böhme, C. C., Müssigbrodt, A., Maes, L., Sergheraert, C., Grellier, P., Schirmer, R. H. & Becker, K. (2001). A prodrug form of a *Plasmodium falciparum* glutathione reductase inhibitor conjugated with a 4-anilinoquinoline. *J. Med. Chem.* 44, 4268-4276.

38. Becker, K., Schirmer, M., Kanzok, S. & Schirmer, R. H. (1999). Flavins and flavoenzymes in diagnosis and therapy. *Methods Mol. Biol.* 131, 229-245.

39. Krauth-Siegel, R. L. & Coombs, G. H. (1999). Enzymes of parasite thiol metabolism as drug targets. *Parasitol. Today* 15, 404-409.

40. Ginsburg, H. (2002). A double-headed pro-drug that overcomes chloroquine resistance. *Trends Parasitol.* 18, 103.

41. Anstey, N. M., Hassanali, M. Y., Mlalasi, J., Manyenga, D. & Mwaikambo, E. D. (1996). Elevated levels of methaemoglobin in Tanzanian children with severe and uncomplicated malaria. *Trans. R. Soc. Trop. Med. Hyg.* 90, 147-151.

42. Pai, E. F., Karplus, P. A. & Schulz, G. E. (1988). Crystallographic analysis of the binding of NADPH, NADPH fragments, and NADPH analogues to glutathione reductase. *Biochemistry* 27, 4465-4474.

43. Karplus, P. A. & Schulz, G. E. (1987). Refined structure of glutathione reductase at 1.54 Å resolution. *J. Mol. Biol.* 195, 701-729.

44. Keese, M. A., Böse, M., Mülsch, A., Schirmer, R. H. & Becker, K. (1997). Dinitrosyl-dithiol-iron complexes, nitric oxide (NO) carriers in vivo, as potent inhibitors of human glutathione reductase and glutathione-S-transferase. *Biochem. Pharmacol.* 54, 1307-1313.

45. Becker, K., Savvides, S. N., Keese, M., Schirmer, R. H. & Karplus, P. A. (1998). Enzyme inactivation through sulfhydryl oxidation by physiologic NO-carriers. *Nat. Struct. Biol.* 5, 267-271.

46. Savvides, S. N., Scheiwein, M., Böhme, C. C., Arteel, G. E., Karplus, P. A., Becker, K. & Schirmer, R. H. (2002). Crystal structure of the antioxidant enzyme glutathione reductase inactivated by peroxynitrite. *J. Biol. Chem.* 277, 2779-2784.

47. Karplus, P. A. & Schulz, G. E. (1989). Substrate binding and catalysis by glutathione reductase as derived from refined enzyme:substrate crystal structures at 2 Å resolution. *J. Mol. Biol.* 210, 163-180.

48. Karplus, P. A., Pai, E. F. & Schulz, G. E. (1989). A crystallographic study of the glutathione binding site of glutathione reductase at 0.3-nm resolution. *Eur. J. Biochem.* 178, 693-703.

49. Savvides, S. N. & Karplus, P. A. (1996). Kinetics and crystallographic analysis of human glutathione reductase in complex with a xanthene inhibitor. *J. Biol. Chem.* 271, 8101-8107.

50. Schönleben-Janas, A., Kirsch, P., Mittl, P. R., Schirmer, R. H. & Krauth-Siegel, R. L. (1996). Inhibition of human glutathione reductase by 10-arylisoalloxazines: crystalline, kinetic, and electrochemical studies. *J. Med. Chem.* 39, 1549-1554.

51. Becker, K., Christopherson, R. I., Cowden, W. B., Hunt, N. H. & Schirmer, R. H. (1990). Flavin analogs with antimalarial activity as glutathione reductase inhibitors. *Biochem. Pharmacol.* 39, 59-65.

52. Luond, R. M., McKie, J. H., Douglas, K. T., Dascombe, M. J. & Vale, J. (1998). Inhibitors of glutathione reductase as potential antimalarial drugs. Kinetic cooperativity and effect of dimethyl sulphoxide on inhibition kinetics. *J. Enzyme Inhib.* 13, 327-345.

53. Schirmer, M. (2000). Ein Doppelhelix-Motiv als Zielstruktur des Drug Design: Die Interface-Domäne obligat dimerer Disulfidreduktasen. M. D. thesis, Heidelberg University.

54. Nordhoff, A., Tziatzios, C., van den Broek, J. A., Schott, M. K., Kalbitzer, H. R., Becker, K., Schubert, D. & Schirmer, R. H. (1997). Denaturation and reactivation of dimeric human glutathione reductase. An assay for folding inhibitors. *Eur. J. Biochem.* 245, 273-282.

55. Nordhoff, A., Bücheler, U. S., Werner, D. & Schirmer, R. H. (1993). Folding of the four domains and dimerization are impaired by the Gly446Glu exchange in human glutathione reductase. Implications for the design of antiparasitic drugs. *Biochemistry* 32, 4060-4066.

56. Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst* 5, 802-810.

57. Kleywegt, G. J. & Brunger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure* 4, 897-904.

58. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D. Biol. Crystallogr.* 57, 122-133.

59. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* 53, 240-255.

60. Kleywegt, G. J. & Jones, T. A. (1996). Phi/psi-chology: Ramachandran revisited. *Structure* 4, 1395-1400.

61. Janes, W. & Schulz, G. E. (1990). Role of the charged groups of glutathione disulfide in the catalysis of glutathione reductase: crystallographic and kinetic studies with synthetic analogues. *Biochemistry* 29, 4022-4030.

62. Vennerstrom, J. L., Makler, M. T., Angerhofer, C. K. & Williams, J. A. (1995). Antimalarial dyes revisited: xanthenes, azines, oxazines, and thiazines. *Antimicrob. Agents Chemother.* 39, 2671-2677.

63. Kirsch, P., Schönleben-Janas, A. & Schirmer, R. H. (1995). Synthesis and characterization of water-soluble and photolabile 10-arylisoalloxazines: Tools for studying the mechanism of flavin-type antimalarials. *Liebigs Ann.* 1995, 1275-1281.

64. Antonov, L., Gergov, G., Petrov, V., Kubista, M. & Nygren, J. (1999). UV-Vis spectroscopic and chemometric study on the aggregation of ionic dyes in water. *Talanta* 49, 99-106.

65. Bhatia, A. & Charet, P. (1984). Action of chloroquine on glutathione metabolism in erythrocytes parasitized by *Plasmodium berghei. Ann. Parasitol. Hum. Comp.* 59, 317-320.

66. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* 54, 905-921.

67. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A.* 47, 110-119.

68. Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* 50, 760-763.

69. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. App. Crystallogr.* 26, 283-291.

70. Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). TLSANL: TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *J. App. Crystallogr.* 26, 622-624.

71. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.

72. Bruns, C. M., Hubatsch, I., Ridderstrom, M., Mannervik, B. & Tainer, J. A. (1999). Human glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *J. Mol. Biol.* 288, 427-439.

73. Mittl, P. R. & Schulz, G. E. (1994). Structure of glutathione reductase from *Escherichia coli* at 1.86 Å resolution: comparison with the enzyme from human erythrocytes. *Protein Sci.* 3, 799-809.

74. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D. Biol. Crystallogr.* 50, 178-185.

75. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.

76. Bondi, A. (1964). van der Waals volumes and radii. *J. Phys. Chem.* 68, 441-451.

77. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.

78. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic Molecular Graphics. *Methods Enzymol.* 277, 505-524.

79. Gromer, S., Merkle, H., Schirmer, R. H. & Becker, K. (2002). Human placenta thioredoxin reductase: preparation and inhibitor studies. *Methods Enzymol.* 347, 382-394.

80. Trager, W. & Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science* 193, 673-675.

81. Lambros, C. & Vanderberg, J. P. (1979). Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.* 65, 418-420.

82. Hodel, A., Kim, S.-H. & Brunger, A. T. (1992). Model Bias in Macromolecular Crystal Structures. *Acta Crystallogr. A.* 48, 851-858.

83. Diederichs, K. & Karplus, P. A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269-275.

84. Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475.

85. Thieme, R., Pai, E. F., Schirmer, R. H. & Schulz, G. E. (1981). Three-dimensional structure of glutathione reductase at 2 Å resolution. *J. Mol. Biol.* 152, 763-782.

86. Becker, K. (1988). Glutathionreduktase und ihr Apoenzym: Beiträge zur Chemotherapie der Malaria und zur Diagnostik von FAD-Mangelzuständen. M. D. thesis, Heidelberg University.

A2.6 **Footnotes**

Abbreviations used: hGR, human glutathione reductase; *Pf*GR, *Plasmodium falciparum* glutathione reductase; *Ec*GR, *Escherichia coli* glutathione reductase; GSSG, oxidized glutathione; arilloxazines, 10-arylisoalloxazines; TLS, translation, libration and screw-rotation; ρ_{rms}, root-mean-square electron density of map, often reported as σ.