Structural Studies of Microbial Proteins - From *Escherichia coli* and Herpesviruses

Daniel Gurmu

ISBN 978-91-7155-995-1, pp. 1-64

Stockholms universitet

Structural Studies of Microbial Proteins

- From *Escherichia coli* and Herpesviruses

Doctoral Thesis

Daniel Gurmu

ISBN 978-91-7155-995-1, pp. 1-64

Printed in Sweden by Universitetsservice US-AB, Stockholm 2010
Distributor: Department of Biochemistry and Biophysics, Stockholm University

Tillägnas min Hustru och mina Döttrar

Abstract

Structure biology concerns the study of the molecular structures of biological macromolecules, such as proteins, and how these relate to the function. Protein structures are also of importance in structure-based drug design. In this thesis, the work has been carried out in two different projects. The first project concerns structural studies of proteins from the bacterium *Escherichia coli* and the second of proteins from five different herpesviruses.

The *E. coli* project resulted in the structural characterization of three proteins: CaiB, RibD, and YhaK. CaiB is a type-III CoA transferase involved in the metabolism of carnitine. Its molecular structure revealed a spectacular fold where two monomers were interlaced forming an interlocked dimer. RibD, a bi-functional enzyme, catalyzes two consecutive reactions during riboflavin biosynthesis. In an attempt to characterize the mechanism of action of the N-terminal reductase domain, the structure of RibD was also determined in two binary complexes with the oxidized cofactor, $NADP^+$, and with the substrate analogue ribose-5-phosphate. YhaK is a protein of unknown function normally found in low abundance in the cytosol of *E. coli* and was previously annotated to be a member of the Pirin family. However, some structural features seem to distinguish YhaK from these other Pirin proteins and we showed that YhaK might be regulated by reactive oxygen species.

The Herpesvirus project resulted in the structural determination of two proteins, the SOX protein and ORF60 from Kaposi's sarcoma associated herpesvirus (KSHV). SOX, a bi-functional shutoff and exonuclease protein, is involved in the maturation and packaging of the viral genome into the viral capsid and in the host shutoff of cellular proteins at the mRNA level. The SOX structure was also used for modeling DNA binding. The crystallization and preliminary structural studies of ORF60, the small R2 subunit of the ribonucleotide reductase (RNR) from KSHV is also discussed.

Table of Contents

List of Publications

I. Stenmark P, **Gurmu D**, Nordlund P.
Crystal structure of CaiB, a type-III CoA transferase in carnitine metabolism.
Biochemistry. 2004, 43(44):13996-4003

II. Stenmark P*, Moche M*, **Gurmu D**, Nordlund P.
The crystal structure of the bifunctional deaminase/reductase RibD of the riboflavin biosynthetic pathway in Escherichia coli*: implications for the reductive mechanism.*
J Mol Biol. 2007, 373(1):48-64

III. **Gurmu D***, Lu J*, Johnson KA, Nordlund P, Holmgren A, Erlandsen H.
The crystal structure of the protein YhaK from Escherichia coli *reveals a new subclass of redox sensitive enterobacterial bicupins.*
Proteins. 2009, 74(1):18-31

IV. Dahlroth SL*, **Gurmu D***, Haas J, Erlandsen H, Nordlund P.
Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus.
FEBS J. 2009, 276(22):6636-45

V. **Gurmu D,** Dahlroth SL, Haas J, Nordlund P, Erlandsen H.
Expression, purification, crystallization and preliminary X-ray analysis of ORF60, the small subunit (R2) of ribonucleotide reductase from Kaposi's sarcoma-associated herpesvirus (KSHV)
Manuscript.

* Shared first authorship

Abbreviations

HTP	High-throughput
NMR	Nuclear magnetic resonance
IMAC	Immobilized metal affinity chromatography
SEC	Size exclusion chromatography
CoA	Coenzyme A
FRC	Formyl-CoA transferase
FMN	Flavin mononucleotide
FAD	Flavin adenine dinucleotide
DAROPP	2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5′-phosphate
AROPP	5-amino-6-ribosylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate
ARIPP	5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate
RP5	Ribose 5-phosphate
GTP	Guanosine triphosphate
ATP	Adenosine triphosphate
GLP	Germin-like proteins
SCOP	Structural classification of proteins
Q23D	Quercetin 2,3-dioxygenase
Pfam	Protein family database
PDB	Protein data bank
PCD	Programmed cell death
Bcl-3	B-cell lymphoma 3-encoded protein
GSNO	Nitrosoglutathione
DTT	Dithiothreitol
TrxR	Thioredoxin reductase
Trx	Thioredoxin
GR	Glutathione reductase
GSH	Glutathione
Grx	Glutaredoxin
ROS	Reactive oxygen species
LTTR	LysR-type transcriptional cofactor
HTH	Helix-turn-helix
ORF	Open reading frame
HSV	Herpes simplex virus
KSHV	Kaposi’s sarcoma-associated herpesvirus
EBV	Epstein-Barr virus
CMV	Cytomegalovirus
mCMV	Murine cytomegalovirus
HHV	Human herpesvirus
KS	Kaposi’s sarcoma
PEL	Pleural effusion lymphoma
MCD	Multicentric Castleman’s disease
dNTP	Deoxyribonucleotides
AE	Alkaline exonuclease

VHS	Virion host shutoff
eIF4H	eukaryotic initiation factor H
SOX	Shutoff and Exonuclease
HLA	Human leukocyte antigen
RNR	Ribonucleotide reductase

Introduction

The role of proteins in living organisms is of fundamental importance since these macromolecules are involved in nearly all processes occuring in the cell. Proteins are the executors of various tasks in accordance to what is stipulated by the information encoded in the DNA of every cell (1). Based on structure, can proteins be broadly divided in three categories, globular, fibrous, and membrane proteins. Membrane proteins often serve as receptors or channels for compounds impermeable to the plasma membrane that separates the cell from its surroundings. Fibrous proteins are typically structural proteins, such as α-keratin, the major component of hair and nails. The cytoskeleton, which grants cells their shape and rigidity, is also composed of structural fibrous proteins. The third group of proteins, the globular proteins often play a catalytic or regulatory role. An important group of proteins is the *enzymes* that catalyze reactions that often are essential for cellular life. Typically, enzymes are highly specific and may only catalyze one or a few related chemical reactions. Upon catalysis, enzymes bind their substrate, the molecule they act on, in the *active site*, a region of the protein containing the catalytic residues needed for the reaction. Enzymes, like most other proteins, generally contain hundreds of amino acid residues. However, only a few of these bind the substrate, and even fewer participate directly in the catalysis. Besides substrate binding, proteins may also bind other molecules. Typically, the binding site of substrates as well as other molecules is located in a "pocket", a depression on the protein surface. The binding partner may be a small compound or another protein molecule. Often, proteins interact in carrying out their functions by the formation of protein complexes, and enzymatic activity may be regulated by such protein-protein interactions.

Proteins

Proteins are polypeptides composed of amino acids, molecules containing an amino group, a carboxyl group, both bonded to a central carbon atom (C_α) (1). Every amino acid contains a specific side chain, also known as an R-group, that distinguishes it from the others. These building blocks of proteins are attached to one another by the formation of peptide bonds between the carboxyl and amino groups of adjacent residues in the biosynthesis of proteins. In this process, known as translation, the growing chain of a polypeptide is synthesized from the amino end, the N-terminus, to the carboxy end, the C-terminus. Every protein has a distinctive amino acid sequence, which is defined by the nucleotide sequence of the corresponding gene.

In general, proteins are folded to form three-dimensional structures. The description of the structure and architechture of proteins often involves four different levels, primary, secondary, tertiary and quaternary structure (2). A protein's primary structure simply is its sequence of amino acid residues linked to one another by peptide bonds. Next, distinct regions of the sequence comprising local structures, such as α-helices and β-sheets, are described in the secondary structure of proteins. Such secondary structural elements are primarily stabilized by the formation of hydrogen bonds. The tertiary

structure of a protein concerns the spatial arrangement of the various secondary structure elements of the polypeptide chain. Typically, this is the level most commonly referred to when describing protein structure. Generally, the tertiary structure, or the fold of the protein, depicts the essential function of a certain protein. The formation of a hydrophobic core, in which hydrophobic amino acids are buried in the interior of the protein, is a major event in the stabilization of a protein fold. Other interactions that may contribute to the stabilization of the tertiary structure are salt bridges, hydrogen bonds and disulfide bonds. The description of the protein structure on a quaternary level concerns the formation of protein oligomers or complexes, in which two or more protein molecules, or subunits, bind to one another. The determination of a protein's tertiary structure and/or the quaternary structure of a protein complex may yield vital insights concerning its biological function. However, protein molecules are far from being rigid. Typically, proteins have the ability to switch between different structural states, conformations, as they carry out their functions. For instance, substrate binding in the active site of an enzyme often induces a conformational change of the protein, which may affect its activity. Structural biology concerns the study of the molecular structure of biological macromolecules, such as proteins and nucleic acids. In this thesis, emphasis will be on protein structure. The study of protein structure is important, since proteins, which are directly involved in most cellular functions, is dependent on their fold to carry out their functions.

The Structural Genomics Approach

The vast quantity of data generated from various sequencing projects has paved the way for the emerging fields of proteomics and structural genomics (3). In general, structural genomics initiatives aim at increasing the protein structure universe by generating three-dimensional models of the greater part of all protein families (3,4). The construction of a comprehensive and accurate map of the protein structure space is dependent on experimental data of structural representatives from various protein families being combined with computational data, such as structure predictions (3). Needless to say, such an accomplishment will make an immense contribution to all branches in the life sciences. In practice, such an effort will require the experimental determination of at least one representative structure for every protein family, defined at a level of 30-35% sequence identity (4). Proteins and other biological macromolecules are too small to be observed by light microscopy. As a consequence, structural determination of these biomolecules has been dependent on the development of other methods, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) and Electron Microscopy. Each method has its benefits and drawbacks. However, X-ray crystallography is by far the most commonly method used. It is also the method that has been used in the work described in this thesis, and thus serves as the foundation of this discussion. The process of generating high-quality models of protein structures involves several steps, such as: cloning of the coding sequence into an appropriate expression vector, expressing the protein in sufficient amounts and at purity levels required for structural studies, crystallization, X-ray data collection, as well as determination, refinement and validation of the experimental structure. In the

work that underlies this thesis, I have been involved in all these steps, with the exception of cloning. In the pursuit of going all the way from gene to a determined three-dimensional protein structure, there are numerous methods to choose between. Therefore, a "structural genomics approach" is often adopted, which takes advantage of high-throughput (HTP) methods at various stages to obtain results better, faster and in a systematic manner. Ideally, as each protein is different from the next, the optimal approach would be to design individual protocols for each target. However, this is often not feasible when working with many proteins. A major contribution from structural genomics initiatives is the information and experience that have been generated, especially for protein expression and purification, which has resulted in "consensus starting points" (5). Still, one has to consider that this outcome is predominantly based on studies of non-membrane cytosolic proteins and soluble protein domains. Although problems may arise at each and every step in this process, two major bottlenecks stand out from the rest, the production of soluble protein and protein crystallization.

Protein Production

Protein production includes protein expression as well as protein purification. As an expression host, the bacterium *Escherichia coli* has proved itself as a suitable candidate for the production of globular protein domains of both prokaryotic and eukaryotic origin (5). Apart from being one of the most studied model organisms, *E. coli* also evades some of the problems that often are associated with expression of a protein destined for crystallization, such as post-translational modifications. In addition, as a lot of effort has been invested in the improvement and optimization of this bacterium as an expression host, as much as 50% of full-length proteins from Eubacteria or Archaea may be expressed as soluble proteins in *E. coli*. However, for proteins of eukaryotic origin the corresponding number is 10%. The size of the protein also matters. If the molecular mass of the protein target exceeds 60 kDa, the likelihood to express the protein in a soluble form declines significantly. Failure to express proteins in a soluble form may be due to incorrect folding or to modification. Often, protein molecules in an insoluble form tend to aggregate to form inclusion bodies within the bacterium. External factors, such as induction time and temperature, may also have an influence on expression levels of soluble proteins.

The purification of a protein target often employs the usage of affinity tags located in-frame and in fusion with the protein of interest (5). Optimally, the tag will facilitate the initial capturing of the target protein without interfering with its biological activity. Typically, an N-terminal hexahistidine add-on tag may be combined with a clevage site for a sequence-specific protease allowing the tag to be removed. The choice of a hexahistidine tag offers several advantages, as it is moderately small and normally does not affect the solubility of the target proteins or other characteristics. Furthermore, using a hexahistidine tag permits the protein of interest to be purified using straightforward protocols by the means of immobilized metal affinity chromatography (IMAC) and has also been shown to play a neutral role in protein

crystallization. For the initial capture, IMAC is an advantageous method for several reasons. The method allows strong and specific binding of the target protein in combination with gentle elution conditions and it is also possible to fine-tune the selectivity of the method. As the levels of purity and concentration increases, many proteins may precipitate at physiological concentrations of salt. Thus, in order to maintain protein solubility and stability, IMAC may be carried out in buffer conditions with high ionic strength. In general, the capture of the target protein by IMAC is followed by an additional "polishing" step, typically carried out by size exclusion chromatography (SEC). SEC, which is also known as gel filtration chromatography, may be performed in a variety of buffer conditions, which may be useful in the preparation of a protein sample for crystallization trials. SEC also allows the oligomeric state of the protein sample to be determined, which may be of significance in protein characterization and also influence crystallization. A homogenous protein sample is characteristically attributed with a narrow and symmetric elution profile, while a broad and asymmetric profile indicates an inhomogenous and/or aggregated state. Generally, a protein sample that displays a high degree of monodispersion is an indication that it is less susceptible to protein aggregation, which may be a critical factor in protein crystallization (6).

Crystallization and Structure Determination

It has been shown that two thirds of all proteins submitted to crystallization trials will fail to generate protein crystals (7). In addition, of the proteins that do form crystals, only half may be optimized to form crystals of sufficient quality that will allow structure determination. Thus, in the course of going from a pure protein sample to its determined molecular structure, a success rate of merely 15% is estimated. Therefore, protein crystallization constitutes to be a major obstacle that deserves the utmost attention. A prerequisite for a well-diffracting protein crystal is that the protein molecules form well-ordered lattices in all three dimensions (6). The general belief is that the likelihood to form well-ordered lattices is greatly enhanced by imposing constraints in the conformational motion of the protein of interest. Frequently, the problems associated with protein production and/or crystallization is due to the modular structure of a protein, i.e. the arrangement of the individual protein domains (8). The description of the modular structure of large multifunctional proteins is typically portrayed as "beads on a string", in which the components are composed of protein domains, autonomously folded polypeptide structures (6). In general, a domain comprises 50-150 amino acid residues. Multi-domain proteins are often associated with an inter-domain flexibility, as the individual domains often are connected by flexible and unstructured linker regions, an arrangement that severely decrease the likelihood to form crystals (6,8).

It has since long been established that protein domains will crystallize better as well as produce better diffracting crystals relative to the corresponding full-length protein (7). Being regarded as "evolutionary units", the study of protein domains is often interesting by itself (8). Consequently, large efforts have been made to develop

methods that will improve the production of protein domains amenable for crystallization. Removal of flexible regions, often located at the termini or between two domains, may increase the solubility of the protein of interest. Thus, in the pursuit of producing truncated versions of a protein target, it is of importance that these flexible segments, which may impede both protein production and crystallization, are identified (6).

Traditionally, the domain borders are identified by limited proteolysis, performed under conditions that allow the native form of the investigated protein to be maintained (6). Thus, proteolysis is typically restricted to disordered regions, such as exposed loops and interdomain linkers. As regions that normally are susceptible may demonstrate a dramatic increase in resisting proteolysis upon binding of both small molecules (substrates, cofactors or inhibitors) as well as binding to a macromolecular partner, the method may also be useful in protein characterization (6,8). Alternatively, the domain borders may be identified by computational methods, such as secondary-structure predictions (6). Having identified putative protein domains, the subsequent steps would be to clone the corresponding gene region, express, purify, and finally crystallize the protein fragment of interest (7).

However, crystallization of protein fragments from proteolysis may also be the outcome of non-intended events, such as protease contaminants in the protein or crystalliztion solution (7). Consequently, structural genomics researchers have investigated the effects of *in situ* proteolysis, in which trace amounts of a certain protease, i.e. chymotrypsin, was added to the crystallization trial of the protein of interest. Most interestingly, this method was reported to be successful both for the proteins that previously had failed to form crystals at all as well as proteins that admittedly had formed crystals albeit not suitable for structure determination. As the sites of proteolysis were charted using mass spectrometry, all of the samples examined were found to have been proteolytically digested in the N- and C-terminal regions. It may be wise to optimize the method by varying the choice of protease, protein to protease ratios, digestion times, buffer conditions, as well as the temperature. *In situ* proteolysis was also utilized in part of the work described this thesis and is further discussed in paper V.

Target Selection

It has been estimated that if the structures of about 16,000 carefully selected target proteins were experimentally determined, useful and accurate atomic models could be generated for the greater part of all proteins (3). However, it has to be emphasized that this estimate is based on the assumption that an optimal strategy for target selection is employed. As it may be unlikely to achieve 100% coverage of the protein structure space, this estimate is also based on the aim to obtain correct models for 90% of the proteins. It has been shown that models based on a sequence identity below 30% display significant alignment errors, which results in further errors in the positions of main chain atoms, whereas models with a sequence identity exceeding 30-35% exhibit

comparatively low levels of alignment and structural errors (3,9). The number of experimentally determined structures necessary to cover the protein structure universe also heavily relies on the trustworthiness of the methods for homology modelling (3). As already mentioned, a critical issue is target selection. In the dawn of the era of structural genomics, many target lists focused on what is known as the "low-hanging fruits", moderately small non-membrane proteins with few structural domains (8). Nevertheless, certain regions of the protein structure space will not be easy to cover, such as filamentous and membrane proteins, which typically pose severe problems regarding crystallization. Proteins with low complexity regions may also cause problems, as these regions may be unstructured unless interacting with putative binding partners. For instance, viral proteins have frequently been demonstrated to be partly disordered (10). Thus, the objectives nowadays (post structural genomics) are much more challenging, as the ambition is to confront more relevant albeit difficult biological issues (8).

The work described in this thesis has been carried out in two different projects, each with its own agenda and target list. The first project is composed of proteins from the bacterium *E. coli* and the second is made up of proteins from five different herpesviruses. Thus, prior to the description of the two projects, a brief overview of microorganisms, with an emphasis on bacteria and viruses, will follow.

Microorganisms- An Overview

The term microorganism or microbe refers to organisms that are of microscopic size, typically to small to be seen by the unaided human eye (11). This diverse group of organisms, that are able to sustain life as single cells or cell clusters, include bacteria, algae, fungi, protozoa and also viruses. However, viruses are not cellular and constitute a matter of controversy as they fail to fulfil the criteria to be defined as living organisms. Instead, virus may be viewed upon as genetic elements that have the ability replicate independently of the genome of a host cell but not independently of the host cell itself.

Microorganisms are inhabitants in all parts of the biosphere as long as liquid water is available, even in extreme environments such as hot springs, on the ocean floor, or even deep within Earth (11). Humans encounter microorganisms on a daily basis. Some of these encounters are undesired as certain microbes are pathogenic, causing diseases that may be harmful and even lethal to humans. In addition to human health, pathogens may also cause disease in other animals and in plants, which may have major impacts on economy and welfare of a society. On the other side, much of our daily consumables such as yoghurt, bread and alcoholic beverages, such as beer and wine, are due to the activity of various microorganisms.

In fact, microbial activity constitutes a prerequisite for life on this planet, as microorganisms play a vital role in the decomposition of organic matter, which allows compounds such as carbon, nitrogen and sulfur to be recycled throughout the

biosphere (11). Microorganisms can be associated with all living organisms and often engage to establish close symbiotic relationships with higher plants and animals, with mutual benefits both for host and microbe. For instance, the normal flora of bacteria that reside on the skin and in the intestines of humans are beneficial for both parties (11). From a human perspective, these natural inhabitants may provide us with dietary supplements and also serve as a protection against other unwanted microorganisms. Nowadays, microbes are often exploited in many areas of biotechnologies as sophisticated industrial applications are utilized in the production of pharmaceuticals, food additives and chemicals. However, human exploitation of microorganisms is not a new phenomenon. In fact, humans have used microbes for thousands of years in their daily lives.

The study of microbes may be fundamentally important as these organisms are able to to exist as autonomous entities and thus are equipped with a complete set of functions necessary to sustain life (11). Furthermore, as it has been shown that all living systems display similarities in metabolical pathways, microorganisms have often been used as model systems in life sciences. In order to grow and multiply, all living organisms require, next to access to water, a supply of energy and a carbon source. Microorganisms, such as bacteria, use various nutrients in the environment by transporting these compounds into the cell, either to be utilized in the formation of building blocks necessary for cell growth, or to be used as energy source. In the course of evolution, microorganisms have developed a variety of strategies to provide for the supplies of energy and carbon that are necessary and thus numerous solutions for the same problems have evolved.

Viruses, on the other hand, are completely dependent on their hosts for their survival (11). Viruses are small, typically with a diameter between 20 and 300 nm, and hence their genomes are of limited size. The genome can consist of either DNA or RNA and it may be either single- or double-stranded. It has been suggested that a significant part of animal genomes may have viral origin (12). Most interestingly, some viruses have been shown to encode host proteins that are involved in cellular growth (11,12). This observation is of greatest interest as viral infections may stimulate cellular division and thus be involved in the formation of malignant tumours.

The *E. coli* Project

The main idea behind the *E. coli* project is relatively straightforward: to determine the three-dimensional structure of *E. coli* proteins of which eukaryotic homologues exist. The selection of the targets was primarily based on their involvement in processes of fundamental biological relevance, such as enzymes regulating basal metabolism. In addition, the targets had to fulfil certain other criteria. The proteins should all be soluble, thus not contain any predicted transmembrane helices. Another matter was the size of the targets. Small molecules are more likely to form crystals, as larger and more complex molecules are more inclined to form many different intermolecular contacts (13). Still, it was important that the size of each target was not to small so that

its molecular structure could be determined by other methods, such as NMR. Finally, no structures of the targets or their homologues were to be available in the Protein Data Bank (PDB).

Traditionally, *E. coli* has been the model organism of choice in life sciences (14,15). As all organisms share a common origin, the study of a model organism may aid answering fundamental questions of high biological relevance. Throughout evolution, central metabolic pathways have been conserved to a high degree. The selected life form may be examined in detail, in the hope that the findings may be valuable in a generic context. Thus, the use of model organisms constitute a central issue in research investigating potential causes and/or treatment for human diseases, as the corresponding experiments on humans would be neither practical nor ethical. The vast amount of information gathered from *E. coli* has resulted in a comprehensive understanding about the biology of this model organism (14). For some time now, the attempt to conduct systematic structural investigations of complete prokaryotic proteomes has been part of the agenda of structural genomics (16). As free-living organisms, such as *E. coli*, contain genomes encoding a complete set of functions for maintaining cellular life, the structural determination of the corresponding proteome would be an invaluable asset. Such an achievement would be of great value in the clarification of protein functions that may not be determined from sequence alone. At present, hundreds of bacterial genomes have been sequenced (14). Yet, functional annotation of the corresponding gene products has still not been completed. The genome of *E. coli* is no exception and the functional assignment of the genes of unknown functions constitutes an important challenge in the functional genomics of this organism (14,15). Achievements in this field will not only aid functional assignment in other organisms, but may also result in the revelation of new physiological and biochemical pathways (14). The *E. coli* protein YhaK represents such a protein of unknown function and is further discussed in Paper III.

Furthermore, a number of *E. coli* strains have been identified as pathogens for humans and animals (15). Fortunately, certain metabolic pathways may be exclusive for bacteria and not found in higher organisms. Such pathways specific for bacteria may involve enzymes without counterparts in higher organisms, such as humans, which may serve as targets for structure-based anti-bacterial drug design (16). Thus, a comprehensive understanding of this organism may allow the identification of new key targets amenable for the development of novel therapeutic methods in the treatment of bacterial infections (14-16). The bi-functional enzyme RibD involved in the biosynthesis of riboflavin in *E. coli* has been suggested to be such a target and is further discussed in Paper II. A significant fraction of the *E. coli* proteome is made up of enzymes, many of which play an essential role in cellular life as they interconvert metabolites, produce cofactors, as well as regulate small molecule metabolism (14). The production and conversion of these metabolites may also be interesting in various industrial applications. For example, the carnitine metabolism of *E. coli* is employed in the industrial production of carnitine, an essential compound in the human nutrition. The *E. coli* protein CaiB is highly involved in this process and is further discussed in Paper I.

Mammalian Metabolism of Carnitine

Carnitine is a small and polar molecule that plays a central role in the energy metabolism in humans as well in all other mammals (17). Above all, this compound facilitates the transport of long-chain fatty acids across the inner membrane of mitochondria (18,19). In mammals, the highest concentrations of carnitine are found in heart and skeletal muscles (17). In these tissues, the β-oxidation of fatty acids in the mitochondrial matrix provides a substantial amount of energy for the cell. Carnitine is essential for this event, as the CoA esters of fatty acids are not able to cross the inner membrane of mitochondria (19). This predicament is circumvented by the conversion of the CoA esters to carnitine esters for which transporters exist. As carnitine functions as an acceptor for the acyl group of acyl-CoA, another role for this compound may be in regulating the ratio of CoA:acyl-CoA in the cell (18). Although mammals are endowed with the ability to synthesize carnitine from the amino acids lysine and methionine (20), humans obtain carnitine mainly from the diet, especially from meat products (21). However, humans like other mammals lack the ability to catabolize carnitine by themselves. Instead, this action is dependent on the actions of resident microorganisms in the large intestine acting on carnitine that has not been absorbed in the small intestine. In humans and other nonruminant mammals, this microbial activity on carnitine in the large intestine results in two different degradation products, trimethylamine, excreted via urine, and γ-butyrobetaine, excreted via feces.

Bacterial Metabolism of Carnitine

There are three different ways by which carnitine is metabolized in bacteria. In some bacteria, like *Pseudomonas* species, carnitine undergoes complete degradation and is used as the sole carbon, nitrogen, and energy source (21,22). Other bacteria, such as *Acinetobacter* species, are able to use carnitine as the sole source of carbon, resulting in the formation of trimethylamine. In *Enterobacteriaceae*, such as *E. coli*, carnitine is reduced to γ-butyrobetaine via crotonobetaine, thereby functioning as a putative terminal electron acceptor during anaerobisis. However, these bacteria are dependent on additional sources of carbon and nitrogen, as these are not assimilated from carnitine.

The proteins involved in the carnitine metabolism in *E. coli* are encoded by the *caiTABCDE* operon, which is induced during anaerobisis in the presence of carnitine or crotonobetaine (23). The conversion of carnitine to γ-butyrobetaine in *E. coli* directly involves the activities of three proteins: CaiA, CaiB and CaiD. CaiB, a CoA transferase, transfer the CoA moiety from γ-butyrobetaine-CoA to carnitine, resulting in γ-butyrobetaine and carnityl-CoA (Figure 1) (24). CaiD, carnityl-CoA dehydratase, catalyzes the dehydration of carnityl-CoA resulting in crotonobetainyl-CoA. Crotonobetainyl-CoA is furthermore reduced by the crotonobetainyl-CoA reductase, CaiA, forming γ-butyrobetaine-CoA. However, both the CaiA reduction and the CaiD dehydration reactions are dependent on the participation of CaiB. CaiT is an antiporter, which catalyzes the exchange for the substrate carnitine for the product γ-butyrobetaine (25). CaiC is a highly specific ATP-dependent betaine:CoA ligase, responsible for the formation of CoA derivatives of trimethylammonium compounds

(26). Thus, CaiC is critical for generating sufficient starting material in order for the CaiB reaction cycle to initiate. The function of CaiE is yet to be established. A transcription activator CaiF is encoded by the *caiF* gene downstream of the cai operon (27). The *caiF* gene, which is repressed by oxygen, is only expressed during anaerobic conditions regardless of the occurrence of carnitine. The transcription of the *caif* gene occurs in the opposite direction to the *cai* operon resulting in a monocistronic mRNA (24,27). It has been shown that CaiF, possibly activated by carnitine (27), together with the cyclic AMP receptor protein (CRP) activate the transcription of the *caiTABCDE* operon as well as the f*ixABCX* operon (27,28). Proteins encoded by the f*ixABCX* operon, co-regulated with the *caiTABCDE* operon, have been shown to be essential for anaerobic carnitine reduction by facilitating the transfer of electrons required for the CaiA reaction (29).

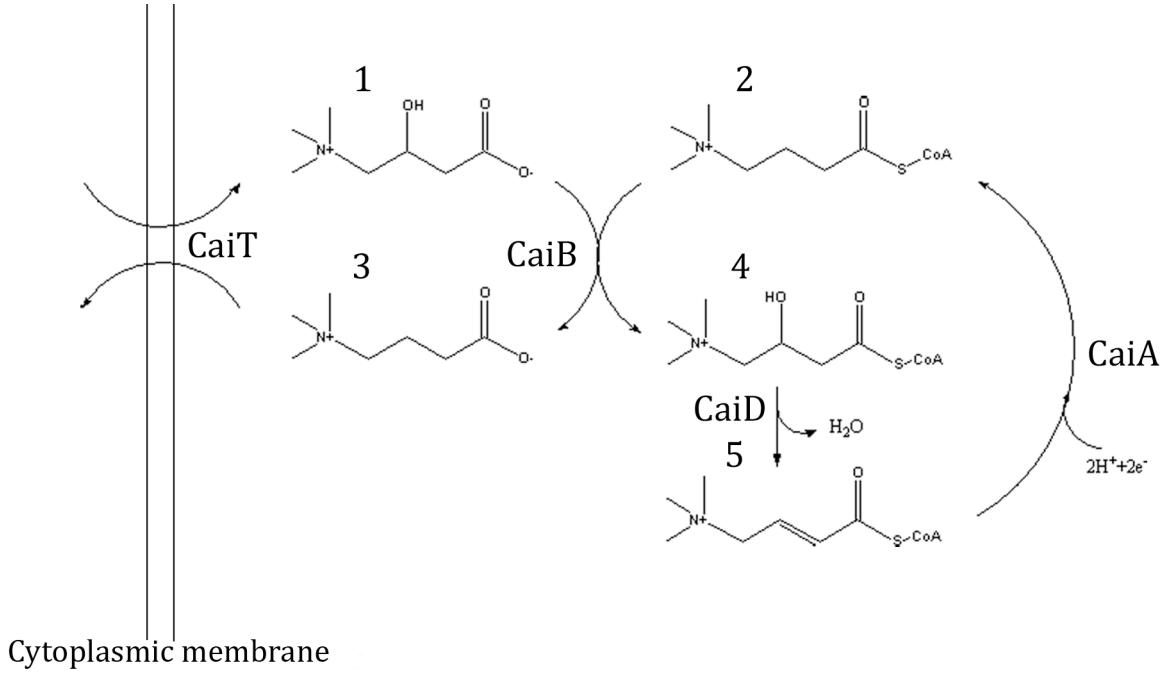


Figure 1 The reaction scheme for carnitine metabolism in *E. coli*. Molecule 1 represents carnitine, 2 γ-butyrobetaine-CoA, 3 γ-butyrobetaine, 4 carnityl-CoA, and 5 crotonobetainyl-CoA.

CaiB- A Type III CoA Transferase

CaiB belongs to the type-III CoA transferase family (30). The function of CoA-transferases is to catalyze the transfer of the coenzyme A moiety from CoA-thioesters to free acids, a reaction that is reversible. At the moment, three enzyme families have been identified and characterized. Especially, the reaction mechanisms for enzymes belonging to families I and II have been well characterized. The reactions of both these families include the formation of thioester and anhydride intermediates. However, the reaction mechanisms of these two families are very different. The members of the type-I family, catalyzes the transfer of the CoA group via a ping-pong mechanism, in which a glutamate residue in the active site serves as acceptor of covalently attached intermediates (30,31). The members of the small type-II family use another mechanism, in which the reaction is executed via a ternary complex without intermediates being covalently bound to the enzyme (30).

However, a third class of CoA-transferases was identified, which were involved in anaerobic metabolic pathways (30). The amino acid sequences of these enzymes differ from the ones of the two other families. Although only bacterial enzymes have been characterized yet, similar genes are also found in eukaryotes and archaea. The first member of the type-III family to be characterized was formyl-CoA transferase (FRC) from *Oxalobacter formigenes*, involved in the catabolism of oxalate in the intestine of mammals (30,32). This enzyme was also the first member of the type-III family to be structurally determined (32). FRC is responsible for catalyzing the first step in oxalate degradation, in which the CoA moiety is transferred from formate to oxalate. Another member of the type-III family, YfdW from *E. coli*, has also been structurally determined (33). Recently, it was shown that this protein also is a formyl-CoA transferase (34). In the structures of both FRC and YfdW, an unusual fold was observed in which the dimers are intertwined (32,35). Due to the fact that the reactions catalyzed by CaiB and CaiD in the carnitine metabolism in *E. coli* are reversible (36), enzymes of the *caiTABCDE* operon may be used in industrial large-scale industrial bioproduction of the essential compound L-carnitine with crotonobetaine, a wasteproduct, as start material (37,38). Thus, in addition to aid in a more comprehensive understanding of this biochemical pathway, structural determination of these enzymes may also be of interest for industrial production of L-carnitine, an essential compound in the human nutrition.

The structure of CaiB (Paper I)

The structure of CaiB was determined both in the apo form and in complex with CoA. Similarly to what was observed for FRC (32) and YfdW (35), the two monomers of the homodimer formed an interlaced ring structure (Figure 2). Out of its 405 amino acids, each monomer could be traced from amino acid 4 to 403. Each monomer is composed of a small and a large domain. The large domain is composed of both the N- and the C-terminal parts of the protein, whereas the small domain is composed of the residue in between (residues 227-323). The dimer interface is primarily made up of residues from the large domain. Sequence analysis of the two domains showed that the large domain was much more conserved in the CaiB homologues, approximately 30% identity to FRC, whereas there was no significant sequence identity between the small domains of the two proteins. The crystal structure of CaiB confirmed that the large domain of CaiB and FRC were highly similar and although the small domains are different in the two protein structures, they share a related topology and most likely a common origin. The largest structural difference between CaiB and FRC is located at the beginning of the small domain. In the small domain of FRC, a loop structure of 36 amino acids is observed (32), whereas the corresponding loop structure of CaiB is only 16 amino acids. This difference results in CaiB lacking the glycine-rich loop that is observed pointing into the active site of the FRC structure. In FRC, this loop structure changes conformation upon CoA binding and have been indicated to play a functional role in the reaction mechanism. The absence of the glycine-rich loop in CaiB further results in that its active site is larger than in FRC, which is reasonable due to the size difference between carnitine and oxalate.

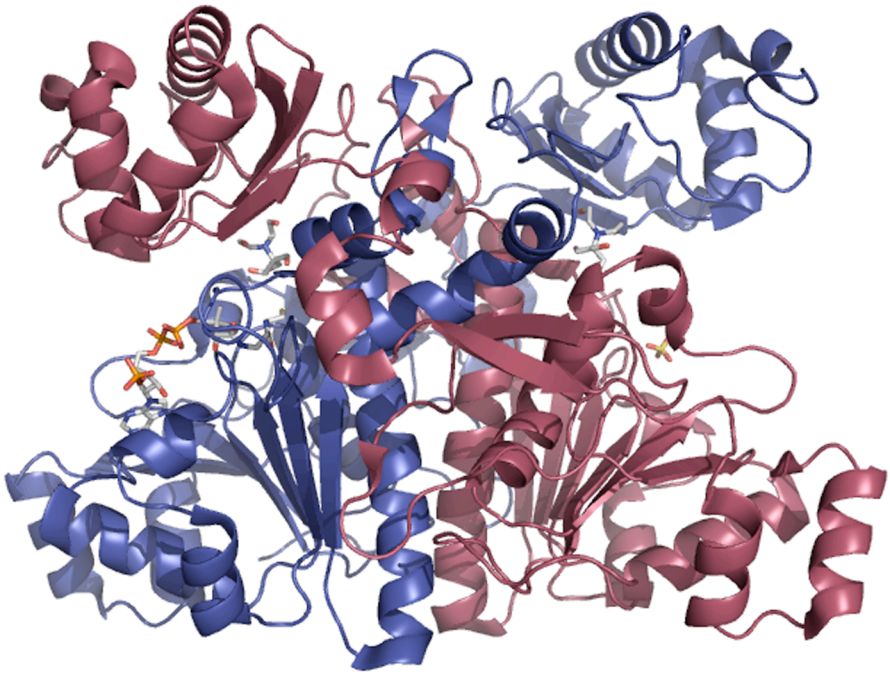


Figure 2 The overall structure of the CaiB homodimer. One monomer is colored red and the other blue. In both active sites a bis-tris molecule is shown in stick representation. In one active site, the CoA molecule is bound whereas a sulfate ion is bound in the other.

The Active Site of CaiB

The active site of CaiB is situated in a large cleft formed in the interface between the small domain of one monomer and the large domain of the other monomer (Figure 2). The homodimer of CaiB is likely to bind CoA in both of its active sites. Still, CoA could only be observed in one of the active sites, which probably results from crystals contacts made with a symmetry-related molecule. Interestingly, CoA formed interactions solely with the residues of the more conserved large domain (Figure 3A). Thus, the observation that the CoA molecule interacts solely with the large domain of CaiB indicates that its higher degree of sequence conservation is related to the binding of CoA, a feature common to all members in this family. As expected from the nature of the CoA molecule, many of the interactions formed with CaiB are of a hydrophobic nature. The SH-group of the CoA molecule interacts with CaiB via two hydrogen bonds, one to the main chain amino group of Ile24 and the other to the carboxyl group of Asp169. Asp169 is a catalytically important residue that is completely conserved amongst the type III CoA transferases (39).

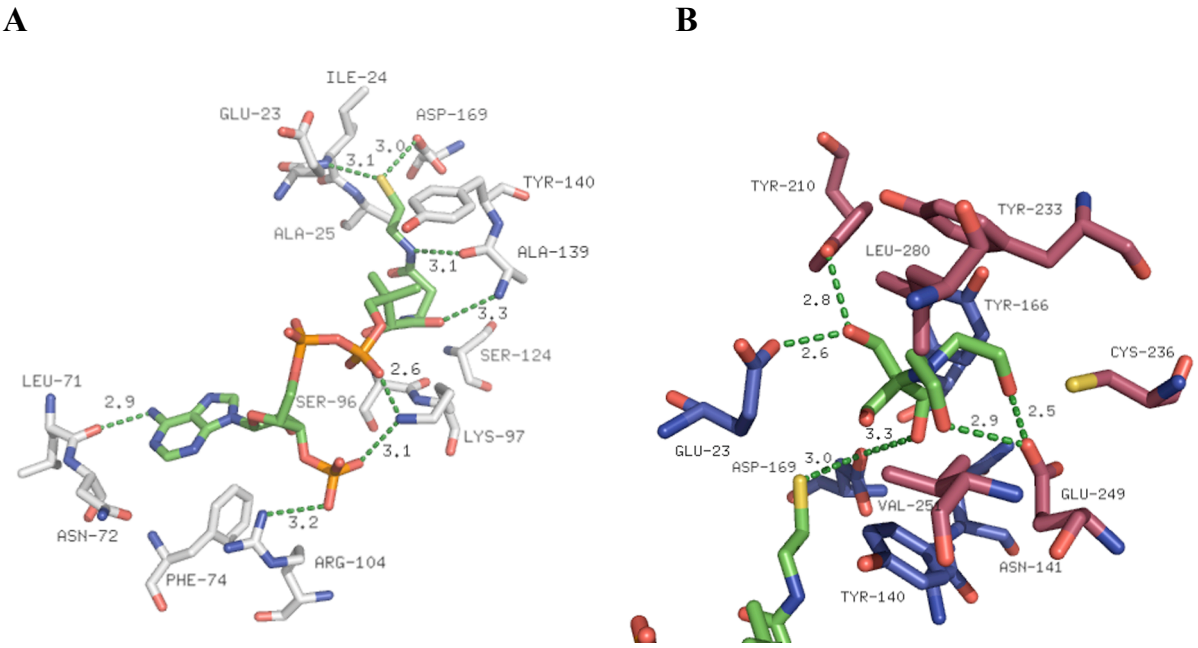


Figure 3 (A) The interactions formed between residues from the large domain of CaiB and the CoA molecule (green). The adenine base of the CoA molecule makes hydrogen bond contact to the main chain carbonyl oxygen of Leu71. Arg104, which coordinates a sulfate ion in the apo structure (not shown), and Lys97 anchor the phosphate groups in position. The SH group of the CoA molecule makes hydrogen bonds to the main chain amino group of Ile24 and to the carboxyl group of the catalytically important Asp169. (B) The interactions formed between residues from the small (red) and the large (blue) domain of CaiB and the Bis-Tris molecule (green). Parts of the CoA molecule (green) are also visible. Hydrogen bonds are illustrated by green dashes and measurements are shown in the units of Ångström (Å).

Although extensive efforts to incorporate the real substrates into the crystal structure of CaiB were unsuccessful, interesting clues regarding substrate binding were obtained from the apo and CoA complex structures. In both these structures, a Bis-Tris molecule from the crystallization solution was found in a hydrophobic pocket in the active site making interactions mainly with residues of the small domain of CaiB (Figure 3B). The position of the Bis-Tris molecule in the CaiB structure corresponds to where the glycine-rich loop is located in the FRC structure. In addition, one of the interactions that the Bis-Tris molecule makes with CaiB is a hydrogen bond to Asp169. Although differences in size, as Bis-Tris is considerably bulkier than both carnitine and γ-butyrobetaine, the binding site pocket identified for the Bis-Tris molecule may constitute the binding site for both of the substrates involved in the transfer of the CoA moiety. Both the real substrates and Bis-Tris are very polar compounds. The positions of Glu249 from the small domain and Glu23 from the large domain made us suggest that these negatively charged residues might serve to neutralize the positive charge on the nitrogen of the substrates. In addition, Glu23 was found to uphold a strained geometry, further implying functional significance. Shortly after our results were published, another structural study of CaiB and its complexes with CoA and carnityl-CoA was reported (40). The two sets of structures are very similar, with rmsd of 0.72Å and 0.73Å between the apo and CaiB-CoA complex structures respectively. It was also found that the locations of the nitrogen atoms of the Bis-Tris molecule and carnitine were in good agreement. In addition, it was established that Glu23 and Glu249 in fact served to compensate for the positive charge of the nitrogen of the substrate in a hydrophobic environment as hypothesised from

our structures. Further, several hydrophobic residues, including Tyr140, Tyr166, Cys236, Val251 and Leu280, were found to interact with the carnityl moiety, in agreement with was found for Bis-Tris molecule, which confirmed that this pocket forms the binding site for the real substrates. As already mentioned, residues from the large domain, which show a higher degree of conservation than the small domain, are likely to be responsible for the binding of CoA that is a common feature for members of this family. On the other hand, the high variability in the small domain of the type III CoA transferases may be due to the variety of substrates that these enzymes act upon.

Interestingly, the binding of CoA induced a domain movement, as the small domain of CaiB approaches the large domain resulting in a more closed active site. The start (amino acid 227) and end (amino acid 323) of the small domain constitute a hinge around which a movement of approximately 3Å is carried out in the presence of CoA. Similar observations were done in the other structural study of CaiB (40). The conformational change observed upon CoA binding may indicate that ligand binding occurs through an induced fit mechanism which allows the active site residues to take on the conformation(s) necessary for product formation. Alternatively, the closure of the active site, resulting from the domain movement, may serve to protect the reaction intermediates during the reaction. In fact, the glycine-rich loop in the active site of FRC has been observed to change conformations at different steps of the reaction and has been indicated to play a central role in the protection of the mixed anhydride and the thioester from hydrolysis (41).

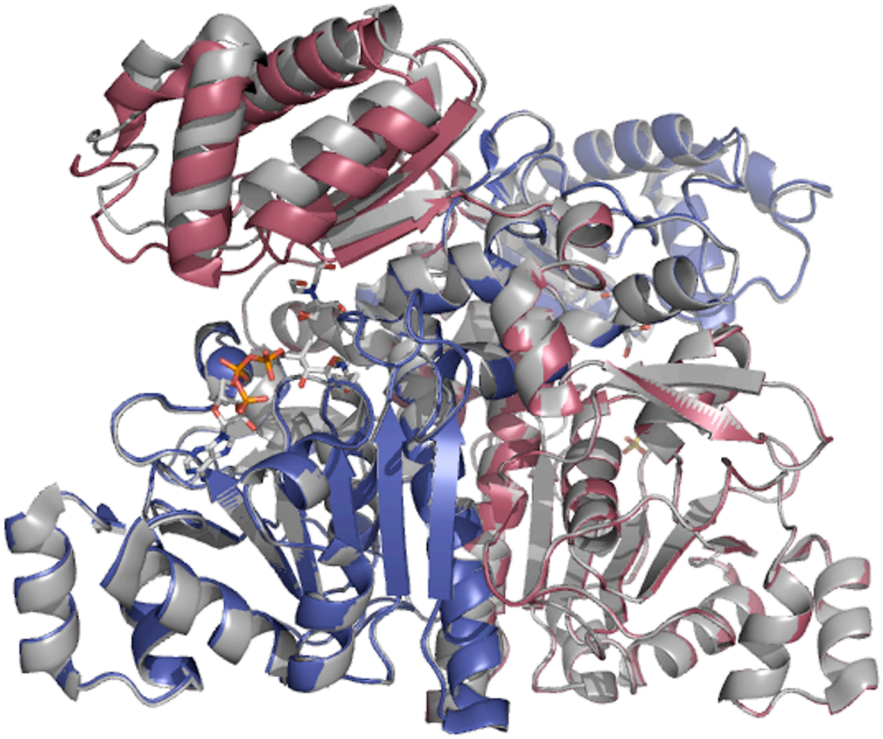


Figure 4 Superposition of the apo structure upon the CaiB structure in complex with CoA. The binding of CoA induces a domain movement in which the small domain moves closer to the large domain, which results in a more closed active site. The apo structure is shown in gray and the complex structure in red and blue.

Updated Reaction Mechanism of Class III CoA-transferases

At the time of our structural studies, the reaction mechanism of the Class III CoA-transferases was not fully known. Recent studies on FRC from *O. formigenes* have provided more clues in how members of the Class III CoA-transferase family catalyze the transfer of the CoA moiety between substrate and product (41). In analogy with the Class I family, the Class III CoA-transferase reaction proceeds by forming aspartyl mixed anhydride intermediates with the oxyacids and covalent thioester intermediates with the CoA molecule. However, in contrast to the Class I family, the Class III family does not seem to carry out the reaction in a ping-pong fashion, in which the CoA donor leaves before the CoA acceptor binds. In addition, as it has been shown that the thioester of the CoA donor can be hydrolyzed and the mixed anhydride can be formed in absence of an acceptor oxyacid, formation of a ternary complex also does not seem to be a requisite. Instead, the CoA donor is most likely released simultaneously as the product.

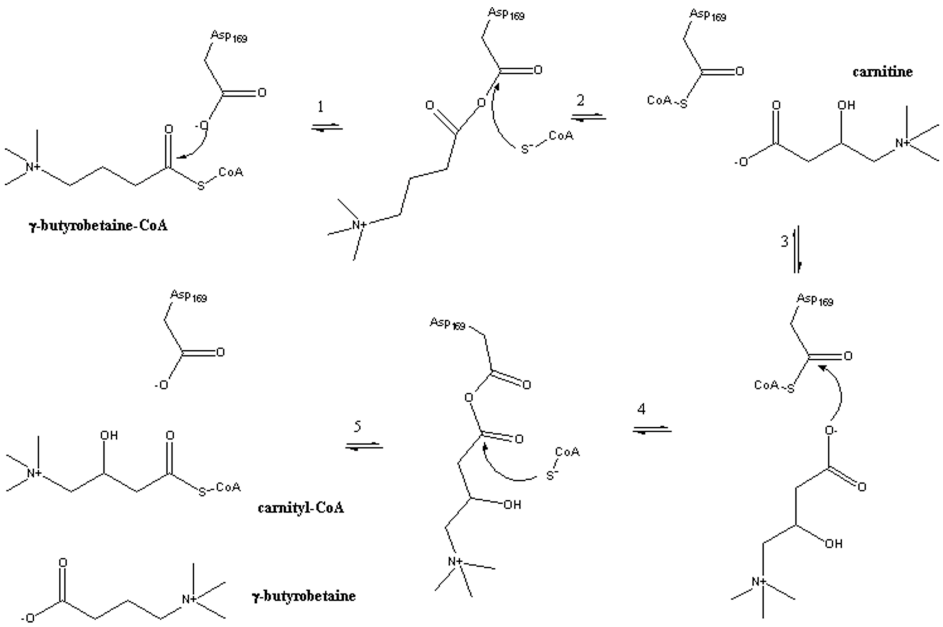


Figure 5 An updated suggestion for the CaiB reaction mechanism based on recent structural studies of Formyl-coenzyme A transferase from *O. formigenes* (41).

Biosynthesis of Riboflavin

Riboflavin, also known as vitamin B2, is an important precursor for the flavin nucleotides, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (42). These flavonucleotides serve as coenzymes for flavoproteins, enzymes involved in electron transfer reactions essential in the energy metabolism of cells (43). While humans obtain riboflavin from the diet, bacteria rely mainly on their own biosynthesis and/or exogenous uptake from the surroundings (42). As many pathogenic bacteria do not have transporters for exogenous uptake and depend on endogenous biosynthesis, bacterial proteins involved in the biosynthesis of riboflavin constitute potential anti-bacterial drug targets (44). For instance, *E. coli* lack transporters for exogenous uptake and is dependent on biosynthesis of riboflavin. Thus, structural determination of enzymes in this pathway may result in structure-based anti-bacterial drug design.

RibD- A Bifunctional Deaminase and Reductase

In order to synthesise riboflavin, a molecule of GTP and two molecules of ribulose 5-phosphate are needed (45). GTP cyclohydrolase II catalyzes the formation of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5′-phosphate (DAROPP) from GTP. This compound is further converted to 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (ARIPP) by two additional reactions, deamination of the pyrimidine ring and NAD(P)H-dependent reduction of the ribose (Figure 6). The sequential order of these two reactions is not the same in all organisms. In higher plants and eubacteria, such as *E. coli*, deamination occurs before reduction, whereas the opposite order takes place in yeast and archaea (46-49). Furthermore, while yeast, plants and most archaea utilize two separate enzymes for these reactions, most eubacteria make use of a bifunctional protein such as the RibD protein of *E. coli* (EcRibD) (50). EcRibD contains two separate domains for the deamination and reduction reactions. In the study of EcRibD described in this thesis, focus was on the reductive reaction that converts the ring structure of 5-amino-6-ribosylamino-2,4(1H,3H)-pyrimidinedione 5′-phosphate (AROPP) into its ribityl form, ARIPP. For this reaction to occur, the enzyme is dependent on the cofactor nicotinamide adenine dinucleotide phosphate, NADPH, or its nonphosphorylated equivalent NADH. The cofactor may provide two electrons and a proton, in the hydride form, while a second proton, needed for the reaction, may originate from either the enzyme itself or the solvent. Either hydrogens, pro-R or pro-S, at the C4 position of the nicotinamide ring of NADPH can be exploited in hydride transfer, but most enzymes are stereospecific and discriminate between the two. The reductive reaction has been suggested to be initiated by the formation of a Schiff base intermediate at C1', by the abstraction of a proton from the adjacent amine, followed by the direct hydride transfer to C1' (51).

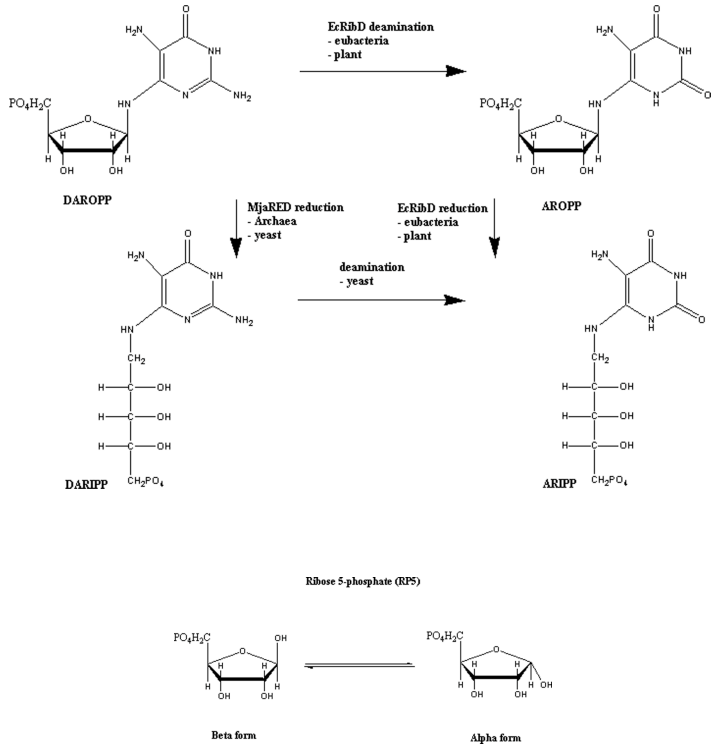


Figure 6 Deamination precedes reduction for the bifunctional RibD proteins of eubacteria as well as in higher plants whereas the opposite order is observed for yeast. The different substrates for the various reactions are shown as well as the substrate analogue, ribose 5-phosphate (RP5).

Structure of EcRibD (Paper II)

The structure of the EcRibD protein was determined in its apo form as well as in two binary complexes with $NADP^+$ and ribose 5-phosphate (RP5), cofactor and a substrate analogue, respectively. The apo structure and both of the binary complexes were determined at 2.6 and 3.0 Å resolution respectively. Although the deaminase domain of RibD supposedly binds Zn^{2+} (49,52), no trace of the metal was found in the structures of either the apo form or the binary complexes. The absence of the zinc ion most likely was a result from using EDTA during purification. As a consequence, the deaminase domain was disordered in the Zn-binding region. Therefore, the emphasis of this structural study of EcRibD was on the characterization of the active site of the reductase domain. The crystal structure of EcRibD revealed a homodimer, in which each monomer is composed of an N-terminal deaminase domain, comprising residues 1-145, and a C-terminal reductase domain that comprises residues 146-367 (Figure 7). The SCOP (Structural Classification of Proteins) database (53) identified the deaminase and the reductase domains to belong to the Cytidine deaminase-like superfamily and the Dihydrofolate reductase-like superfamily respectively.

During our work, the crystal structure of RibG, an ortholog from *Bacillus subtilis* with a 40% sequence identity to EcRibD, was determined (52). Interestingly, while the structure of EcRibD revealed a dimer, RibG from *B. subtilis* (BsRibG) was shown to form a tetrameric structure. Both enzymes form a dimer interface between their reductase domains, but in BsRibG additional interactions between the deaminase domains of the two dimers resulted in the formation of a tetramer. Although the

majority of the secondary structure elements are common to both enzymes, some structural features seem to distinguish the oligomerization mode of EcRibD from BsRibG. For instance, the reductase domain of EcRibD has an insertion, comprising residues 210-221, which is not present in BsRibG. The relative domain orientation of EcRibD also differs from what is observed in BsRibG. These two features are likely to cause steric clashes inhibiting formation of a stable EcRibD tetramer. In addition, hydrophobic residues in BsRibG that are involved in packing interactions between the deaminase domains are in EcRibD replaced by Arg9, Glu40 and Gln44. The position of these charged residues does not allow the formation of salt-bridges in a putative tetramer interface, but is more likely to cause the deaminase domains repelling each other.

In the structures of the binary complexes, the two compounds, $NADP^+$ and RP5, were observed to differ in their binding mode between the two subunits of the homodimer. For both the cofactor and the substrate analogue, the electron density was of good quality in subunit A. RP5 did not bind at all in subunit B. Although most parts of the $NADP^+$ molecule bound identically in the two subunits, the binding of the nicotinamide ring and the ribose were different. However, in contrast to subunit B, subunit A was independent of crystal contacts in addition to the electron density being of higher quality and hence chosen to be trustworthier. An interesting feature of the active site of the reductase domain is the presence of a loop structure, comprising residues 159-173, that interacts with both the cofactor and the substrate analogue. This loop structure was observed to assume four different conformations in the two subunits. In an attempt to be consistent with previous studies of *E. coli* dihydrofolate reductase (EcDHFR) (54), these conformations were termed “substrate occluded”, “cofactor and substrate occluded” (C&S-occluded), “accessible” and “disordered”. Two of these conformations, the “accessible” and the “substrate occluded” states observed in subunit A, were independent of crystal contacts. In the “accessible” state, residues 164-166 were excluded due to disorder, whereas the active site loop is modelled with no gaps in the “substrate occluded” form. The “substrate occluded” state was observed in the binary complex with $NADP^+$, in which the active site loop is ordered to form an α-helix which seems to cover the binding site of the substrate analogue. For both the apo and the binary complex with RP5, the active site loop was observed in the “accessible” conformation.

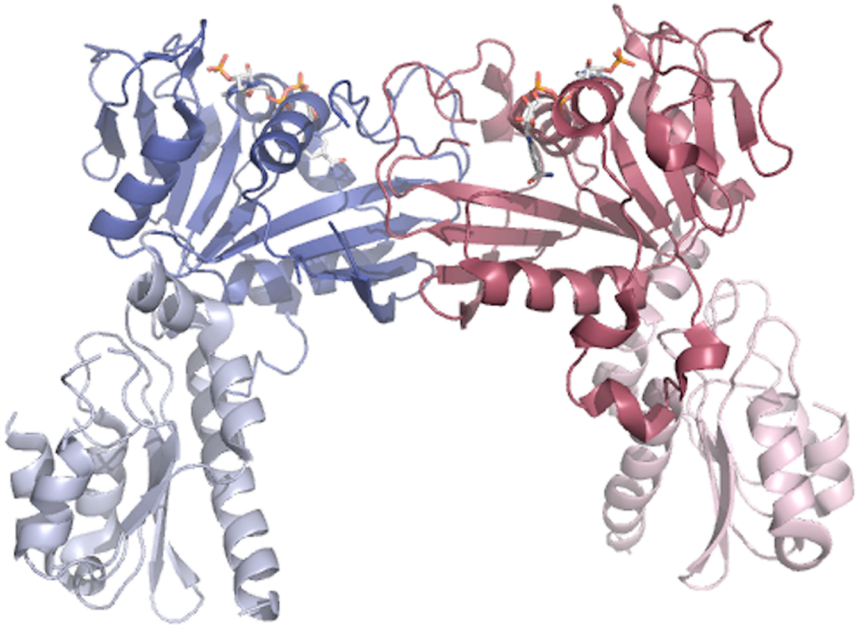


Figure 7 The overall structure of the EcRibD homodimer in complex with the oxidized cofactor $NADP^+$, which is shown in stick representation in the active site of both reductase domains. Only residues from the reductase domain participate in the dimer interface.

The Active Site Loop of EcRibD

In the binary complex of EcRibD together with the substrate analogue RP5, the active site loop retained the "accessible" conformation upon RP5 binding, similar to what was observed for the apo structure (Figure 8A). In the structure of EcRibD in complex with RP5, the substrate analogue was modeled so that the O1 hydroxyl, which corresponds to the amine nitrogen of the real substrate for the reductase reaction adjacent to the proposed hydride acceptor C1' (51), was interpreted to form a short hydrogen bond to Asp200. The interpretation of the 3Å data also resulted in that the O3 hydroxyl was modeled to form hydrogen bonds to the carboxyl group of Glu299 and to the guanidium group of Arg184. The resulting model displayed that both the phosphate group and the hydroxyl groups of the substrate analogues seemed to form several interactions to conserved residues of EcRibD, and was interpreted to reflect the binding mode of the actual substrate.

In the binary complex of RibD together with the oxidized cofactor, $NADP^+$ was observed to bind to the reductase active site in an outstretched conformation (Figure 8B). The binding of $NADP^+$ to the reductase active site of subunit A changed the conformation of the active site loop to the "substrate occluded" state. In this conformation residues 167-170 adopted an α-helical structure, which resulted in that the side chains of Glu167 and Trp170 occupied the binding sites for the ribose respective the phosphate groups of RP5. This conformational change allowed Trp170 to be involved in pi-pi stacking interactions with the nicotinamide ring of $NADP^+$. In addition, the position of Asp200 was shifted due to the formation of a hydrogen bond to the NO2 hydroxyl of the nicotinamide ribose.

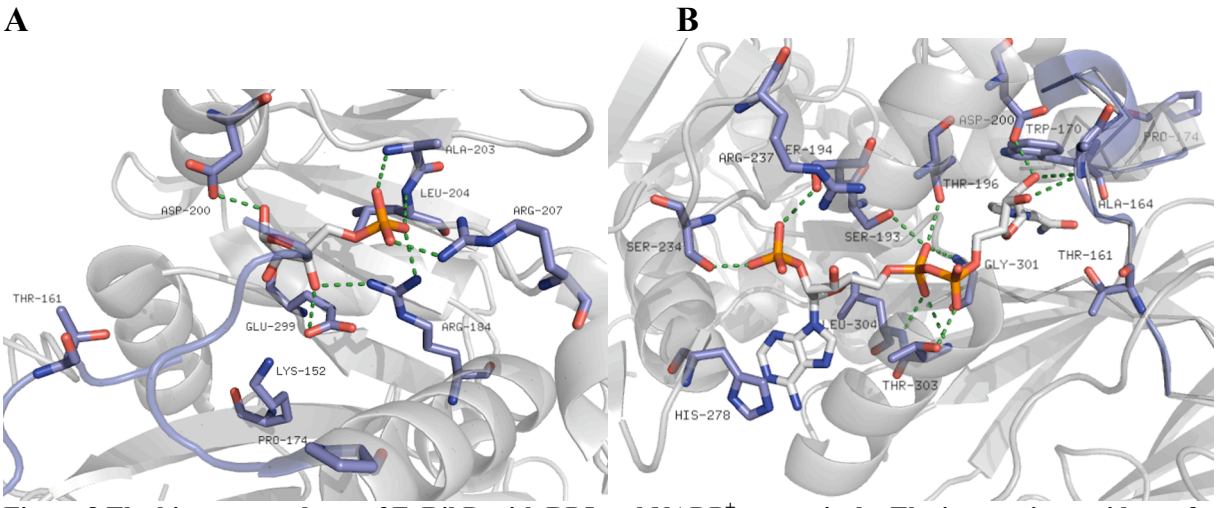


Figure 8 The binary complexes of EcRibD with RP5 and $NADP^+$ respectively. The interacting residues of EcRibD as well as the ligands are shown in stick representation. The EcRibD reductase domain residues are colored blue while RP5 and $NADP^+$ are colored white. Hydrogen bonds are illustrated by green dashes. For clarity the views are different for the two complex structures. (A) The binary complex of EcRibD and the substrate analogue, RP5. The phosphate group of RP5 is anchored in position by the side chains of Arg184 and Arg207 as well by the main chain nitrogen atoms of Ala203 and Leu204. (B) The binary complex of EcRibD and the oxidized cofactor, $NADP^+$. The pyrophosphate group of $NADP^+$ is coordinated by Ser193, Thr196, Gly301, Thr303 and Leu304 and its negative charge also seem to be compensated by the helix dipole effect due to its position close to the N-termini of helix 10. Both hydroxyls of the nicotin ribose form hydrogen bonds to the main chain nitrogen of Ala164. The amine group of the adenine interacts with His278 via pi-pi interactions while its phosphate group interacts with Ser194, Ser234 and Arg237.

Prior to our work, structural studies on *E. coli* dihydrofolate reductase (EcDHFR) had revealed different conformations of an active site loop depending on ligand binding in the binding sites for substrate and cofactor (54,55). The "accessible" and "substrate occluded" conformations (Figure 9), which have been observed in our structural studies of EcRibD, as well as the "occluded" and "closed" conformations of EcDHFR are all independent of crystal contacts (55). Therefore, we attempted to compare the "accessible" and "substrate occluded" states of the EcRibD structures to the "occluded" and "closed" conformations of EcDHFR, which have been indicated to represent different steps in the reaction mechanism (54,55). However, the "closed" conformation of EcDHFR does not have a counterpart in our studies of EcRibD, whereas it is frequently observed among DHFRs from various sources (54).

In the two complex structures of EcRibD, the flexible part of the active site loop appears to be comprised of 12 residues, starting at Thr161 and ending at Pro174. The corresponding active site loop of EcDHFR, on the other hand, involves a stretch of 10 residues, with the start and end residues Ile14 and Pro25, respectively. Despite the differences in length, Thr161 and Pro174 of EcRibD are the structural equivalents of Ile14 and Pro25 of EcDHFR, respectively. In both enzymes, the active site loop is observed to be able to induce a helical conformation. In EcRibD, the active site loop forms an α-helix upon binding of the oxidized cofactor, $NADP^+$, resulting in that binding of the substrate is sterically hindered. In contrast, the formation of a 3_{10}-helix in EcDHFR upon substrate binding consequently blocks the binding of the cofactor. As already mentioned, the "occluded" and "closed" conformations of EcDHFR have been coupled to different steps in the reaction. After catalysis, the product is not

released until the oxidized cofactor $NADP^+$ is exchanged for its reduced form, NADPH, bound to the enzyme (55). The "occluded" conformation interferes with $NADP^+$ binding resulting in that the oxidized cofactor may be "pushed out" of the active site after product formation. Although the catalytic cycle of EcRibD is not fully understood, these similarities may provide clues. It is probable that the active site loop of EcRibD is involved in the binding of substrate and release of the product. In analogy to what has been suggested for EcDHFR, the "substrate occluded" state of EcRibD, in which binding of the substrate is hindered, may reflect that the product is expelled before the release of the oxidized cofactor.

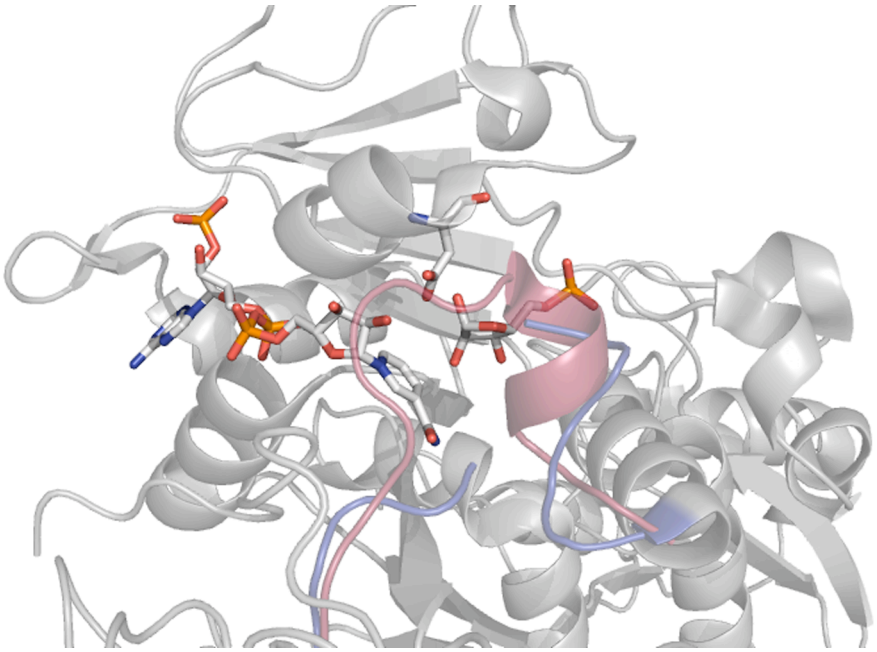


Figure 9 The different conformations of the EcRibD active site loop, which are independent of crystal contacts. The active site loop is colored in blue for the "accessible" conformation and in red for the "substrate occluded" conformation. Asp200, RP5 and $NADP^+$ are shown in stick representation.

Determination of Substrate Specificity

As stated before, the order of the deamination and the reductive reactions may vary in different species. In eubacteria and higher plants, deamination is followed by reduction (48,49). Although no archaeal deaminase enzyme yet has been identified, Archaea is indicated to follow the yeast pathway, in which reduction precedes deamination (47). The sequential order of these reactions is determined by the substrate specificities that the reductase and deaminase domains of different species exhibit. We therefore compared the reductase domains of EcRibD and an archaeal reductase of *Methanocaldococcus jannaschii* (MjaRED) (56). The two enzymes differ in their substrate specificities as the MjaRED substrate, DAROPP (Figure 6), serves as a substrate for the EcRibD deaminase domain while MjaRED does not act on AROPP, the substrate of the reductase domain of EcRibD (47). Thus, our model with RP5 bound to the active site of the reductase domain of EcRibD was compared with a model of MjaRED in which its substrate, DAROPP, had been positioned by *in silico* docking together with NADPH by *in silico* docking to represent a ternary complex (56).

In both models, the phosphate groups of the substrates were found to make similar interactions. The formation of a hydrogen bond between the O3 hydroxyl and the equivalent of Arg184 of EcRibD was also observed in the docking experiments of MjaRED. However, in the binary complex of EcRibD and RP5, Glu299 is modeled to form hydrogen bonds with the O2 and the O3 hydroxyls of RP5, while Asp200 forms hydrogen bonds to the O1 hydroxyl, which corresponds to the amine of the real substrate. In contrast, the MjaRED ternary complex, *in silico*, identified Glu154 of MjaRED, the equivalent of EcRibD Glu299, to interact with the amine group of the substrate. Most interestingly, Lys152, which forms a salt bridge to Glu299 in EcRibD, is not a conserved feature of MjaRED. The corresponding residue of MjaRED, Asn14, was observed to interact with the amino group at position 5 in DAROPP, an interaction that is unlikely to occur with EcRibD Lys152 as such an interaction would be unfavorable. These differences suggest that the two enzymes may differ in their binding modes of the ring structures of the substrates and that Lys152 may form an important factor in discriminating between the two different substrates in EcRibD. Lys152 is highly conserved in bacteria and plant while not present in yeast and Archaea. Recently, the complex structure of BsRibG with its substrate showed that this residue indeed forms favorable interactions with O2 of the pyrimidinedione ring of AROPP, thus discriminating between the substrates for deamination and reduction (57).

The Reductive Reaction– an Updated Reaction Mechanism

The reductive reaction has been suggested to proceed via the formation of a Schiff-base intermediate (51). However, recent kinetic isotope studies suggested that the reductive reaction occurs via a direct hydride transfer from the C4 donor of $NADP^+$ and the C1 acceptor of RP5 (58). Nonetheless, the identity of the base that would initiate such a reaction has sofar been unknown. In the DAROPP model, Glu154 of MjaRED appears to be a suitable candidate (56). In EcRibD, the corresponding Glu299 is located between two positively charged residues, Arg184 and Lys152, which are likely to neutralize its negative charge. Instead, based on our structural study of EcRibD in its apo form as well as in complex with its substrate analogue, RP5, and the oxidized cofactor, $NADP^+$, we proposed that Asp200, whose carboxylate is close to the O1 hydroxyl of RP5 corresponding to the amine of the actual substrate, to be a more likely candidate for initiating the reductive reaction by abstracting a proton from the amine group. However, the RP5 compound that was added to our EcRibD crystals was an equilibrium mixture of the α and β forms that only differ in the position of a single hydroxyl group (Figure 6). It was not possible, at 3 Å resolution, to fully distinguish which of the two that in reality was bound to EcRibD. It may be that both the α and the β forms bound to EcRibD. However, due to the fact that the β form of RP5 resembles the actual substrate more than the α form, the β form was modeled in the electron density of the binary complex of EcRibD and RP5. When the sugar plane of the RP5 β form was "flipped" 180 degrees relative the final model, the B-factors of the RP5 molecule rose significantly, indicating that it was a less likely model of the RP5 β form. In addition, the final modeling of the β form allowed the ribose hydroxyls of RP5 to interact by hydrogen bonds with invariant residues of the RibD enzymes.

Recently, the structure of BsRibG was determined in complex with AROPP, the substrate for the reductive reaction (57). Interestingly, the data indicate the binding of a ribitylimino intermediate rather than the actual substrate, which provides evidence for the involvement of an Schiff base intermediate (51). Although the phosphate groups of the RP5 molecule and the AROPP intermediate in the binary complexes of EcRibD and BsRibG, respectively, are similarly coordinated, their sugar moieties are positioned in opposite directions. Thus, the complex structure of BsRibG provides convincing evidence that Glu290 of BsRibG, which corresponds to Glu299 in EcRibD, instead of Asp200 is involved in the proton transfer (57).

The Cupin Superfamily

The cupin proteins constitute one of the most functionally diverse superfamilies with enzymatic as well as non-enzymatic members (59,60). The Rmlc-like Cupin superfamily was discovered as it was shown that the wheat protein germin, expressed at an early stage of wheat embryo germination, shared sequence similarity to spherulin, a stress-related protein in the slime mould *Physarum polycephalum*. Later on, a group of germin-like proteins (GLPs) in dicotyledonous plants as well as the globulin storage proteins from plant seeds and spores were found to share the similarity. This new class of proteins with their characteristic β-barrel structure was subsequently named cupins referring to "cupa", the latin term for small barrel (60). The cupin domain is characterized by two structural motifs, Motif 1 and Motif 2, both composed of two β-strands (59,60). The two conserved motifs are separated by a segment with lower degree of sequence conservation, corresponding to two additional β-strands and an intervening loop. This intermotif region may vary greatly in size with as little as 11 amino acids (60) in some microbial enzymes up to 50-100 amino acids in seed storage proteins and eukaryotic transcription factors (59-61). The amino acid sequence for Motif 1 and Motif 2 is typically G(x)5HxH(x)3,4E(x)6G and G(x)5PxG(x)2H(x)3N respectively (60). It is in these motifs that residues important for metal binding and catalysis often are found (59,60). Typically, these catalytically important and metal coordinating amino acids are composed of two histidines and one glutamate from Motif 1 and one histidine from Motif 2.

The SCOP database (53) divides the Rmlc-like Cupin superfamily into 24 different subfamilies, but this number may turn out to be even higher. The present knowledge of the cupins is largely based on the structural determination of several cupin members for which no additional biochemical data is available (60). Nevertheless, these structural studies have resulted in the expansion of the Rmlc-like Cupin superfamily, as new members with limited sequence similarity have been identified. The cupin members may also be categorized into different subgroups according to the presence of a single cupin domain (monocupin), two cupin domains (bicupin), or more than two cupin domains (multicupin). In monocupins, the single cupin domain may make up the centre of a simple protein or constitute one out of several domains of a complex protein. Although most monocupins are enzymes, this subclass also includes some microbial transcription factors such as AraC, in which the effector-binding cupin domain is combined with a DNA-binding domain. Proteins endowed with more than one cupin domain, the bicupins and multicupins, are believed to originate from duplication and fusion events of a single monocupin precursor, although the fusion of cupin domains from two different ancestors may also have occurred for some (59-61). Enzymatic cupin members, for which a specific function has been annotated, are most commonly associated with some form of dioxygenase activity, typically oxidizing an organic substrate by the use of a dioxygen molecule (60). Most dioxygenases possess a metal cofactor in the active site, most often iron but other metals such as copper and nickel have also been observed. Typically, in bicupin dioxygenases only one of the two domains contains an active site. For instance, the copper-containing quercetin 2,3-dioxygenase (Q23D) catalyzes the oxidation of quercetin, a plant flavonol, to 2-protocatechuoylphloroglucinol carboxylic acid and carbon monoxide. In this reaction,

Q23D is only dependent on a metal ion and does not require any additional organic cofactors.

The YhaK Protein

In *E. coli*, YhaK is a cytosolic protein usually found in low abundance. To date, YhaK has no annotated function, but based on its amino acid sequence the Protein Family Database (Pfam) suggested it to be a member of the Pirin family, a family of bicupin proteins (62). Identified as a nuclear protein, human pirin (hPirin) is predominantly found in the tissues of heart and liver (63). Although the exact function of hPirin is not known, it has been observed to stabilize the formation of quaternary complexes composed of hPirin, the oncoprotein B-cell lymphoma 3-encoded (Bcl-3), the anti-apoptotic transcription factor NF-κB and the NF-κB DNA-binding site (60,64). As the resulting protein complex targets promoter regions of genes involved in apoptosis, the role of hPirin may involve regulating transcription and/or replication events (60,63,64). Likewise, the expression level of Le-pirin, a pirin homologue in tomato that shares 56% sequence identity with hPirin (60), has been reported to rise significantly during camptothecin-induced programmed cell death (PCD) (60,65). Shortly after the crystal structure of hPirin was determined (66), the structure of YhhW, another *E. coli* member of the Pirin family was also determined (67). As the structures of both these pirin members resembled the one of Q23D from *Aspergillus japonicus* (68), another bicupin, it was investigated whether hPirin and YhhW also could catalyze the same reaction. Interestingly, YhhW and hPirin were both demonstrated to possess quercetinase activity. However, YhaK exhibits a low sequence identity to both YhhW (32%) as well as hPirin (13%) and also lacks the residues important for both metal binding and enzymatic activity that are present in both YhhW and hPirin. Therefore, based on primary sequence, YhaK would not be expected to bind divalent metal ions in contrast to the other structurally determined pirins.

The Structure of YhaK (Paper III)

As the crystal structure of YhaK was determined, YhaK was observed to be in a monomeric state. As expected, YhaK was found to be composed of two cupin domains, arranged face-to-face, like the structures of hPirin and YhhW along with other bicupins (66,67). The N- and C-terminal cupin domains are both composed of two anti-parallel β-sheets with 4-6 strands forming a sandwich together with an adjacent short α-helix (Figure 10). The N-terminal domain (residues 1-134) is connected to the C-terminal domain (140-233) via a short linker of five amino acids (residues 135-139). In the centre of both domains similar cavities are found. The somewhat larger cavity of the N-terminal domain corresponds to the active site of both YhhW and hPirin. In the active sites of YhhW and hPirin, divalent metal ions are found to be coordinated by three histidines and a glutamate. These residues are part of the cupin Motif 1 and Motif 2 and are conserved in many of the enzymatic members of the cupins (60). In contrast, YhaK lack these metal-binding residues. Instead, Arg57, Tyr59, Tyr101 and Glu103 constitute the structurally equivalent residues. As such a

combination of amino acids generally do not coordinate metal ions, it was not surprising that no metal ion could be observed in the putative active site of YhaK. Still, the presence of several well-ordered water molecules in the cavity of the N-terminal domain may imply that this region might constitute an active site and/or binding site for a presently unidentified substrate or ligand. Furthermore, no signs of metals were observed in the slightly more compact C-terminal domain cavity.

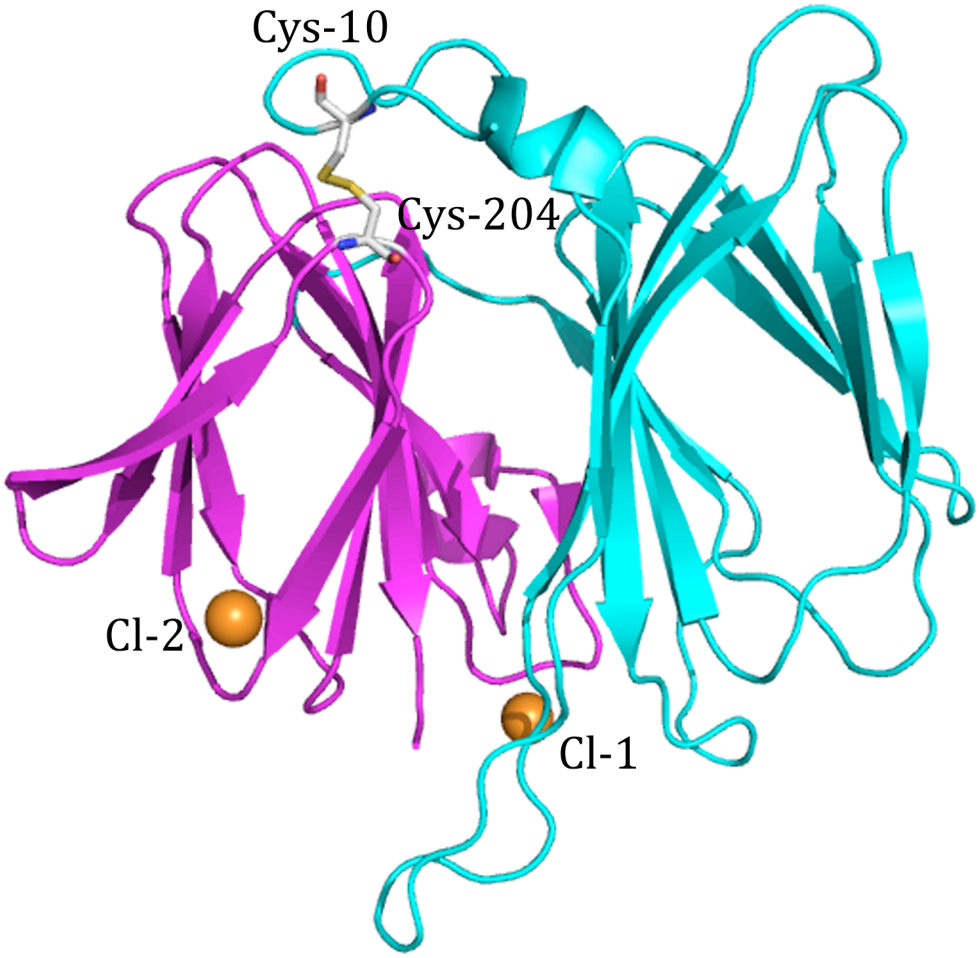


Figure 10 The overall structure of YhaK. The N-terminal domain is colored in cyan, and the C-terminal domain in magenta. The disulfide bond between Cys10 of the N-terminal domain and Cys204 of the C-terminal domain is shown in stick representation and the two modelled chloride ions are shown as sand colored spheres.

In the interface between the N- and C-terminal domains, most of the interactions are due to hydrophobic packing. In addition to the short linker, the two domains are also connected to one another via a disulfide bridge. This disulfide cross-link is formed between Cys10 from the N-terminal domain and Cys204 from the C-terminal domain. YhaK's third cysteine, Cys122, was also covalently modified to a sulfenic acid (Figure 11). Such a covalent modification suggests that Cys122 may be a reactive cysteine. In the crystal structure of YhaK, Cys122 is situated on the surface of the protein, in close proximity to a symmetry related molecule. An arginine, Arg125, is observed to pack against Cys122 and may function to activate the cysteine. Moreover, two ions were observed, which have been interpreted as chloride ions from the crystallization condition. The first chloride ion, Cl-1 interacts with the main-chain amides of Ala38 of

the N-terminal domain and Glu158 of the C-terminal domain, apparently serving as a bridge between the two domains. Interesting, this chloride ion is positioned in close proximity to modified Cys122 (5.3Å). The second chloride ion, Cl-2, is situated at the protein surface of YhaK, coordinated by the side-chain amine of Lys199 and to the main-chain amide of Arg211. In addition, the guanido groups of Arg180 and Arg211 of a symmetry related molecule also coordinate Cl-2. These structural features differ very much to what has been observed for hPirin and YhhW, and neither the cysteine residues nor the interdomain ions are observed in the YhhW or the hPirin structures.

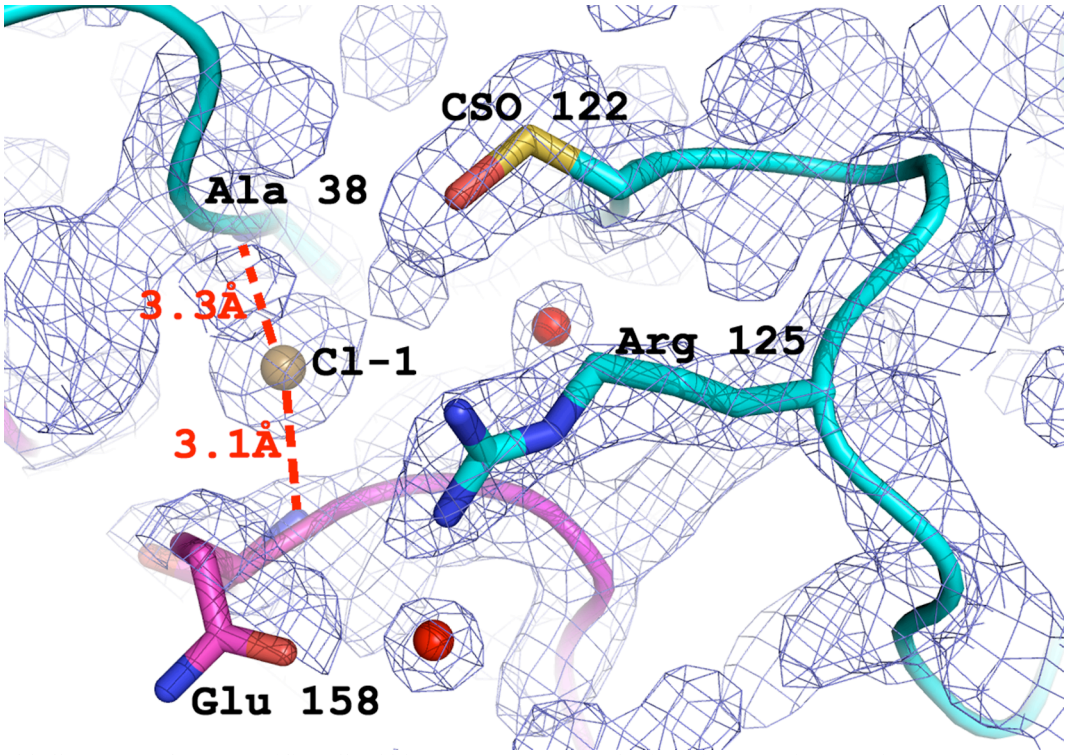


Figure 11 Close-up of the modified Cys122, modeled as a sulfenic acid, showing the $2F_{obs}$-F_{calc} electron density at 1σ. Cl-1 (sand-colored sphere) is situated close to Cys122 and the distances are shown in Å. The N-terminal domain is colored in cyan and the C-terminal domain in magenta. Two water molecules are shown as red spheres.

YhaK vs Quercetinases

As stated above, YhaK had previously been annotated to belong to the Pirin family. The determination of the YhaK structure led to that our expectations rose that its structure may shed some more light on the function of YhaK. Therefore, the YhaK structure was used to search for closely related protein structures in the PDB, using the DALI server (69). Again, the proteins with the highest structural similarity to YhaK turned out to be YhhW and hPirin. All of YhaKs structural relatives identified in the DALI search were bicupins, some with annotated functions and others without. However, most of the proteins with an annotated function did in fact display the metal binding motif, which is not present in YhaK. As mentioned before, the YhhW and hPirin members of the Pirin family have been shown to exhibit quercetinase activity (67). In addition to these Pirins and two additional Q23D's, Q23D from *A. japonicus* and YxaG from *B. subtilis*, other bicupins with annotated functions were composed of

5-keto 4-deoxyuronate isomerase from *E. coli*, oxalate decarboxylase from *B. subtilis*, seed storage proteins from plants, and mannose 6-phosphate isomerases.

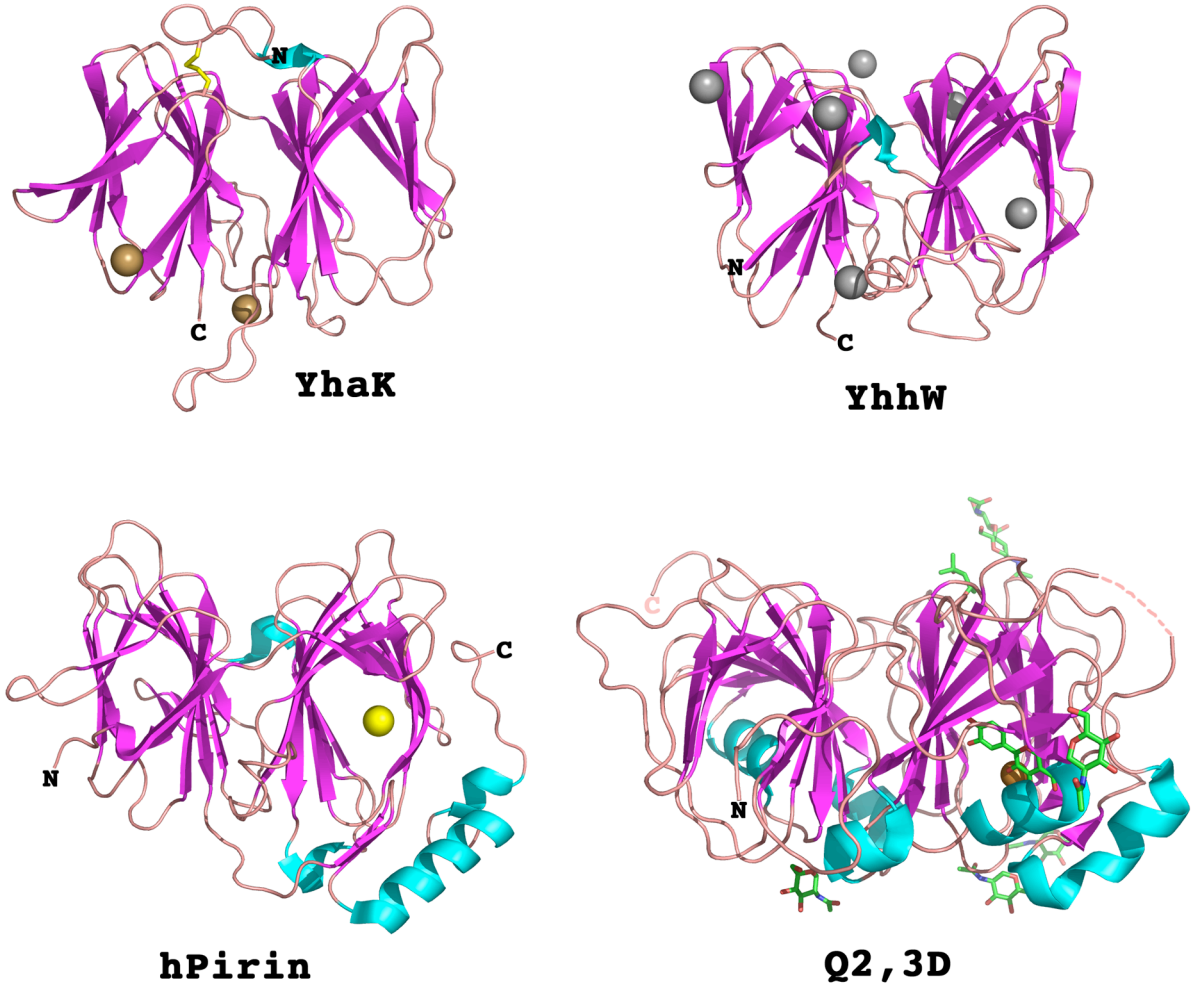


Figure 12 The overall structures of YhaK and three of its structural relatives. Top left: the monomeric structure of YhaK. Top right: the monomeric structure of YhhW (PDB ID 1TQ5). Bottom Left: the monomeric structure of hPirin (PDB ID 1J1L). Bottom Right: Molecule A of Q23D with bound quercetin (PDB ID 1H1I). The β-strands are colored in magenta, α-helices in cyan, and coiled regions in salmon. In YhaK, the two chloride ions are sand-colored and its disulfide bond is colored in yellow. In YhhW, Cadmium ions are colored in gray. In hPirin and Q23D, iron and cupper ions are colored in yellow and orange, respectively. The substrate of Q23D, quercetin, is colored in green.

Due to the Pfam identification of YhaK as a member of the Pirin family, in addition to the high degree of structural similarities between YhaK and YhhW, hPirin as well as other proteins exhibiting quercetinase activity (Figure 12), it was decided to investigate whether YhaK also possessed any enzymatic activity. In the quercetinase reaction, quercetin is converted to 2-2-protocatechuoylphloroglucinol carboxylic acid and carbon monoxide (68). Using the same enzymatic assay that was used for YhhW and hPirin (67), it was established that YhaK does not catalyze the quercetinase reaction. The result was not unexpected, as the quercetinase reaction is a metal-dependent event (68). The putative active site of YhaK, in contrast to YhhW, hPirin and the Quercetin 2,3-dioxygenases, lacks the residues required for metal binding of the conserved bicupin Motifs 1 and 2, and hence a divalent metal ion, which is critical for activity (Figure 13). Although it is evident that YhaK shares the same fold as these

enzymes, YhaK is unlikely to share the same enzymatic activities as other members in the Pirin family due to the differences in its putative active site.

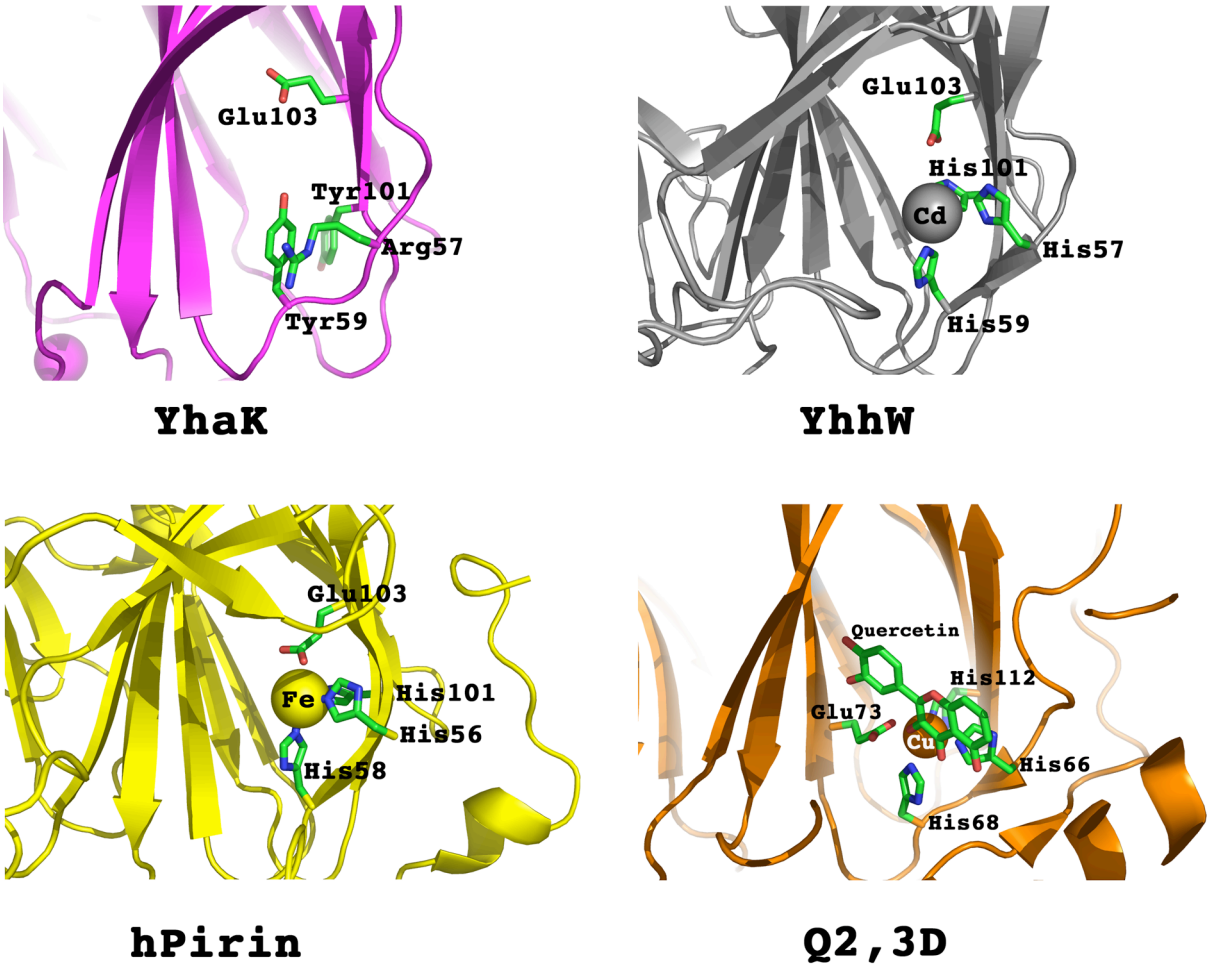


Figure 13 Close-up of the active sites of YhaK and three of its structurally relatives. Top left: The YhaK structure in the Q23D active site region. Top right: The YhhW active site (PDB ID 1TQ5). Bottom left: The active site of hPirin (PDB ID 1J1L). Bottom right: The active site of molecule A of Q23D with bound quercetin (PDB ID 1H1I). In the quercetinase enzymes, only residues involved in metal binding, as well as the equivalent residues in YhaK, are shown.

YhaK- A Redox-Sensitive Protein?

Although crystallization of the YhaK protein was carried out in excess of reductant, its crystal structure revealed the presence of three oxidized cysteines. In addition, the expression levels of YhaK were reported to be strongly upregulated when *E. coli* cells were exposed to nitroso-glutathione (GSNO) or sodium nitrite ($NaNO_2$) (70). Thus, the function of YhaK may be associated with redox regulation in *E. coli.* Accordingly, the redox state of YhaK was monitored under reduced and oxidized settings. Treatment with dithiothreitol (DTT), led to the detection of three free thiols per monomer, indicating that all of the three cysteines in YhaK could be reduced to form free thiols. Yet, reduced YhaK was observed to be very susceptible to oxidation, as only one thiol per monomer could be detected as the solution containing reduced YhaK was exposed to air for 40 min. Most likely, this is due to spontaneous formation of the disulfide bond between Cys10 and Cys204, which is observed in the crystal

structure. Interestingly, oxidation of YhaK coincided with the formation of oligomers, especially dimers or trimers and to some degree also tetramers (Figure 14). The exposure to hydrogen peroxide led to further oxidation of YhaK and additional formation of higher oligomers. Simultaneously, the amount of reduced YhaK in its monomeric form was observed to decrease, adding further support to that YhaK oligomerizes upon oxidation. However, the oligomerization that was observed for the YhaK protein may be a consequence of its protein concentration. When the same experiments were conducted at a lower protein concentration (5 μM versus 55 μM), YhaK was again reduced and oxidized by DTT and H_2O_2 respectively. In contrast to what was observed in the experiments conducted with the higher protein concentration (55 μM), YhaK was observed to remain primarily in a monomeric state in experiments carried out with a lower protein concentration (5 μM). This observation is in agreement with the monomeric form of YhaK in its crystal structure.

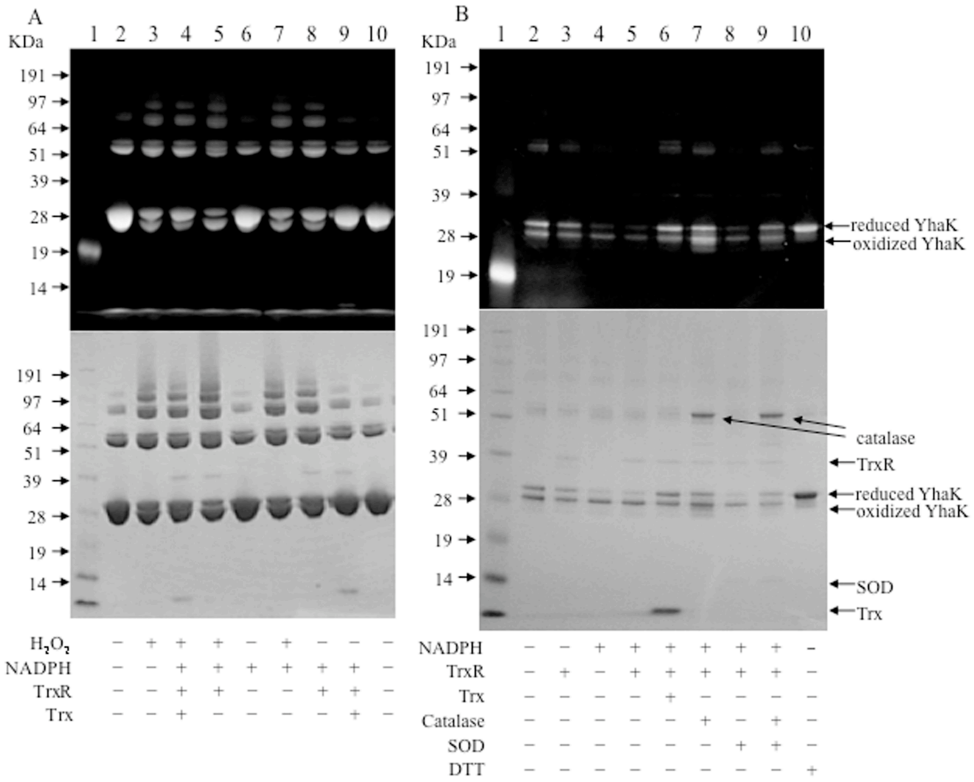


Figure 14 Effect of the *E. coli* Trx system on the redox change of YhaK (paper III). (A) 55 μM reduced YhaK was incubated with H_2O_2 (lane 3), NADPH, *E. coli* TrxR, *E. coli* Trx1, and H_2O_2 (lane 4); NADPH, *E. coli* TrxR, and H_2O_2 (lane 5); NADPH (lane 6); NADPH, and H_2O_2 (lane 7); NADPH, *E. coli* TrxR (lane 8); NADPH, *E. coli* TrxR, *E. coli* Trx 1 (lane 9) at room temperature for 40 min. The reduced YhaK exposed under air for 40 min was used as a control (lane 2 and 10). (B) 5 μM reduced YhaK was incubated with *E. coli* TrxR, (lane 3); NADPH (lane 4); NADPH, *E. coli* TrxR (lane 5); NADPH, *E. coli* TrxR, *E. coli* Trx1 (lane 6); NADPH, *E. coli* TrxR, catalase (lane 7); NADPH, *E. coli* TrxR, SOD (lane 8); NADPH, *E. coli* TrxR, catalase, and SOD (lane 9); DTT (Lane 10) at room temperature for 40 min. The reduced YhaK without any further DTT-treatment (as compared with the YhaK in Lane 10) was used as a control (lane 2). Lane 1 was the protein marker. The free cysteine in the protein was detected by alkylation with 5-IAF (top panel) and the protein was stained with Coomassie Brilliant Blue (bottom panel).

The finding that YhaK seemed to be susceptible to oxidative stress by hydrogen peroxide made us investigate whether the thioredoxin and glutathione-glutaredoxin systems would affect YhaK. In *E. coli* these two systems are the two protein disulfide reductase systems that are involved in numerous fundamental activities in the cell (71,72). The thioredoxin system is composed of NADPH, thioredoxin reductase (TrxR), and thioredoxin (Trx) whereas the glutathione-glutaredoxin system comprises NADPH, glutathione reductase (GR), glutathione (GSH), and glutaredoxin (Grx). Thus, in order to determine whether the Trx and the GSH/Grx systems could reduce YhaK, NADPH oxidation assays (73,74) were performed and analyzed by SDS-PAGE (Figure 15). The full Trx system was seemingly able to attenuate H_2O_2-induced oxidation of YhaK (Figure 14). Most interestingly, in the absence of Trx, TrxR together with NADPH was observed to increase the oxidation of YhaK. The addition of Trx decreased this oxidation. Thus, it seems that TrxR, in the absence of Trx, may be involved in the production of superoxides and reactive oxygen species (ROS). Again, discrepencies were observed between experiments conducted with different protein concentrations. At high concentrations (55 μM) of YhaK, NADPH alone did not appear to affect the redox state of YhaK in the absence of TrxR whereas at low protein concentrations of YhaK (5 μM) NADPH alone also oxidized the reduced form of YhaK. For both protein concentrations of YhaK, the addition of Trx resulted in the protection of YhaK against oxidation. The GSH/Grx system was also observed to reduce the oxidized form of YhaK (75). It was shown that reduced GSH was essential for the YhaK reduction and that GSH alone could to some degree protect YhaK against oxidative stress induced by hydrogen peroxide. However, the full GSH/Grx system resulted in a significant enhancement in the protection of YhaK against oxidative stress, demonstrated by both the NADPH oxidation assay and analysis of SDS-PAGE. In contrast to the oxidative effect induced by TrxR and NADPH, GR and NADPH was not observed to oxidize YhaK.

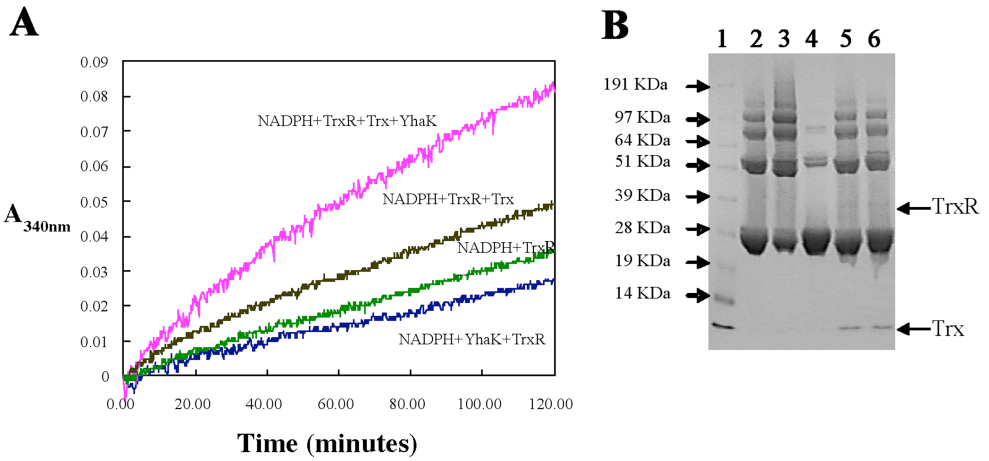


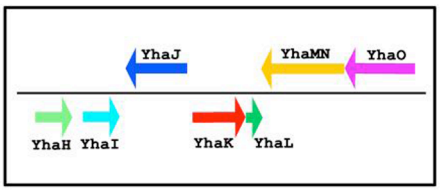
Figure 15 (A) The NADPH oxidation assay was used to test the effect of the Trx system on the oxidative state of YhaK. The consumption was enhanced by the addition of oxidized YhaK, providing additional evidence that the *E. coli* Trx system can reduce YhaK. (B) The Trx system did not reduce YhaK to the same extent as DTT (Lane 4) as the addition of DTT seemed to result in higher levels of monomeric YhaK.

YhaK- A Transcription Factor?

The outcome of the experiments conducted on the YhaK protein led us to propose that YhaK is not a member of the Pirin family, but instead is the first identified member in a new subclass of bicupins. The finding that YhaK does not display similar enzymatic activity to YhhW and hPirin, other members in the Pirin family, is supportive of this statement. The lack of quercetinase activity is most likely due to YhaK, in contrast to YhhW and hPirin, lacking residues required for metal binding and consequently does not bind a divalent metal ion. In addition, other structural features distinguish YhaK from YhhW and hPirin. The crystal structure of YhaK revealed the presence of three modified cysteine residues, two of which were involved in the formation of a disulfide bond and the third oxidized to a sulfenic acid. Experiments have demonstrated these cysteines to be responsive to alterations in the redox state of the surroundings. Interestingly, BLAST searches conducted with the YhaK sequence resulted in the identification of numerous enterobacterial sequences with 68% sequence identity or higher. In all of these all three cysteines were conserved and none contained any of the metal-binding residues of the cupin Motif1 and Motif2. In addition, the residues involved in the binding of Cl-1 near Cys122 were completely conserved, implying that this region may be of functional importance. Interestingly, Arg125, which was found to stack against the modified Cys122 in the crystal structure of YhaK, was also conserved. This structural feature resembles the active site motif of peroxiredoxins (76), although YhaK was shown not to possess peroxiredoxin activity (data not published).

Chemical modification of reactive cysteines might represent one of the means by which proteins are regulated, as the redox state of protein thiols may act as a molecular switch, activating or deactivating protein function. Various cysteine modifications such as S-nitrosylation, S-glutathionylation, formation of disulfides, and S-hydroxylation (oxidation to sulfenic acid) have been indicated to be utilized in the control of gene expression (77). For instance, OxyR, a LysR-type transcriptional cofactor (LTTR), is regulated by and dependent on the redox state of specific cysteine residues. Similar to YhaK, OxyR is responsive to hydrogen peroxide. OxyR switches from inactive to active conformation by the oxidation of two distant cysteines, resulting in a disulfide bond and subsequently tetramer formation followed by DNA binding. The function of YhaK may resemble the one found for OxyR. In fact, the *yhaK* gene is positioned back-to-back, in the opposite direction, to the *yhaJ* gene, which encodes a putative LYSR-type transcriptional regulator (LTTR) (Figure 16). In the cyanobacterium *Synechocystis sp. PCC 6803*, a similar genetic structure exists with *pirA*, another YhhW homologue, positioned back-to-back with *pirR*, another LTTR gene, in divergent directions (78). The expression of the *pirA* gene and an adjacent *pirB* gene, whose gene product is a putative DNA binding protein with two zinc finger motifs, are induced by high salt concentrations and other stress conditions and negatively regulated by PirR. These findings may provide very interesting leads regarding the physiological function of YhaK, especially as YhaK and YhaJ have been observed to form protein-protein interactions (79). YhaK and YhaJ may form a protein-protein complex that able to bind DNA as YhaJ contains a helix-turn-helix (HTH) motif, a common feature of the LTTRs (80), in response to changes in the

redox state and/or salinity of the surroundings. Although speculative, such a scenario could explain the existence of a chloride ion near the reactive Cys122 in the crystal structure of YhaK. In addition, the role as a transcriptional cofactor, in which YhaK forms a protein complex with a DNA binding protein, is in agreement with the formation of a quaternary complex composed of hPirin, NF-κB, BCL-3 and the NF-κB DNA binding site (60,64).



	MG1655 Gene Identifier[#]	Gene Name	UniProt accession number	Predicted function of hypothetical protein
1	b3103	YhaH	P64590	Cell inner membrane; multi-pass membrane protein.[a]
2	b3104	YhaI	P64592	Cell inner membrane; multi-pass membrane protein.
3	b3105	YhaJ	P67660	Putative HTH-type (LysR) transcriptional regulator.
4	b3106	YhaK	P42624	Yhak is a low abundance protein. yhhW and yhaK paralogs, and homologs of human Pirin nuclear protein. Transcriptional induction by 1mM $NaNO_2$ (25-fold) and 1mM GSNO (6-fold).
5	b3107	YhaL	P42625	No function annotated.
6	b4470	YhaM/N	P42626	Cell membrane; multi-pass membrane protein (Potential). Upregulated by GSNO in defined medium (aerobic).
7	b3110	YhaO	P42628	Cell inner membrane; multi-pass membrane protein. Belongs to the amino acid/polyamine transporter 2 family. SdaC/tdcC subfamily. Upregulated by GSNO in defined medium.

[#] Ecogene at http://ecogene.org
[a] UniProt at http://www.pir.uniprot.org/

Figure 16 The *yhaHIJKLMNO* gene cluster. Upper panel: Direction of genes in the cluster. Botton panel: Present knowledge of functions of gene products in the cluster.

The Herpesvirus Project

The main intention of the Herpesvirus project was to utilize a structural genomics (high-throughput) approach to produce, crystallize, and eventually determine the molecular structure of as many target proteins as possible where structures could explain protein function. In the initial expression screening, all Open Reading Frames (ORFs) of five different herpesviruses, Herpes simplex virus-1 (HSV-1), Varicella zoster virus (VZV), Kaposi's sarcoma-associated herpesvirus (KSHV), Epstein-Barr virus (EBV), and murine Cytomegalovirus (mCMV), made up the target list. However, due to practical reasons an initial target list was composed of 65 targets from all of these five herpesviruses. The elucidation of the molecular structure of many of the targets may be of important medical relevance and may eventually lead to structure-based drug design against herpesviral infections. Herpesviruses constitute a large family of DNA viruses that are known to cause disease in humans as well as other vertebrates (81). The most exceptional trait of these viruses is the ability to establish and reactivate from latency (82,83). Many herpesviruses exist in the latent state during most of their lifecycle, characterized by the expression of a limited number of viral genes and the lack of infectious viral particles in the tissues or cells containing the virus (82,84). Thus, all herpesviruses may be present in one out of two states, lytic or latent. The family members may be categorized into one out of three subfamilies, α-, β-, and γ-herpesviruses, based on differences in host range, length of life cycle, and target cells (85). Presently, eight members of the known herpesviruses have been recognized to be disease causing in humans, representing all three subfamilies. Members of the α-herpesvirus subfamily are neurotropic and comprise varicella zoster virus (VZV) and herpes simplex virus 1 and 2 (HSV-1 and HSV-2). The members of the γ-herpesvirus subfamily, Epstein-Barr virus (EBV) and Kaposi's sarcoma associated herpesvirus (KSHV), are lymphotropic, whereas the site of latency still has not been established for the β-herpesviruses, human cytomegalovirus (CMV) and human herpesvirus 6 and 7 (HHV-6 and HHV-7). All members of the herpesviridae contain the same structural elements of the viral particle (86). The double-stranded DNA genome is comprised within a capsid, composed of an icosahedral protein arrangement. The tegument, an amorphous layer composed of several proteins, surround the capsid. In turn, the tegument is encapsulated by the outermost structural element of the virion, the envelope. The envelope consists of a lipid bilayer membrane typically embedded with glycosylated proteins. The herpesvirus replication is initiated as the viral envelope adsorbs to the plasma membrane of the host cell, a process mediated by viral glycoproteins and proteoglycans and specific surface receptors of the host (83). The consequential fusion is followed by the internalization and dismantling of the virion as the capsid and the elements of the tegument are released in the cytoplasm. Subsequently, the viral DNA, in a circularized form, is transferred into the cell nucleus where both viral replication and transcription occur. In the lytic state, the expression of many herpesvirus genes results in that the viral DNA is capable to replicate independent of the host's cell cycle (84). In the latent state, on the other hand, the herpesvirus may only replicate in tandem with the chromosomal DNA of the host. The ability of the members in the herpesvirus family to switch between latent and lytic

states may be critical in their evasion of the host immune responses, a main feature of the herpesviruses (87).

KSHV and EBV are oncoviruses, as they are able to cause lymphoproliferation and cancers (88). EBV, the first oncovirus to be discovered, is the causal agent of several cancer forms such as Burkitt and Hodgkin lymphoma (89). However, although almost all adults are infected by EBV, the infection is in most cases not associated with tumor formation (88). The ability of latently infected cells to avoid immune responses is one of the factors that influence the carcinogenic effect of a viral infection. In addition, it has been established that parallel infections, toxic compounds, and the genetic predisposition of the host may be important aspects. Although Kaposi's sarcoma (KS) was known already in the late 19^{th} century, it was not until 1994 that KSHV, a previously unidentified herpesvirus, was identified as a causative agent (90). In addition, KSHV has also been associated with other human malignancies such as pleural effusion lymphoma (PEL) and multicentric Castleman's disease (MCD) (91). All DNA viruses, such as herpesviruses, are dependent on that the building blocks for their genome, the deoxyribonucleotides (dNTPs), are available in sufficient supply to ensure their replication (92). In the host cell, the level of the dNTP pool varies depending on cell type and the stage of the cell cycle and reaches a maximum during cell division. However, in quiescent and terminally differentiated cells, the concentration of intracellular dNTPs is very low. As most cells in an adult organism have undergone differentiation and remain in such a dormant state, DNA viruses have been forced to apply various strategies to ensure adequate amounts of dNTPs. Large DNA viruses, such as some of the herpesviruses, have adopted a strategy to circumvent the problem. Due to the large size of their genome, these viruses are able to encode functional homologues of cellular enzymes involved in nucleotide metabolism. Such a strategy has resulted in that herpesviruses rely less on the host's enzymatic machinery and are able to replicate regardless of the state of host cell. The α- and γ-herpesviruses appear to have adopted this strategy and encode functional homologues of cellular enzymes involved in nucleotide metabolism. In contrast, the β-herpesviruses seem to lack many genes required for biosynthesis of the DNA precursors. Instead, β-herpesviruses seem to be able to force dormant cells to express cellular nucleotide anabolic enzymes to ensure their requirement for the dNTPs in resting cells. Nonetheless, the study of the herpesviral DNA metabolism is of great importance in the development of antiviral drugs.

Alkaline Exonucleases of the Herpesviruses

All herpesviruses encode a member of the alkaline exonuclease (AE) family, which has been annotated to belong to the PD-(D/E)XK superfamily of Mg^{2+}-dependent nucleases (93,94). Structural studies of members of type II endonucleases revealed that, despite a low degree of sequence similarity, certain protein domains were conserved (93). A common feature of the members of the PD-(D/E)XK superfamily is the formation of a structural core, composed of a mixed β-sheet comprising four or five strands flanked by α-helices (93,94). This structural feature serves as scaffold for the formation of an active site for metal binding and catalysis, by spatially bringing

together two or three acidic residues, Asp or Glu, and a Lysine which constitute part of the signature residues of the PD-(D/E)XK superfamily. The lack of sequence similarity is further complicated by the fact that some members in this superfamily have developed different variants of the active site (93). For instance, the acidic residues, which constitute signature residues for members of this superfamily, have been noticed to be exchangeable with Asn or Gln residues. In other variants, these residues have been observed to "migrate" to other regions in the sequence while maintaining their spatial orientation. These aspects result in that the identification of new members solely based on sequence is hampered and often requires the determination of their tertiary structure. Furthermore, in many cases, the fold of the core itself may display variations, as insertions and terminal extensions may result in that additional subdomains are formed. These structural variations may result in different preferences regarding oligomerization and substrate specificities (93,94). Thus, although enzymes in this diverse superfamily, which often are involved in DNA recombination and repair events, vary in their substrate specificities, the prerequisite for catalysis seem to have been conserved (95). Constituting a subfamily in the PD-(D/E)XK superfamily, the AE family members have been found to be close to exclusively present in viruses (94). In herpesviruses, the AEs appear to be essential in the maturation and packaging of the viral genome into the viral capsids. The role of the AEs has been suggested to be involved in resolving branched genomic structures, a critical event in the formation of new infectious virions. In the amino acid sequences of the herpesviral AEs, seven conserved motifs have been found. The members of the AE family have been shown to exhibit 5'-3' exonuclease activity *in vitro*, even though endonuclease and 3'-5' exonuclease activities also have been observed to occur *in vitro* (94,96-98). However, despite the ability to act upon both single- and double-stranded substrates *in vitro*, the real substrate *in vivo* is still not established. For their activity, the AEs display an absolute requirement for a divalent cation, preferably a Mg^{2+} ion, although Mn^{2+} also functions to some extent (96-99). The enzymes also exhibit an alkaline pH optimum, hence the name, as well as a sensitivity to high salt concentrations (96,97,99).

Host Shutoff of the Herpesviruses

Viral infection is often accompanied by a global shutoff of the host cellular protein synthesis (100). For many viruses, the inhibition of cellular gene expression is crucial in their life cycle. Host shutoff not only allows maximum viral gene expression, but may also result in an evasion of the host immune defense, which may be vital for viral replication. In herpesviruses, host shutoff is an intrinsic feature of the α- and γ-subfamilies. However, although inhibition of cellular protein synthesis on the mRNA-level is a shared attribute of both the α- and γ-herpesviruses, the two subfamilies utilize different approaches to attain host shutoff (100). In the α-herpesviruses, the virion host shutoff (VHS) protein, a tegument protein encoded by the UL41 gene in HSV, is primarily responsible for host shutoff (100-102). VHS has been shown to be a RNase, degrading cellular as well as viral mRNA, with a preference for cleaving its substrate near the 5' end where the translation is initiated (100,103). Although VHS appears to be a functional RNase on its own, it may depend on cellular factor(s) for proper mRNA targeting. In mammalian cells, VHS has been observed to interact with

a eukaryotic initiation factor (eIF4H), which normally serves as a RNase helicase accessory factor in cellular translation (100). In addition, co-expression in bacteria of VHS and eIF4H resulted in the formation of a protein complex (103). Such a protein complex may be of significant relevance in the targeting of VHS to the initiation regions of translation of mRNA. The effect of the VHS protein is further enhanced by the activity of the immediate-early ICP27 protein, which inhibits splicing of pre-mRNA (103-105). As HSV transcripts, in contrast to cellular transcripts, are generally not spliced, the actions of ICP27 result in that cellular mRNA ready for translation is further decreased (104,105). As observed in the α-herpesviruses, host shutoff of cellular proteins in the γ-subfamily also takes place at the mRNA level (106,107). It has been demonstrated that lytic replication of KSHV and EBV is accompanied by a shutoff of host gene expression by the actions of ORF37 and BGLF5 respectively. Interestingly, ORF37 and BGLF5 are not VHS homologues, but are members of the AE family. The additional shutoff function of the AEs in the γ-herpesvirus subfamily was first discovered for KSHV ORF37, which subsequently was named **s**hut**o**ff and e**x**onuclease (SOX) protein (106).

The Bi-functional Shutoff and Exonuclease protein

In contrast to VHS, ORF37 and BGLF5 are not part of the tegument of the infecting virion, but are expressed during the early lytic phase (107). As already mentioned, in herpesviruses, members of the AE family are involved in processing of the viral genome. However, the AEs of the γ-subfamily have also been demonstrated, *in vivo*, to be responsible for host shutoff, which occurs at the mRNA level in the early lytic phase (106,107). Mutation studies of KSHV SOX and EBV BGLF5 have revealed that the DNase and shutoff functions of these bifunctional proteins are genetically separable and that these two activities also may take place in different cellular compartments (108,109). The degradation of cellular mRNAs may take place in the cytoplasm while the DNase activity occurs in the nucleus, in contrast to UL12, the α-herpesvirus AE, which is strictly localized to the nucleus (108). It was also shown that mutations that eliminated only the shutoff function were located outside the seven conserved sequence motifs, whereas mutations affecting both or merely the DNase function were located within these motifs. These findings indicate the additional shutoff function of the γ-herpesvirus AEs has evolved more recently. Most interestingly, in contrast to the other herpesviral AEs, the shutoff activities of SOX and BGLF5 have been implicated to cause down regulation of human leukocyte antigen (HLA) molecules, which may reduce T cell recognition (107,109).

The Structure of KSHV SOX (Paper IV)

In the crystal structure, SOX was found to be a monomer. Two domains make up the SOX protein, an N-terminal domain, containing residues 7-164 as well as 446-481, and the larger C-terminal domain, containing residues 179-445 (Figure 17). The N-terminal domain is composed of 10 α-helices varying in length whereas a 5-stranded β-sheet with five flanking α-helices constitute the C-terminal domain, which is also denoted as the core domain, In some areas, the structure suffers from disorder and for

that reason residues 1-7, 165-178, 315-318 and 482-487 have been excluded in the final model. Together, the two domains endow the SOX structure with a bowl-shaped appearance in which a crevice forms in between the two domains. The two domains are connected via a loop region, designated the "bridge", which starts with Gln154 and ends with Gly180. However, as mentioned above, the majority of the bridge, residues 165-178, has been left out in the structure due to disorder. Some parts of this large crevice are positively charged which may be indicative of where DNA binding occurs. In addition, two sulfates, sulfate 1 (S1) and sulfate 2 (S2), were found to bind in the central region of the crevice, and a third, sulfate 3 (S3), at the edge of the crevice constituted by the core domain. In the bottom of the crevice, the conserved catalytic residues, Asp221 and Glu244, are situated in close proximity to the sulfates, S1 and S2. Asp221 and Glu244 corresponds to the active site carboxylate residues in the PD-(D/E)XK motif of the type II nucleases. Together with the main chain oxygen of Ile245 and a water molecule, Asp221 and Glu244 coordinate an active site magnesium ion. In order to find structural relatives to SOX, a DALI search against PDB was conducted. During this search we found that the closest structural relative of SOX is the monomer of the bacteriophage λ-exonuclease (PDB ID 1AVQ) (110). In addition to the λ-exonuclease, other members of the type II PD-(D/E)XK nuclease superfamily were also identified as structural homologues. Despite the fact that SOX is noticeably larger than the λ-exonuclease and that the amino acid sequences of the two proteins do not show any significant sequence identity, the structures of SOX and λ-exonuclease can be successfully superimposed upon each other (Figure 18A). The superposition reveals that the two proteins are structurally very similar both in the overall structure as well as in the active site. However, while the active site of the SOX structure is characterized by disordered loops, the corresponding region in the λ-exonuclease structure appears to have fairly ordered secondary structural elements. Still, the magnesium ion coordinating residues in the SOX structure can be superimposed very well on top of the residues that coordinate a manganese ion in the λ-exonuclease structure (110). In addition, the position of a phosphate ion in the λ-exonuclease structure correlates well to S1 found at active site crevice in SOX. The phosphate and S1 interact with both main chain and side chain atoms from five residues that appear to be completely or partly conserved in SOX and the λ-exonuclease (Figure 18B). Thus, the positions of the sulfate/phosphate ions may point out where the terminal phosphate of the nucleic acid substrate binds. In addition to having a larger N-terminal domain, the SOX structure also has a different oligomeric state than the λ-exonuclease. The structure of the λ-exonuclease revealed a trimer while SOX was determined as a monomer. On the other hand, the trimeric state of the λ-exonuclease has been suggested to be the active form of the protein playing an important role in the processivity of the enzyme. Shortly before our results were published, a low-resolution structure of the BGLF5 protein from EBV was published (111). The two structures are very similar with minor variations in some loop regions as well as in some secondary structure elements of the N-terminal domain.

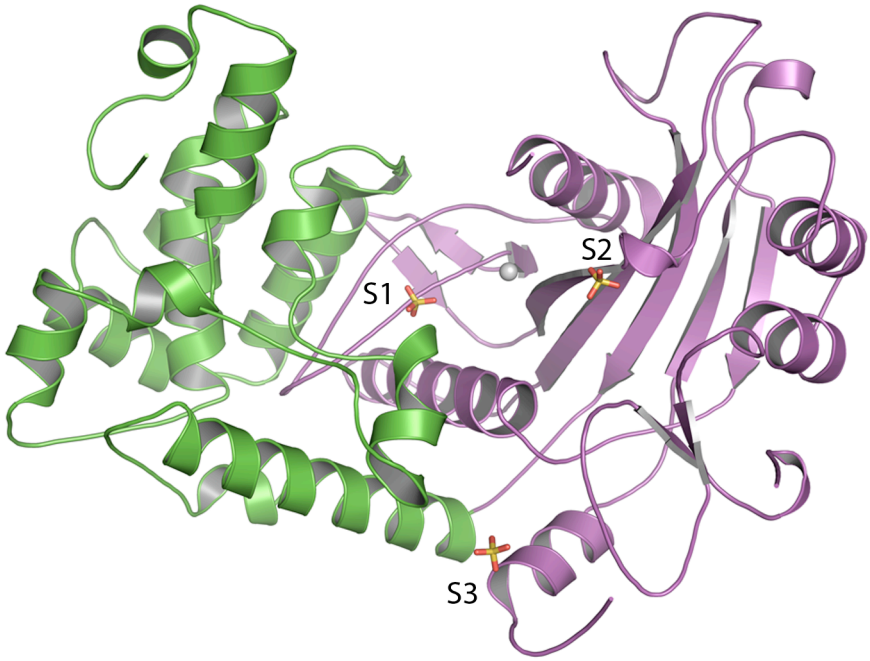


Figure 17 The overall structure of SOX. The N-terminal domain is colored in green and the core domain in purple. The three sulfate ions, designated S1-S3, are shown in stick representation, and the magnesium ion is shown as a gray sphere.

A **B**

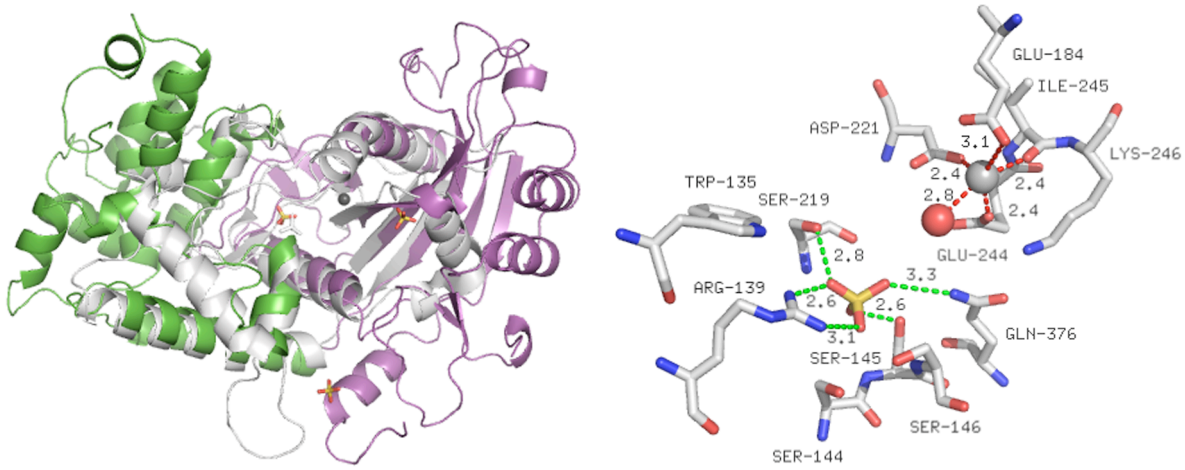


Figure 18 (A) Superimposition of SOX (green and purple) and the λ-exonuclease monomer (PDB ID 1AVQ). (B) A stick representation of active site residues in SOX that are involved in the coordination of sulfate 1 and and the magnesium ion. The magnesium ion is shown as a gray sphere and in (B), a coordinated water molecule is shown as a red sphere.

SOX Mechanism in DNA Processing

As previously mentioned, members of the AE family have been shown to hydrolyse both single- and double-stranded DNA *in vitro*, the true substrate(s) *in vivo* is still not known. Thus, in an attempt to elucidate how SOX interacts with the nucleic acid, the positions of the sulfates were used as both single- and double-stranded nucleic acids were modelled in the active site crevice. The positions of two of the sulfates, S1 and S2, were used as reference points in an attempt to model a short single-stranded

oligonucleotide in the active site crevice of SOX. These attempts resulted in the finding that SOX is likely to accommodate 4 nucleotides in between the binding sites of S1 and S2. As shown in Figure 19A, the positions of S1 and S2 correspond to the binding sites of the phosphates of the oligonucleotide, P1 and P4. In addition, two other putative phosphate binding sites, P2 and P3, were identified. The position of P2 in close proximity to the Mg^{2+} site makes it a likely candidate for positioning the phosphodiester bond at the site of hydrolysis. The hydrolysis may occur via a direct attack from an activated water molecule on the phosphodiester bond as has been suggested for other type II nucleases (112,113). These parts of the cavity region of SOX display a high degree of sequence conservation in all herpes AEs (Figure 19A). However, the size of the active site crevice indicates that SOX may also be able to interact directly with double-stranded DNA. In fact, the modelling of double-stranded DNA 11-mer using the same reference points, S1 and S2, implies that SOX is able to encompass a double-stranded DNA substrate into its large cavity. Interestingly, the modelling also suggests that the binding site of the third sulfate, S3, may be involved in the interaction with the nucleic acid.

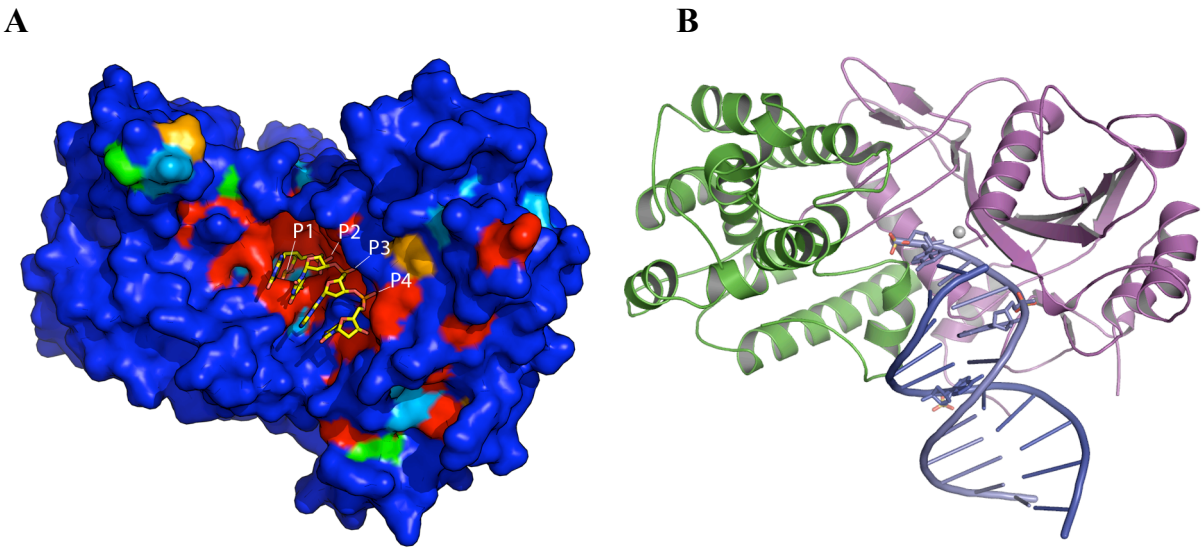


Figure 19 Modeling of DNA binding to SOX. (A) A conserved residue map, based on sequence alignments with the other human herpesvirus AEs. Completely conserved residues are colored red and nonconserved residues are colored dark blue. The degree of conservation spans from red to orange to green to light blue to dark blue. A short oligonucleotide (4 bp) was modelled in the highly conserved crevice on top of the positions of sulfate 1 and sulfate 2. The phosphates in the DNA backbone of this oligonucleotide are labelled P1-P4, following the 5'-3' direction. (B) Modeling of duplex DNA binding to SOX, based on the positions of all three sulfates. SOX is viewed in a different orientation compared to (A) for clarity. In similarity to what is seen for the single-stranded substrate in (A), sulfate 1 and sulfate 2 align with P1 and P4 of one of the strands (light blue) and sulfate 3 aligns with a phosphate situated further downstream in the complementary strand (dark blue).

In an attempt to provide further clues regarding the binding mode of nucleic acids to SOX, ligand-induced thermal shifts were measured using a fluorescence-based thermal melt assay in the presence of different non-hydrolysable oligonucleotides, both single- and double-stranded, (Table 1). For all oligonucleotides that were examined, significant thermal shifts could be measured. Interestingly, a single-stranded oligonucleotide with a phosphate at the 5'-position gave a significantly higher shift than oligonucleotide with a phosphate at its 3'-position, in agreement with that SOX

exhibits a stronger 5'-3'-exonuclease activity (114,115). Furthermore, double-stranded oligonucleotides with a phosphate at the 5'-position but with different modes of base-pairing also gave rise to various significant shifts. Although it cannot be ruled out that the slight differences in length of the different oligonucleotides may influence the results, it was interesting to notice that the most significant shift was that of a fully complementary 14-mer duplex, which was significantly higher than the shifts observed for duplexes with a overhang or partial non-complementary base-pairing. Nevertheless, the data suggest that SOX is able to bind a double-stranded DNA substrate and that the binding of the 5'-phosphate is of importance, in agreement with the position of the 5'-phosphate at the P1 pocket in our model. Interestingly, all of the seven conserved motifs of the herpesvirus AEs were located in the active site or near regions suggested to be involved in DNA binding.

Complex	**Oligonucleotide**	**TM**	**SD (TM)**
SOX without oligonucleotide	None	50.0	0.2
SOX + Duplex with overhang: 9 complementary bp and 5 single stranded overhang bp	5'PAATTCGATCCCGGT-3' 3'P-GCTAGGGCCA-5'	55.4	0.3
SOX + Non-complementary Duplex: 9 complementary bp and 5 noncomplementary bp	5'PAGTTAGATCTTAAA-3' / CTAGAATTT-5' 3'-CATGA/	53.4	0.3
SOX +Complementary Duplex: 14 complementary bp	5'-TGAG CGATCCGAGT-3' 3'-ACTCGCTAGGCTCA-P5'	58.9	0.3
SOX + Single stranded DNA with 3'P	5'-ACCGGGATCG-3'P	52.9	0.3
SOX + Single stranded DNA with 5'P	5'P-ACTCGGATCGCTCA-3'	54.5	0.5

Table 1 Results from the fluorescence-based thermal melt assay using different non-hydrolysable oligonucleotides, both single- and double-stranded. TM: Thermal melting point. SD: Standard deviation.

Mapping Mutations onto the SOX Structure

Prior to our structural studies, two rounds of mutagenesis studies of SOX had been conducted in order to characterize this bifunctional protein (106,108). The first study

was carried out using a somewhat focused strategy (106) whereas the second employed random mutagenesis (108). The effects of various amino acid substitutions were analysed using an *in vitro* DNase assay and an *in vivo* shutoff assay. These studies identified both double function mutants, in which both functions were affected (106), and single function mutants, in which either the DNase or the shutoff activity was affected (108). The double function mutations that affected both functions correspond to a Gly217/Ala substitution and a Asp221/Glu substitution (106). A substitution of Gln129 to a histidine residue was found to result in a complete loss of DNase activity while not affecting the shutoff activity (108). Several amino acid substitutions were observed to affect the shutoff activity: Thr24/Ile, Ala61/Thr, Pro176/Ser, Val369/Ile, Asp474/Asn and Tyr477/Stop codon. Interestingly, as these substitutions were mapped on the SOX structure it was noticed that the mutations that affect the DNase function are located in the active site region whereas mutations only affecting the shutoff function are mostly found close to the protein surface of the N-terminal domain (Figure 20). Still, two of the amino acid substitutions, Gly217/Ala and Asp221/Glu, that affect the shutoff, as well as the DNase activity, are located in the active site region of SOX. An interesting finding as Asp221 is observed to coordinate the magnesium ion in the SOX structure.

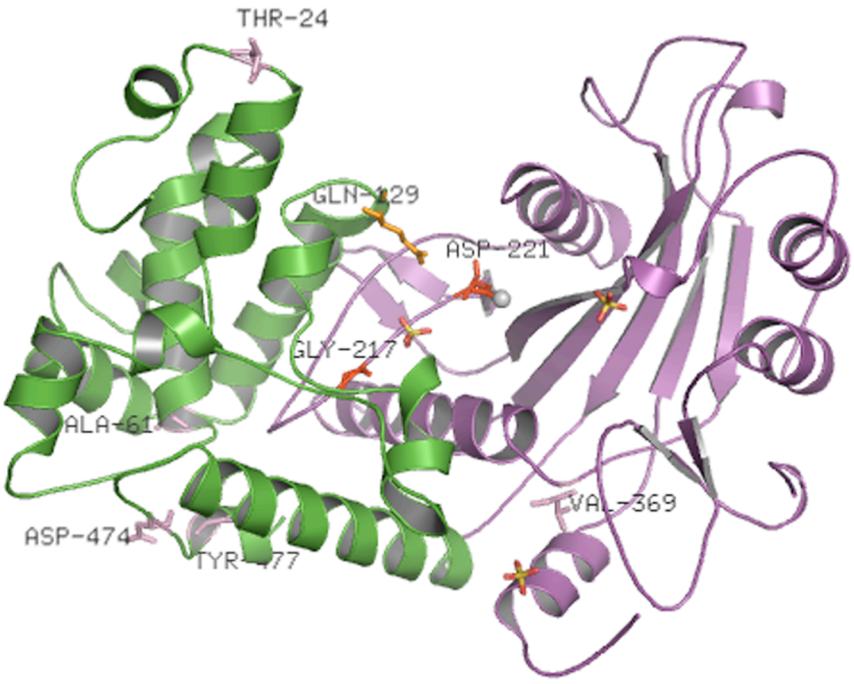


Figure 20 Substitutions that abolish DNase and shutoff activities have been mapped in the crystal structure of SOX. Double-function substitutions, which abolish both DNase and shutoff activities, are colored in red. A single-function substitution that only affects DNase activity is shown in orange, while single-function substitutions that only abolish the shutoff activity of SOX are colored in pink.

Shutoff Function of the Bi-functional SOX Protein

The structural work conducted in this study still leaves some question marks regarding how the SOX protein mediates the host shutoff. For instance, SOX may bind and degrade mRNA directly. Alternatively, as the shutoff activity only has been observed

in vivo, SOX-mediated mRNA degradation may be dependent on additional binding partners, viral or cellular. It may even be so that SOX mediates host shutoff indirectly by activating an intrinsic mRNA degradation pathway. The mapping of the single- and double function substitutions on the SOX structure gives some interesting clues to SOX role in host shutoff. For instance, a direct function of SOX as an RNase is supported by the positions of the double function substitutions in the active site region, which indicate that the shutoff activity is dependent on the same active site as the DNase activity. In addition, the predominant location of the single function substitutions, which affect the shutoff activity, in the N-terminal domain implies that this domain may be involved in making essential interactions with RNA and/or an unidentified protein, resulting in the direct degradation of RNA or the activation of a mRNA degradation pathway. As mentioned before, shortly before our results were published, a low-resolution structure of the BGLF5 protein from EBV was published (111). In this study, the authors found that the SOX homologue exhibited RNase activity in the presence of manganese ions. It would be of great interest to test SOX in the same manner, but this has not yet been done.

ORF60 of KSHV- The R2 Subunit of the Ribonucleotide Reductase

The enzyme ribonucleotide reductase (RNR) is present in cells from all life forms as well as in some large DNA viruses, such as herpesviruses (116). This allosterically regulated enzyme is responsible for the reduction of all four ribonucleotides to their corresponding deoxyribonucleotides (dNTPs) (Figure 21), the building blocks of DNA. Thus, the activity of this enzyme regulates the cellular levels of the dNTP pool, ensuring that accurate DNA replication and repair may occur. Based on the source of the radical cofactor used in the reductive reaction, RNRs are divided into three classes, I, II and III. Class I is further divided into subclasses Ia and Ib. Class I RNRs are composed of a heterotetramer, which in turn is composed of two homodimers of the R1 and R2 subunits. The R1 subunit contains the active site as well as the sites for allosteric regulation. The R2 subunit contains a dinuclear iron center as well as a stable tyrosyl radical per monomer. The di-iron site generates a free radical, typically located at a tyrosine residue, by the reductive cleavage of molecular oxygen. The radical is subsequently transferred to the R1 subunit activating the nucleotide substrate for catalysis. In eukaryotes the activity of the RNR enzyme is cell cycle dependent, and reaches its highest levels during the S phase, when DNA synthesis and replication occur (117). The regulation of the two subunits differs and the R2 subunit is rate limiting for the activity of the RNR enzyme. The regulation of the R2 subunit occurs on both the transcriptional level and by protein degradation. The R2 gene is not transcribed in resting cells. However, certain events, such as DNA repair and synthesis of mitochondrial DNA, require that the synthesis of deoxyribonucleotides also occur in resting cells. Thus, in order to synthesise dNTPs in response to DNA damage, quiescent cells are able to synthesize another cellular R2, p53R2, which together with R1 form a functional RNR complex (116,117). The name reflects the fact that it is controlled/regulated by the transcription factor and tumor suppressor p53. In addition, viral replication in infected cells may also trigger the production of dNTPs. Interestingly, members of the α- and the γ-herpesviruses encode two class Ia subunits,

R1 and R2, of the RNR enzyme, while members of the β-subfamily only encode an inactive R1 subunit (92). The presence of a fully functional RNR in the α- and the γ-herpesviruses, although the host already encodes a functional enzyme of its own, indicates that the virus may have evolved to ensure a sufficient supply of nucleotides even if the host is in a quiescent state. The regulation of the viral and host RNR may also differ. The oncovirus KSHV, together with the Epstein-Barr virus (EBV), belongs to the γ-subfamilies of human herpesviruses. The R2 subunit of herpesviruses, encoded by the ORF60 gene in KSHV, shares a low sequence identity to other R2s of both bacterial and eukaryotic origin (117). Still, no structure of any R2 from herpesviruses has been published. An interesting feature of ORF60 is that a glutamate, Glu64, is found in the position usually occupied by a metal-coordinating aspartate. Studies of this enzyme may be of significant importance as it may serve as a potential drug target for the treatment of herpesviral infections and its associated cancer forms.

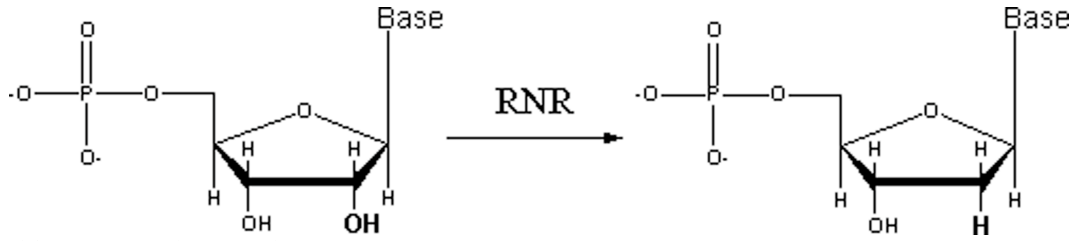


Figure 21 The reductive reaction catalyzed by RNR.

Crystallization and Structure Determination of ORF60 (Paper V)

Initial crystallization trials were unsuccessful in producing any crystals of the ORF60 protein. Thus, in an attempt to promote crystal formation, *in situ* proteolysis was undertaken in which a small amount of chymotrypsin was added (7). This resulted in that crystal platelets were formed (Figure 22A). However, crystallization optimization failed to produce three-dimensional crystals suitable for structural determination. The crystallization was further improved by the addition of hexamine cobalt (III) chloride as an additive, which induced the formation of three-dimensional crystals suitable for diffraction experiments (Figure 22B).

The structure of ORF60 has been determined but the structure is still undergoing model building and structure refinement. Still, its crystal structure reveals a homodimer in which each monomer is composed of an α-helix bundle similar to other R2s but with some interesting features in the metal site (Figure 23). Based on the electron density, there seems to be no iron atoms present at the dinuclear iron center. Instead it is more likely to be occupied by two water molecules, since no electron density is present above 3 σ. However, the short distances and the geometry between the metal coordinating residues and one of these putative water molecules do suggest metal binding. These factors together with the fact that magnesium ions were present in the crystallization condition make us suggest that a Mg^{2+} ion occupies at least one of the di-iron metal sites. In addition, there seems to be no clear electron density for the putative metal coordinating Glu64 (see below). Further experiments will be required to determine the identity of the atoms in the di-nuclear metal site in ORF60.

A B

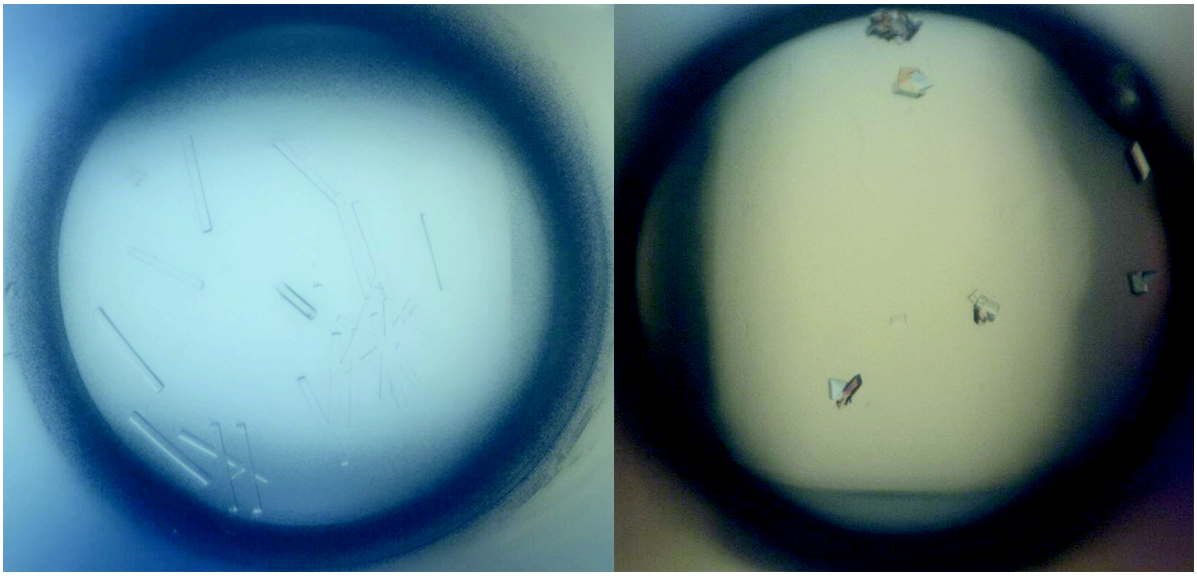


Figure 22 Crystals of ORF60 from KSHV. A) Crystal plates were formed as a result of *in situ* proteolysis with chymotrypsin as the proteolytic agent. B) The addition of (Hexamine) cobalt (III) chloride resulted in the formation of more three-dimensional crystals.

A B

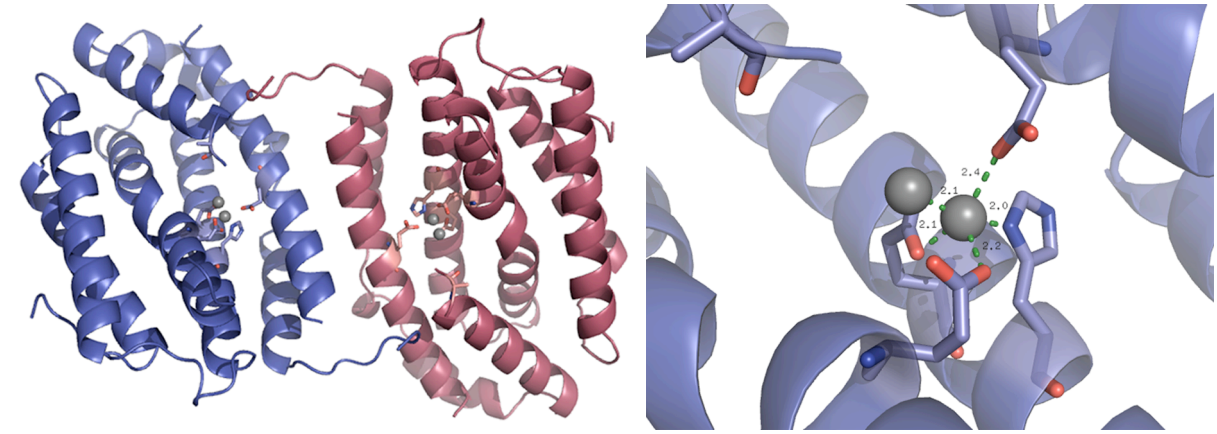


Figure 23 (A) Overall structure of the homodimeric ORF60, the RNR R2 subunit from KSHV. The different monomers are colored in blue and red. Metal coordinating residues are shown in stick representation and putative magnesium ions shown as gray spheres. (B) Close-up of the metal binding site of one of the monomers. Again, metal coordinating residues are shown in stick representation and putative magnesium ions shown as gray spheres.

In both subunits, the residues 64-81 and 110-142 are disordered, which results in the di-iron metal site being unusually accessible for the solvent. Interestingly, the first disordered residue in the first missing segment of both subunits corresponds to Glu64, which is supposed to be a metal-coordinating residue of the di-iron site. Whether or not the observed disorder is due to proteolysis during crystallization or if it is an intrinsic feature of the ORF60 protein remains to be seen. At the moment, we are not able to give an answer to this subject. However, in the structures of human p53R2 (PDB ID: 2VUX and 3HF1) (118) as well as in other R2 structures (*Plasmodium vivax* (PDB ID: 2O1Z), and *Saccharomyces cerevisiae* (PDB ID: 1JK0)) (119) similar disorder and/or displacement of the structural equivalents of metal-coordinating Glu64

can also be observed. At present, it is not known whether such unstructured regions of the ORF60 protein may be of any physiological significance.

Future Prospects

The work described in this thesis has resulted in structural determination of five different proteins from two projects, the *E. coli* project and the herpesvirus project. However, for both projects, further efforts may prove to be necessary to clarify their functional role. Structural determination of the tertiary structure of a protein molecule may reveal a great deal about its molecular function. However, the elucidation of the quaternary structure of a protein complex may in some cases be necessary for a comprehensive understanding of a protein's mechanism. Thus, for the *E. coli* project, it would be of great interest to elucidate the protein-protein interactions that have been observed to occur between YhaK and YhaJ. For instance, the determination of a putative YhaK-YhaJ protein complex would be of great value in the functional annotation of the YhaK protein. Formation of such a protein complex may indicate that YhaK plays the role of a transcription factor in which DNA binding may be facilitated through the HTH motif of YhaJ. In addition, in the functional annotation of YhaK it would be interesting to determine its physiological response, alone or in a putative complex with YhaJ, to cellular changes in the redox environment and/or salt (i.e chloride or other halides) concentrations.

In the herpesvirus project, the ORF60 structure will be completed and analyzed. It would be of great interest to determine if the structural disorder that was observed around the metal binding site of both R2 monomers is an intrinsic feature of this protein or a consequence of adding a protease during crystallization. In addition, many targets from both the *E.coli* and herpesvirus projects remain to be structurally characterized. The pursuit of these interesting, but not trivial, targets will require a significant effort. The *in situ* proteolysis approach that was used in crystallization of ORF60 may be helpful in this endeavour. Furthermore, the quaternary structure of SOX in complex with different non-hydrolysable oligonucleotides would shed some light on the substrate specificity of this bifunctional protein. Initially, the aim has been focused on the determination of the structure of the viral proteins alone, but eventually the focus should shift towards determining complex structures, composed of two or more viral protein partners as well as between viral and human protein partners. Such complexes will most likely have biological implications as well as being of medical relevance, since they may elucidate important interactions between host and pathogen.

Acknowledgements

Most often, this is the hardest part of the thesis to write. Hopefully I will remember to thank all of the many persons that have helped during these years.

First of all I would like to thank my supervisors, Heidi and Pär. Heidi, Your support and guidance have been greatly appreciated during all these years. Thanks for always being so positive and helpful. Pär, thanks for providing for all the equipment and resources a PhD student can ask for and also for sharing your knowledge. I have learnt a lot during these years.

I would like to thank all the present and former members of the PN lab: Albert, Amin, Agnes, Anders, Benita, Karl-Magnus, Maria, Marie, Tobias, Martin Hö, Martin Hä, Jessica, Victoria, Monica, Andreas, Herwig, Pelle, Hanna, Henrik, Elisabeth, Marina, Ulrika, Audur, Malin, Mikaela, Cedric, Christian, Christine, Rebecka, Esben, and Kerstin.

Said, many thanks for all the help whenever needed and for all fun discussions. Good luck in Singapore.

Damian, thanks for being a good colleague, office companion and former neighbor as well as a good friend both inside and outside the lab. You always seem to have the time to help out, and I appreciate it. I have the highest expectations for you.

Daniel MM, thanks for teaching me that the inside counts as well. Seriously, I appreciate your eagerness to help others and for making the lab more enjoyable.

Lola, thanks for always being positive and happy. You have a strong spirit that will take you a long way.

Sue-Li, thanks for a good collaboration on the Herpesvirus project. It was an interesting journey. It was fun to work together with you.

Pål, thanks for a good collaboration on the *E. coli* project and for a good "examensarbete". It was a fun and successful experience.

Martin Moche, thanks for a good collaboration on the EcRibD project and for that you always take your time to help out. It is appreciated.

Karin and Anna-Karin, there was always a positive and cheerful spirit in our office.

All the people on SGC Stockholm, especially Martin, Helena, Tomas and Susanne.

The administrators on KI and DBB, especially Maria and Ann at DBB. Thank you for always helping me out when I am in trouble.

Stefan at DBB, thanks for being a good support during my time as a PhD student.

Mina vänner (nya såväl som gamla): Renato och Sunil, för intressant och trevligt umgänge, Micke R, vi har känt varandra sedan barnsben och det är roligt att det går bra för dig, Kristian, Gidden, för att du alltid ställer upp, Eli, roligt att återförenas, Pelle, ett gott hjärta i en sund kropp, Anders R, alltid lika roligt och inspirerande att träffa dig, Christer och Marina, tack för all hjälp och alla trevliga stunder, Joakim, tack för all hjälp och för trevligt samarbete.

To the Houston family: James and David with families, aunt Mary, I am truly happy for your progress and I wish you all the best and hope to see you soon again.

Moscoso Torres-familjen samt Smith familjen: Lennie and Dave and Will, Guille + Barbara, Lotta och Stephanie + Peter. Rosa och Memo, tack för all er hjälp och stöd (samt piroger) under alla dessa år.

Två vänner som sticker ut: Micke B och Maxi, tack för er vänskap och för trevligt umgänge både på och utanför mattan trots alla tempoväxlingar.

Kärstin, tack för all ditt stöd och för vad du har gjort under alla år, goda såväl som svåra. Jag uppskattar allt du gör.

Min bror Ambo med familj, Helen och Jordan. Tack ska ni ha för all hjälp och stöd under dessa år. Min bror och tredje förälder, tack för att du alltid ställer upp och för din osjälviska brödrakärlek.

Mina föräldrar, tack för allt ni har gjort för mig både som barn och som vuxen. Jag inser att jag aldrig kommer att kunna återgälda detta till er, men hoppas att jag kan betala tillbaka genom att försöka göra detsamma för mina barn. Pappa, du har lärt mig att kämpa och har alltid varit en inspirationskälla. Mamma, du har alltid stått bakom mig och trott på mig. Tillsammans med Martha, Geneth och Esther är ni mina största "fans".

Mina hustru och mina döttrar: Martha, min evigt unga och vackra hustru och min stora kärlek, tack för att du alltid lyfter upp mig när jag behöver det och för att du tar ned mig på jorden när det behövs. Det har varit en lång resa och jag ser fram emot nästa fas. Geneth, min förstfödda, livet förändrades totalt till det bättre när du kom in i vårt liv. Jag är enormt stolt över dig och lovar att jag kommer vara en mycket roligare pappa efter detta. Esther, blåöga, du har vänt upp och ned på allting. Du får mig alltid att skratta och bli på gott humör. Solen går upp och ned med er alla tre.

References

1. Lehninger, A. L., Nelson, D. L., and Cox, M. M. (2005) *Lehninger principles of biochemistry*, 4th ed., W.H. Freeman, New York
2. Brändén, C.-I., and Tooze, J. (1999) *Introduction to protein structure*, 2nd ed., Garland Pub., New York
3. Vitkup, D., Melamud, E., Moult, J., and Sander, C. (2001) *Nature Structural Biology* **8**, 559-566
4. Burley, S. K. (2000) *Nature Structural Biology* **7**, 932-934
5. Graslund, S., Nordlund, P., Weigelt, J., Bray, J., Hallberg, B. M., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., Dhe-Paganon, S., Park, H. W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S. H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C. Y., Hung, L. W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D. Y., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., Gunsalus, K. C., Consortium, S. G., Macromol, A. F., Ctr, B. S. G., Consortium, C. S. G., Function, I. C. S., Ctr, I. S. P., Genomics, J. C. S., Genomics, M. C. S., Ctr, N. Y. S. G. R., Consortium, N. S. G., Facility, O. P. P., Facility, P. S. P., Med, M. D. C. M., Proteomics, R. S. G., and Complexes, S. (2008) *Nature Methods* **5**, 135-146
6. Cohen, S. L. (1996) *Structure* **4**, 1013-1016
7. Dong, A. P., Xu, X. H., and Edwards, A. M. (2007) *Nature Methods* **4**, 1019-1021
8. Quevillon-Cheruel, S., Leulliot, N., Gentils, L., van Tilbeurgh, H., and Poupon, A. (2007) *Current Protein & Peptide Science* **8**, 151-160
9. Sanchez, R., and Sali, A. (1997) *Current Opinion in Structural Biology* **7**, 206-214
10. Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N., and Tawfik, D. S. (2009) *Trends in Biochemical Sciences* **34**, 53-59
11. Madigan, M. T., Martinko, J. M., and Parker, J. (2000) *Brock biology of microorganisms*, 9th ed., Prentice Hall, Upper Saddle River, NJ
12. Kalland, K. H., Ke, X. S., and Oyan, A. M. (2009) *Apmis* **117**, 382-399
13. Durbin, S. D., and Feher, G. (1996) *Annual Review of Physical Chemistry* **47**, 171-204
14. Mori, H. (2004) *Journal of Biochemistry and Molecular Biology* **37**, 83-92
15. Matte, A., Jia, Z., Sunita, S., Sivaraman, J., and Cygler, M. (2007) *Journal of Structural and Functional Genomics* **8**, 45-55
16. Frishman, D. (2003) *OMICS: A Journal of Integrative Biology* **7**, 211-224
17. Bremer, J. (1983) *Physiological Reviews* **63**, 1420-1480
18. Bieber, L. L. (1988) *Annual Review of Biochemistry* **57**, 261-283
19. McGarry, J. D., and Brown, N. F. (1997) *European Journal of Biochemistry* **244**, 1-14
20. Vaz, F. M., and Wanders, R. J. A. (2002) *Biochemical Journal* **361**, 417-429
21. Rebouche, C. J., and Seim, H. (1998) *Annual Review of Nutrition* **18**, 39-61
22. Kleber, H. P. (1997) *FEMS Microbiology Letters* **147**, 1-9
23. Eichler, K., Bourgis, F., Buchet, A., Kleber, H. P., and Mandrandberthelot, M. A. (1994) *Molecular Microbiology* **13**, 775-786

24. Engemann, C., Elssner, T., Pfeifer, S., Krumbholz, C., Maier, T., and Kleber, H. P. (2005) *Archives of Microbiology* **183**, 176-189
25. Jung, H., Buchholz, M., Clausen, J., Nietschke, M., Revermann, A., Schmid, R., and Jung, K. (2002) *Journal of Biological Chemistry* **277**, 39251-39258
26. Bernal, V., Arense, P., Blatz, V., Mandrand-Berthelot, M. A., Canovas, M., and Iborra, J. L. (2008) *Journal of Applied Microbiology* **105**, 42-50
27. Eichler, K., Buchet, A., Lemke, R., Kleber, H. P., and MandrandBerthelot, M. A. (1996) *Journal of Bacteriology* **178**, 1248-1257
28. Buchet, A., Nasser, W., Eichler, K., and Mandrand-Berthelot, M. A. (1999) *Molecular Microbiology* **34**, 562-575
29. Walt, A., and Kahn, M. L. (2002) *Journal of Bacteriology* **184**, 4044-4047
30. Heider, J. (2001) *FEBS Letters* **509**, 345-349
31. Selmer, T., and Buckel, W. (1999) *Journal of Biological Chemistry* **274**, 20772-20778
32. Ricagno, S., Jonsson, S., Richards, N., and Lindqvist, Y. (2003) *EMBO Journal* **22**, 3210-3219
33. Gogos, A., Gorman, J., and Shapiro, L. (2004) *Acta Crystallographica Section D-Biological Crystallography* **60**, 507-511
34. Toyota, C. G., Berthold, C. L., Gruez, A., Jonsson, S., Lindqvist, Y., Cambillau, C., and Richards, N. G. J. (2008) *Journal of Bacteriology* **190**, 2556-2564
35. Gruez, A., Roig-Zamboni, V., Valencia, C., Campanacci, V. R., and Cambillau, C. (2003) *Journal of Biological Chemistry* **278**, 34582-34586
36. Canovas, M., Maiquez, J. R., Obon, J. M., and Iborra, J. L. (2002) *Biotechnology and Bioengineering* **77**, 764-775
37. Obon, J. M., Maiquez, J. R., Canovas, M., Kleber, H. P., and Iborra, J. L. (1999) *Applied Microbiology and Biotechnology* **51**, 760-764
38. Bernal, V., Sevilla, A., Canovas, M., and Iborra, J. L. (2007) *Microbial Cell Factories* **6**, 1-17
39. Jonsson, S., Ricagno, S., Lindqvist, Y., and Richards, N. G. J. (2004) *Journal of Biological Chemistry* **279**, 36003-36012
40. Rangarajan, E. S., Li, Y. G., Iannuzzi, P., Cygler, M., and Matte, A. (2005) *Biochemistry* **44**, 5728-5738
41. Berthold, C. L., Toyota, C. G., Richards, N. G. J., and Lindqvist, Y. (2008) *Journal of Biological Chemistry* **283**, 6519-6529
42. Bacher, A., Eberhardt, S., Eisenreich, W., Fischer, M., Herz, S., Illarionov, B., Kis, K., and Richter, G. (2001) *Vitamins and Hormones - Advances in Research and Applications, Vol 61* **61**, 1-49
43. Massey, V. (2000) *Biochemical Society Transactions* **28**, 283-296
44. Gerdes, S. Y., Scholle, M. D., D'Souza, M., Bernal, A., Baev, M. V., Farrell, M., Kurnasov, O. V., Daugherty, M. D., Mseeh, F., Polanuyer, B. M., Campbell, J. W., Anantha, S., Shatalin, K. Y., Chowdhury, S. A. K., Fonstein, M. Y., and Osterman, A. L. (2002) *Journal of Bacteriology* **184**, 4555-4572
45. Bacher, A., Eberhardt, S., Fischer, M., Kis, K., and Richter, G. (2000) *Annual Review of Nutrition* **20**, 153-167
46. Hollander, I., and Brown, G. M. (1979) *Biochemical and Biophysical Research Communications* **89**, 759-763
47. Graupner, M., Xu, H. M., and White, R. H. (2002) *Journal of Bacteriology* **184**, 1952-1957
48. Burrows, R. B., and Brown, G. M. (1978) *Journal of Bacteriology* **136**, 657-667

49. Fischer, M., Romisch, W., Saller, S., Illarionov, B., Richter, G., Rohdich, F., Eisenreich, W., and Bacher, A. (2004) *Journal of Biological Chemistry* **279**, 36299-36308
50. Richter, G., Fischer, M., Krieger, C., Eberhardt, S., Luttgen, H., Gerstenschlager, I., and Bacher, A. (1997) *Journal of Bacteriology* **179**, 2022-2028
51. Keller, P. J., Levan, Q., Kim, S. U., Bown, D. H., Chen, H. C., Kohnle, A., Bacher, A., and Floss, H. G. (1988) *Biochemistry* **27**, 1117-1120
52. Chen, S. C., Chang, Y. C., Lin, C. H., Lin, C. H., and Liaw, S. H. (2006) *Journal of Biological Chemistry* **281**, 7605-7613
53. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *Journal of Molecular Biology* **247**, 536-540
54. Sawaya, M. R., and Kraut, J. (1997) *Biochemistry* **36**, 586-603
55. Schnell, J. R., Dyson, H. J., and Wright, P. E. (2004) *Annual Review of Biophysics and Biomolecular Structure* **33**, 119-140
56. Chatwell, L., Krojer, T., Fidler, A., Romisch, W., Eisenreich, W., Bacher, A., Huber, R., and Fischer, M. (2006) *Journal of Molecular Biology* **359**, 1334-1351
57. Chen, S. C., Lin, Y. H., Yu, H. C., and Liaw, S. H. (2009) *Journal of Biological Chemistry* **284**, 1725-1731
58. Magalhaes, M. L. B., Argyrou, A., Cahill, S. M., and Blanchard, J. S. (2008) *Biochemistry* **47**, 6499-6507
59. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R., and Goodenough, P. W. (2001) *Trends in Biochemical Sciences* **26**, 740-746
60. Dunwell, J. M., Purvis, A., and Khuri, S. (2004) *Phytochemistry* **65**, 7-17
61. Dunwell, J. M., Khuri, S., and Gane, P. J. (2000) *Microbiology and Molecular Biology Reviews* **64**, 153-+
62. Sonnhammer, E. L. L., Eddy, S. R., and Durbin, R. (1997) *Proteins-Structure Function and Genetics* **28**, 405-420
63. Wendler, W. M. F., Kremmer, E., Forster, R., and Winnacker, E. L. (1997) *Journal of Biological Chemistry* **272**, 8482-8489
64. Dechend, R., Hirano, F., Lehmann, K., Heissmeyer, V., Ansieau, S., Wulczyn, F. G., Scheidereit, C., and Leutz, A. (1999) *Oncogene* **18**, 3316-3323
65. Orzaez, D., de Jong, A. J., and Woltering, E. J. (2001) *Plant Molecular Biology* **46**, 459-468
66. Pang, H., Bartlam, M., Zeng, Q. H., Miyatake, H., Hisano, T., Miki, K., Wong, L. L., Gao, G. F., and Rao, Z. H. (2004) *Journal of Biological Chemistry* **279**, 1491-1498
67. Adams, M., and Jia, Z. C. (2005) *Journal of Biological Chemistry* **280**, 28675-28682
68. Fusetti, F., Schroter, K. H., Steiner, R. A., van Noort, P. I., Pijning, T., Rozeboom, H. J., Kalk, K. H., Egmond, M. R., and Dijkstra, B. W. (2002) *Structure* **10**, 259-268
69. Holm, L., and Sander, C. (1997) *Nucleic Acids Research* **25**, 231-234
70. Mukhopadhyay, P., Zheng, M., Bedzyk, L. A., LaRossa, R. A., and Storz, G. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101**, 745-750
71. Holmgren, A. (1989) *Journal of Biological Chemistry* **264**, 13963-13966
72. Holmgren, A. (1985) *Annual Review of Biochemistry* **54**, 237-271
73. Lundstrom, J., and Holmgren, A. (1990) *Journal of Biological Chemistry* **265**, 9114-9120
74. Holmgren, A. (1984) *Methods in Enzymology* **107**, 295-300
75. Gurmu, D., Lu, J., Johnson, K. A., Nordlund, P., Holmgren, A., and Erlandsen, H. (2009) *Proteins-Structure Function and Bioinformatics* **74**, 18-31

76. Wood, Z. A., Schroder, E., Harris, J. R., and Poole, L. B. (2003) *Trends in Biochemical Sciences* **28**, 32-40
77. Filomeni, G., Rotilio, G., and Ciriolo, M. R. (2005) *Cell Death and Differentiation* **12**, 1555-1563
78. Hihara, Y., Muramatsu, M., Nakamura, K., and Sonoike, K. (2004) *FEBS Letters* **574**, 101-105
79. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) *Genome Research* **16**, 686-691
80. Maddocks, S. E., and Oyston, P. C. F. (2008) *Microbiology* **154**, 3609-3623
81. Boehmer, P. E., and Nimonkar, A. V. (2003) *IUBMB Life* **55**, 13-22
82. Croen, K. D. (1991) *Annual Review of Medicine* **42**, 61-67
83. Mossman, K. L. (2002) *Viral Immunology* **15**, 3-15
84. Bowie, A. G., Zhan, J., and Marshal, W. L. (2004) *Journal of Cellular Biochemistry* **91**, 1099-1108
85. Nicholas, J. (2000) *Journal of Clinical Pathology-Molecular Pathology* **53**, 222-237
86. McGeoch, D. J., Rixon, F. J., and Davison, A. J. (2006) *Virus Research* **117**, 90-104
87. DavisPoynter, N. J., and Farrell, H. E. (1996) *Immunology and Cell Biology* **74**, 513-522
88. Conti, M. (2007) *Cancer and Metastasis Reviews* **26**, 215-220
89. Beltran, B., Salas, R., Quinones, P., Morales, D., Hurtado, F., Cotrina, E., Riva, L., and Castillo, J. (2009) *Infect Agent Cancer* **4:10**, (Open Access Article http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2713202/)
90. Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M., and Moore, P. S. (1994) *Science* **266**, 1865-1869
91. Verma, S. C., and Robertson, E. S. (2003) *FEMS Microbiology Letters* **222**, 155-163
92. Lembo, D., and Brune, W. (2009) *Trends in Biochemical Sciences* **34**, 25-32
93. Kosinski, J., Feder, M., and Bujnicki, J. M. (2005) *BMC Bioinformatics* **6:172**, (Open Access Article http://www.biomedcentral.com/1471-2105/1476/1172)
94. Bujnicki, J. M., and Rychlewski, L. (2001) *Virus Genes* **22**, 219-230
95. Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L., and Ginalski, K. (2007) *BMC Structural Biology* **7:40**, (Open Access Article http://www.biomedcentral.com/1472-6807/1477/1440)
96. Bronstein, J. C., and Weber, P. C. (1996) *Journal of Virology* **70**, 2008-2013
97. Stolzenberg, M. C., and Ooka, T. (1990) *Journal of Virology* **64**, 96-104
98. Hoffmann, P. J., and Cheng, Y. C. (1979) *Journal of Virology* **32**, 449-457
99. Hoffmann, P. J., and Cheng, Y. C. (1978) *Journal of Biological Chemistry* **253**, 3557-3562
100. Glaunsinger, B. A., and Ganem, D. E. (2006) *Advances in Virus Research* **66**, 337-394
101. Kwong, A. D., and Frenkel, N. (1989) *Journal of Virology* **63**, 4834-4839
102. Kwong, A. D., Kruper, J. A., and Frenkel, N. (1988) *Journal of Virology* **62**, 912-921
103. Everly, D. N., Feng, P. H., Mian, I. S., and Read, G. S. (2002) *Journal of Virology* **76**, 8560-8571
104. Hardwicke, M. A., and Sandri-Goldin, R. M. (1994) *Journal of Virology* **68**, 4797-4810
105. Hardy, W. R., and Sandrigoldin, R. M. (1994) *Journal of Virology* **68**, 7790-7799
106. Glaunsinger, B., and Ganem, D. (2004) *Molecular Cell* **13**, 713-723

107. Rowe, M., Glaunsinger, B., van Leeuwen, D., Zuo, J. M., Sweetman, D., Ganem, D., Middeldorp, J., Wiertz, E. J. H. J., and Ressing, M. E. (2007) *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3366-3371
108. Glaunsinger, B., Chavez, L., and Ganem, D. (2005) *Journal of Virology* **79**, 7396-7401
109. Zuo, J., Thomas, W., van Leeuwen, D., Middeldorp, J. M., Wiertz, E. J. H. J., Ressing, M. E., and Rowe, M. (2008) *Journal of Virology* **82**, 2385-2393
110. Kovall, R., and Matthews, B. W. (1997) *Science* **277**, 1824-1827
111. Buisson, M., Geoui, T., Flot, D., Tarbouriech, N., Ressing, M. E., Wiertz, E. J., and Burmeister, W. P. (2009) *Journal of Molecular Biology* **391**, 717-728
112. Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005) *Cellular and Molecular Life Sciences* **62**, 685-707
113. Pingoud, A., and Jeltsch, A. (2001) *Nucleic Acids Research* **29**, 3705-3727
114. Sheaffer, A. K., Weinheimer, S. P., and Tenney, D. J. (1997) *Journal of General Virology* **78**, 2953-2961
115. Hoffmann, P. J., and Cheng, Y. C. (1978) *Journal of Biological Chemistry* **253**, 3557-3562
116. Eklund, H., Uhlin, U., Farnegardh, M., Logan, D. T., and Nordlund, P. (2001) *Progress in Biophysics & Molecular Biology* **77**, 177-268
117. Nordlund, N., and Reichard, P. (2006) *Annual Review of Biochemistry* **75**, 681-706
118. Smith, P., Zhou, B. S., Ho, N., Yuan, Y. C., Su, L., Tsai, S. C., and Yen, Y. (2009) *Biochemistry* **48**, 11134-11141
119. Sommerhalter, M., Voegtli, W. C., Perlstein, D. L., Ge, J., Stubbe, J., and Rosenzweig, A. C. (2004) *Biochemistry* **43**, 7736-7742