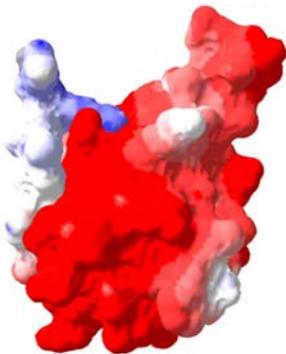
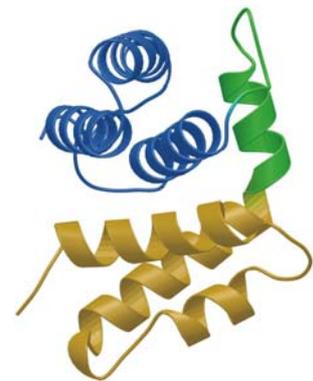
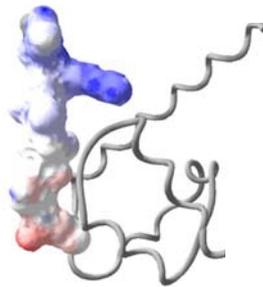


**Structural Analysis of Quinoline 2-Oxidoreductase from
Pseudomonas putida 86**

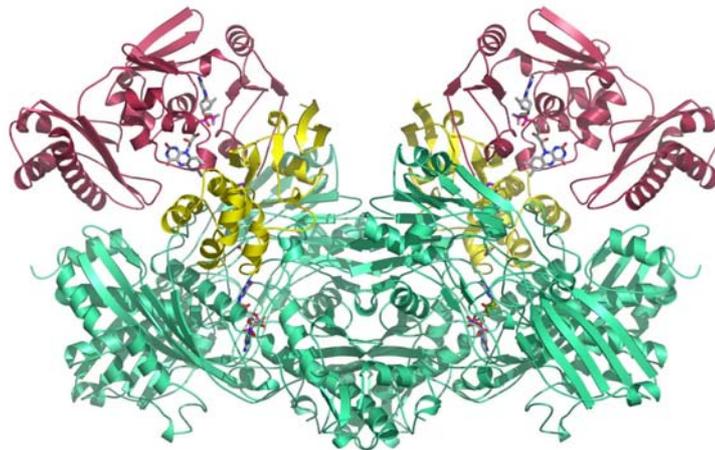
**Structural and Biochemical Studies of Two Nus Family Proteins:
NusB and NusA AR1- λ N Complex**



NusA AR1- λ N Complex



NusB



Quinoline 2-Oxidoreductase

Irena Bonin

Max-Planck-Institut für Biochemie

Abteilung Strukturforschung

D-82152 Martinsried, München

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

**Structural Analysis of Quinoline 2-Oxidoreductase from
Pseudomonas putida 86**

**Structural and Biochemical Studies of Two Nus Family Proteins:
NusB and NusA AR1- λ N Complex**

Irena Bonin

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität
München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr. St. J. Glaser

Prüfer der Dissertation:

1. apl. Prof. Dr. Dr. h.c. R. Huber

2. Univ.-Prof. Dr. Dr. A. Bacher

Die Dissertation wurde am 22.09.2004 bei der Technischen Universität München eingereicht
und durch die Fakultät für Chemie am 11.11.2004 angenommen.

Ai miei genitori

Acknowledgments

The work, here presented, was carried out in the Abteilung für Strukturforschung at the Max-Planck-Institut für Biochemie, Martinsried bei München, under the supervision of Prof. Dr. Dr. h.c. Robert Huber.

I would like to start thanking Prof. Dr. Robert Huber for giving me the opportunity to work in his department, his financial support, and for providing the necessary conditions for the success of my projects. To Dr. Markus C. Wahl who helped me in the initial struggle and promptly accepted to share his projects with me. A special thank goes to Dr. Holger Dobbek and Dr. Berta Martins for their patience and for introducing me to the world of molybdenum enzymes.

I would like to acknowledge the secretaries Renate Rüller, Gina Beckmann, Monika Schneider and Monika Bumann for their help during these years.

To my office colleagues Otto Kyrieleis, Arne Ramsperger and Li-Chi Chang thank you for the always interesting conversations.

To all the friends who are not around, a big kiss, to each of you made this possible: Dr. Martin Augustin, Dr. Gerd Bader, Dr. Iris Fritze, Dr. Stefan Gerhardt, Dr. Daniela Jozic, Dr. Jens T. Kaiser, Dr. Chrystelle Mavoungou. And of course also those who are still here: Dr. Christine Breitenlechner, Mireia Comellas, Dr. Yuri Cheburkin, Mekdes Debela, Michael Engel, Dr. Rainer Friedrich, Dr. Peter Goettig, Dr. Stefan Henrich, Joannis Joannidinis, Dr. Dorota Ksiazek, Dr. Holger Niessen, Anna Tochowicz and Rasso Willkomm. For the nice atmosphere in the lab I would like to thank: Dr. Uwe Jacob, Dr. Stefan Krapp, Dr. Sofia Gadeiro Macieira, Snezan Marinkovic, Dr. Berta Martins, Dr. Paolo Perani and Emina Savarese.

Un grazie a Marta, Paolo e Benito per i bei momenti passati insieme. Un grazie a Francesca per la “zara-terapy”, i giri in macchina e la tua amicizia. Fede grazie per il tuo ottimismo e l’incondizionata amicizia. Ai miei genitori, a mio fratello, alla nonna e alle zie Miranda e Pina grazie per il vostro supporto e affetto in questi anni di Germania.

Table of Contents

<i>Acknowledgments</i>	4
<i>Table of Contents</i>	5
<i>1 Summary</i>	7
<i>2 Zusammenfassung</i>	9
<i>3 Introduction</i>	11
3.1 Quinoline 2-Oxidoreductase	11
3.1.1 Physical and Chemical Properties of Molybdenum	11
3.1.2 Molybdenum-containing Enzymes	12
3.1.3 Molybdenum Hydroxylases	14
3.1.4 <i>Pseudomonas putida</i> 86 Quinoline 2-Oxidoreductase	16
3.1.5 Molecular Properties of Qor	17
3.1.6 The Molybdenum Cofactor	18
3.1.7 The Iron-sulfur Clusters	19
3.1.8 The FAD Moiety	21
3.1.9 Purpose of this Study	22
3.2 The NusA AR1-λN Complex, and NusB	23
3.2.1 Transcription in Prokaryotes and Eukaryotes	23
3.2.2 Prokaryotic Transcription: Regulation at Transcription Initiation	25
3.2.3 Prokaryotic Transcription: Regulation After Initiation	26
3.2.4 Using Termination to Regulate Gene Expression	27
3.2.5 NusA, NusB, and the Family of Transcription Factors	29
3.2.6 Purpose of this Study	33
3.3 Methods for Structural Determination	33
<i>4 Experimental Procedures</i>	36
4.1 Protein Biochemistry	36
4.1.1 Crystal Structure of the Quinoline 2-Oxidoreductase	36
4.1.1.1 Cultivation of <i>P. putida</i> 86	36
4.1.1.2 Purification of Qor and Qor Variants	37
4.1.1.3 Assay for Qor Activity, and Estimation of Protein Concentrations	37
4.1.1.4 Polyacrylamide Gel Electrophoresis	38
4.1.2 Crystal Structure of the NusA AR1- λ N Complex, and NusB	38
4.1.2.1 Isothermal Titration Calorimetry	39
4.1.2.2 Circular Dichroism Measurements	39
4.1.2.3 Identification of the Oligomeric State in Solution	40
4.2 Crystallography	41
4.2.1 Phasing	41
4.2.1.1 The Phase Problem	41
4.2.1.2 The Patterson Function	42
4.2.1.3 Patterson Search or Molecular Replacement (MR)	42
4.2.1.4 Multi-wavelength Anomalous Dispersion (MAD)	43
4.2.2 Crystal Structure of the Quinoline 2-Oxidoreductase	44
4.2.2.1 Crystal Growth and Data Collection	44
4.2.2.2 Structure Determination	44
4.2.3 Crystal Structure of the NusA AR1- λ N Complex, and NusB	45
4.2.3.1 Crystal Growth and Data Collection	45
4.2.3.2 Structure Determination	45
4.2.3.3 Molecular Modeling Studies	46
4.2.4 Structural Analysis and Graphical Representation	46

5 Results and Discussion	47
5.1 Crystal Structure of the Quinoline 2-Oxidoreductase	47
5.1.1 Protein Biochemistry	47
5.1.2 Crystal Growth and Data Collection	48
5.1.3 Patterson Search and Model Refinement	50
5.1.4 Description of the Structure	52
5.1.4.1 Overall Structure	52
5.1.4.2 Structure of the Three Subunits	53
5.1.4.3 Substrate Channel	58
5.1.4.4 Qor active site. Moco and the Ligands around the Molybdenum Ion	60
5.1.4.5 Qor Active Site and Catalytic Pocket	61
5.1.4.6 Putative Substrate Binding Mode	64
5.1.4.7 Active Site Protein Variants	66
5.1.4.8 Conclusions and Future Perspectives	67
5.2 Crystal Structure of the NusA AR1-λN Complex	68
5.2.1 Protein Biochemistry	68
5.2.1.1 NusA-λN Binding Reactions	68
5.2.1.2 Analysis by Circular Dichroism Spectroscopy	69
5.2.2 Crystallography	70
5.2.2.1 Crystal Growth and Data Collection	70
5.2.2.2 Structure Determination	71
5.2.3 Description of the Structure	74
5.2.3.1 Crystal Content and Overall Structure	74
5.2.3.2 Structure of the NusA Acidic C-Terminal Repeats	76
5.2.3.3 Structure of the (NusA AR1) ₂ -λN Peptide Complex	78
5.2.3.4 λN Interaction is Stoichiometric in Solution	80
5.2.3.5 Identification of the Biologically Relevant Complex by Alanine Scanning	83
5.2.3.6 AR2 Contributes to λN Binding	86
5.2.3.7 λN Binding to NusA Does Not Enhance Its RNA Affinity	87
5.2.3.8 Summary and Conclusions	88
5.3 Crystal Structure of the Antitermination Factor NusB	90
5.3.1 Crystallography	90
5.3.1.1 Crystal Growth and Data Collection	90
5.3.1.2 Structure Determination	91
5.3.1.3 Oligomerization State in Solution	93
5.3.2 Description of the Structure	93
5.3.2.1 Overall Structure	93
5.3.2.2 Phylogenetic Comparisons	94
5.3.2.3 Quarternary Structures	96
5.3.2.4 <i>tmNusB</i> is Monomeric in Solution	101
5.3.2.5 Binding of <i>tmNusB</i> to <i>boxA</i> -like RNA	102
5.3.2.6 A Putative RNA Binding Region at the NusB N-terminus	105
5.3.2.7 Hypothesis - Dimerization as a Silencing Mechanism for Some NusB Proteins	106
6 References	109
7 Appendix	117
7.1 Abbreviations	117
7.2 Index of Figures	119
7.3 Index of Tables	120
7.4 Curriculum vitae	120

1 Summary

- **Quinoline 2-Oxidoreductase from *Pseudomonas putida* 86**

Quinoline 2-Oxidoreductase (Qor) is a member of the molybdenum hydroxylase family, catalyzing the first metabolic step of the quinoline degradation pathway converting quinoline to 2-oxo-1,2-dihydroquinoline. Still controversial for molybdenum hydroxylases is the nature of the molybdenum apical ligand. In the CO-dehydrogenase (*ocCODH*) crystal structure, the sulfido-ligand was found in the equatorial position while the oxo-ligand occupied the apical position. In contrast, the re-sulfurated *dgALO* structure reveals an equatorial sulfido-ligand and an apical oxo-ligand. A crystallographic study on Qor was therefore initiated, and the equatorial position for the sulfido-ligand was unambiguously identified. The available spectroscopical and structural data suggested that a sulfido-ligand in the equatorial position is a conserved feature in the molybdenum hydroxylase family. The structural comparison of Qor with the allopurinol inhibited xanthine dehydrogenase from *Rhodobacter capsulatus* allows direct insight into the mechanism of substrate recognition and the identification of putative catalytic residues. Furthermore, additional studies of substrate-binding and -conversion by Qor could help to predict the feasibility to biologically oxidize toxic organic compounds.

- **The NusA AR1- λ N Complex**

Transcription factor NusA from *E. coli* carries duplicated C-terminal domains, which bind to protein N from phage λ . In order to elucidate the structural basis of the NusA- λ N interaction, NusA C-terminal repeats in complex with a λ N peptide (residues 34-47) were crystallized. The two NusA units became proteolytically separated during crystallization and the crystal exclusively contained two copies of the first domain in contact with a single λ N fragment AR1- λ N. The NusA modules employ identical regions to contact the peptide, but approach the ligand from opposite sites. In contrast to the α -helical conformation of the λ N N-terminus in complex with *boxB* RNA, the present region of λ N remains extended and aligns with loops of the NusA domains. Mutational analysis showed that only one of the observed AR1- λ N contacts is biologically significant supporting an equimolar ratio of NusA and λ N in antitermination complexes. Solution studies showed that the additional interactions are fostered by the second NusA repeat unit, consistent with known compensatory mutations in

NusA and λ N. Contrary to the RNA polymerase α subunit, λ N does not stimulate RNA interaction of NusA upon binding. These results suggest that λ N recruits NusA to antitermination complexes by binding to its C-terminal extension and serves as a scaffold to closely appose NusA and the mRNA.

- **NusB from *Thermotoga maritima***

In order to better understand the role of transcription factor NusB during transcriptional antitermination, a structure-function analysis of the protein from *Thermotoga maritima* was carried out. The structures of NusB from *E. coli* (*ecNusB*) and from *Mycobacterium tuberculosis* (*mtNusB*) have been determined previously. *ecNusB* was found to be monomeric and *mtNusB* dimeric. The significance of the *mtNusB* dimers and the existence in other organisms of NusB dimers are unknown. A crystallographic study of NusB from *Thermotoga maritima* (*tmNusB*) has been initiated and five different crystal forms of the protein have been obtained. Three crystal forms harbored monomeric NusB, two forms showed dimers which were similar to *mtNusB* dimers. Solution studies hinted to a preference for the monomeric form of *tmNusB*. Sequence and structural characteristics pointed to an important RNA binding site in the N-terminal region of NusB. The fact that this RNA binding site is occluded in dimeric form of the protein indicated that the dimerization might be used by some organisms to package the protein into an inactive form.

Part of this thesis have been or will be published in due course:

Bonin, I., Martins, B.M., Purvanov, V., Fetzner, S., Huber, R. and Dobbek, H. (2004). Active Site Geometry and Substrate Recognition of the Molybdenum Hydroxylase Quinoline 2-xidoreductase. *Structure*, **12**, 1425-1435.

Bonin, I., Mühlberger, R., Bourenkov, G.P., Huber, R., Bacher, A., Richter, G. and Wahl, M.C. (2004). Structural basis for an interaction of *Escherichia coli* NusA with Protein N of Phage λ . *PNAS*, in press

Bonin, I., Robelek, R., Benecke, H., Urlaub, H. Bacher, A., Richter, G. and Wahl, M.C. (2004). Crystal structures of the antitermination factor NusB from *Thermotoga maritima* and implications for RNA Binding. *Biochem. J.*, in press

2 Zusammenfassung

- **Chinolin-2-Oxidoreduktase aus *Pseudomonas putida* 86**

Chinolin-2-Oxidoreduktase (Qor) gehört zur Familie der Molybdän-Hydroxylasen und katalysiert den ersten Schritt im Chinolinabbau durch Umsetzung von Chinolin zu 2-Oxo-1,2-dihydrochinolin. Für die Molybdän-Hydroxylasen ist die Frage nach der Natur des apikalen Liganden noch immer ungeklärt. In der Kristallstruktur der CO-Dehydrogenase (*ocCODH*) wurde der Sulfidoligand in äquatorialer Stellung gefunden, während der Oxoligand die apikale Bindungsstelle besetzte. Daher wurde eine kristallographische Studie über Qor unternommen, in der die äquatoriale Stellung des Sulfidoliganden eindeutig zugewiesen werden konnte. Die bekannten spektroskopischen und strukturellen Daten sprechen dafür, dass ein äquatorial gebundener Sulfidoligand ein konserviertes Merkmal der Molybdän-Hydroxylasefamilie ist. Der strukturelle Vergleich von Qor mit durch Allopurinol inhibierter Xanthindehydrogenase von *Rhodobacter capsulatus* erlaubt Einblick in die Substraterkennung und die Identifizierung der vermuteten katalytischen Reste. Des weiteren könnten zusätzliche Studien zur Substratbindung und –umsetzung durch Qor bei Vorhersagen über die Machbarkeit der biologischen Oxidation toxischer organischer Verbindungen helfen.

- **Der NusA AR1- λ N-Komplex**

Der Transkriptionsfaktor NusA aus *E. coli* besitzt duplizierte C-terminale Domänen, welche das Protein N des Phagen λ binden. Um die strukturelle Grundlage der NusA- λ N-Wechselwirkung zu erhellen, wurden die C-terminalen Wiederholungssequenzen im Komplex mit einem λ N-Peptid (Reste 34-47) kristallisiert. Die beiden NusA-Einheiten wurden bei der Kristallisation proteolytisch gespalten und der Kristall enthielt ausschließlich zwei Kopien der ersten Domäne in Kontakt mit einem einzigen λ N-Fragment (AR1- λ N). Die NusA-Module bedienen sich identischer Bereiche zur Bindung des Peptids, nähern sich jedoch dem Liganden von der entgegengesetzten Seite. Im Gegensatz zur α -helikalen Konformation des λ N-N-Terminus im Komplex mit *boxB*-RNA bleibt die vorliegende λ N-Region in einer gestreckten Konformation und aligniert mit *Loops*/Schleifen der NusA-Domänen. Eine Mutationsanalyse zeigte, dass nur einer der beobachteten AR1- λ N-Kontakte biologische

Bedeutung hat, indem ein äquimolares Verhältnis von NusA und λ N in Antiterminationskomplexen gefördert wird. Untersuchungen mit Lösungen ergaben, dass die zusätzliche Wechselwirkung durch die zweite NusA-Einheit verstärkt wird, was im Einklang mit bekannten kompensatorischen Mutationen in NusA und λ N steht. Im Gegensatz zur α -Untereinheit der RNA-Polymerase stimuliert λ N bei der Bindung nicht die Wechselwirkung mit RNA. Diese Ergebnisse legen nahe, dass λ N NusA durch C-terminale Bindung zur Bildung von Antiterminationskomplexen befähigt und als Gerüst bei der Anlagerung von NusA und mRNA dient.

- **NusB aus *Thermotoga maritima***

Für ein besseres Verständnis der Rolle des Transkriptionsfaktors NusB bei der transkriptionalen Antitermination wurde eine Struktur-Funktionsanalyse des Proteins aus *Thermotoga maritima* durchgeführt. Die Strukturen von NusB aus *E. coli* (*ecNusB*) und aus *Mycobacterium tuberculosis* (*mtNusB*) wurden vor kurzem aufgeklärt. *ecNusB* erwies sich als Monomer und *mtNusB* als Dimer. Die Bedeutung der *mtNusB*-Dimere und eine mögliche Existenz von NusB Dimeren in anderen Organismen sind noch unbekannt. An NusB aus *Thermotoga maritima* (*tmNusB*) wurde eine kristallographische Studie unternommen, wobei fünf verschiedene Kristallformen des Proteins erhalten wurden. Drei Kristallformen besitzen monomeres NusB und zwei Formen Dimere, ähnlich den *mtNusB*-Dimeren. Untersuchungen in Lösung sprechen für ein Überwiegen der monomeren Form von *mtNusB*. Sequenz und strukturelle Charakteristika weisen auf eine wichtige RNA-Bindungsstelle in der N-terminalen Region von NusB hin. Da die RNA-Bindungsstelle in der dimeren Form des Proteins abgeschirmt ist, erscheint eine Verwendung des Dimeren als inaktive Lagerungsform naheliegend.

3 Introduction

3.1 Quinoline 2-Oxidoreductase

3.1.1 Physical and Chemical Properties of Molybdenum

Molybdenum (Mo) is the only element from the second (4d) period of the transition metals with known biological function. Molybdenum has been recognized since the 1930s for its role in the biological nitrogen fixation and Mo-enzymes play intimate roles in the global cycles of nitrogen, carbon, and sulfur (Johnson et al., 1996; Stiefel, 1997).

Mo is quite scarce in the earth's crust, comprising only about 1.2 ppm. Like other heavy metals from this part of the periodic table, it is quite soluble in water at pH 7. Its most stable form has an oxidation state of +6, as in the trioxides, and forms a series of compounds known as the molybdates. The oxometalates of the heavier homologue tungsten, and those of the Group IIIB-VB metals to the left of Mo are nearly insoluble in water at pH 7 (Kaim and Schwederski, 1994).

Molybdenum chemistry, in oxidation states +5 to +2 is characterized by cluster formation. In aqueous solutions dimeric and trimeric ion systems are formed with the metal ions being liganded by hydroxo or oxo bridges and coordinatively saturated by water ligands. Such aggregation can be suppressed by the presence of proteinaceous multidentate chelating agents or specialized enzyme cofactors. The result is that the relevant oxidation states for molybdenum lie between +4 and +6 and the corresponding redox potential is about -300 mV, which is in the physiologically acceptable range (Kaim and Schwederski, 1994).

While molybdenum is used by mesophilic organisms, certain thermophilic organisms and hyperthermophiles utilize tungsten (W) instead of molybdenum. This might be due to the increased solubility of the tungsten sulfides. However, this substitution could also be due to the different electrochemistry of W and Mo compounds. Since the redox potentials for tungsten complexes are known to be higher than those of the molybdenum complexes they could more effectively reduce the Fe-S sites within a protein. Thus, the use of W may be explained by the simple increase in redox equivalents. Another explanation may be that tungsten enzymes are the evolutionary forerunners of the Mo-enzymes found in modern organisms (Abdel-Rassol, 2003).

3.1.2 Molybdenum-containing Enzymes

Molybdenum, an essential transition metal for over fifty redox enzymes, is available to biological systems because of the solubility of its high-valence oxides in water. It can be either part of the multinuclear center as in nitrogenases or in the mononuclear active centers as in the oxomolybdenum enzymes. The oxomolybdenum enzymes catalyze the incorporation of oxygen into a variety of organic and inorganic compounds. This large class of enzymes can be subdivided, according to their reaction mechanism, into molybdenum hydroxylases and oxotransferases.

Molybdenum hydroxylases possess an LMoOS unit (Figure 1) and catalyze the oxidative hydroxylation of diverse aldehydes and aromatic heterocycles. The hydroxylation reaction carried out by molybdenum hydroxylases is substantially different from other type of redox reactions, since water is the source of the oxygen atom incorporated into the product, instead of molecular oxygen (Hille, 1996).

The oxotransferases can be further subdivided in the sulfite oxidase and DMSO reductase enzyme families. The eukaryotic sulfite oxidase and the assimilatory nitrate reductases possess an LMoO₂ unit and catalyze oxygen transfer to or from an available electron lone pair of a sulfur or nitrogen atom of the substrate (Kisker et al., 1997b). DMSO reductase family members on the other hand, are found in a variety of bacteria and include, among other enzymes, the dissimilatory nitrate reductases, formate dehydrogenases, trimethylamine-N-oxide reductases, and biotin sulfoxide reductases (Figure 1). Enzymes of this group possess a L₂MoX unit and catalyze an oxo-transfer reaction, general redox reaction, or other more complex reactions (Hille et al., 1999). In the last 10 years several three-dimensional structures of the oxomolybdenum enzymes have become available (Figure 1).

INTRODUCTION

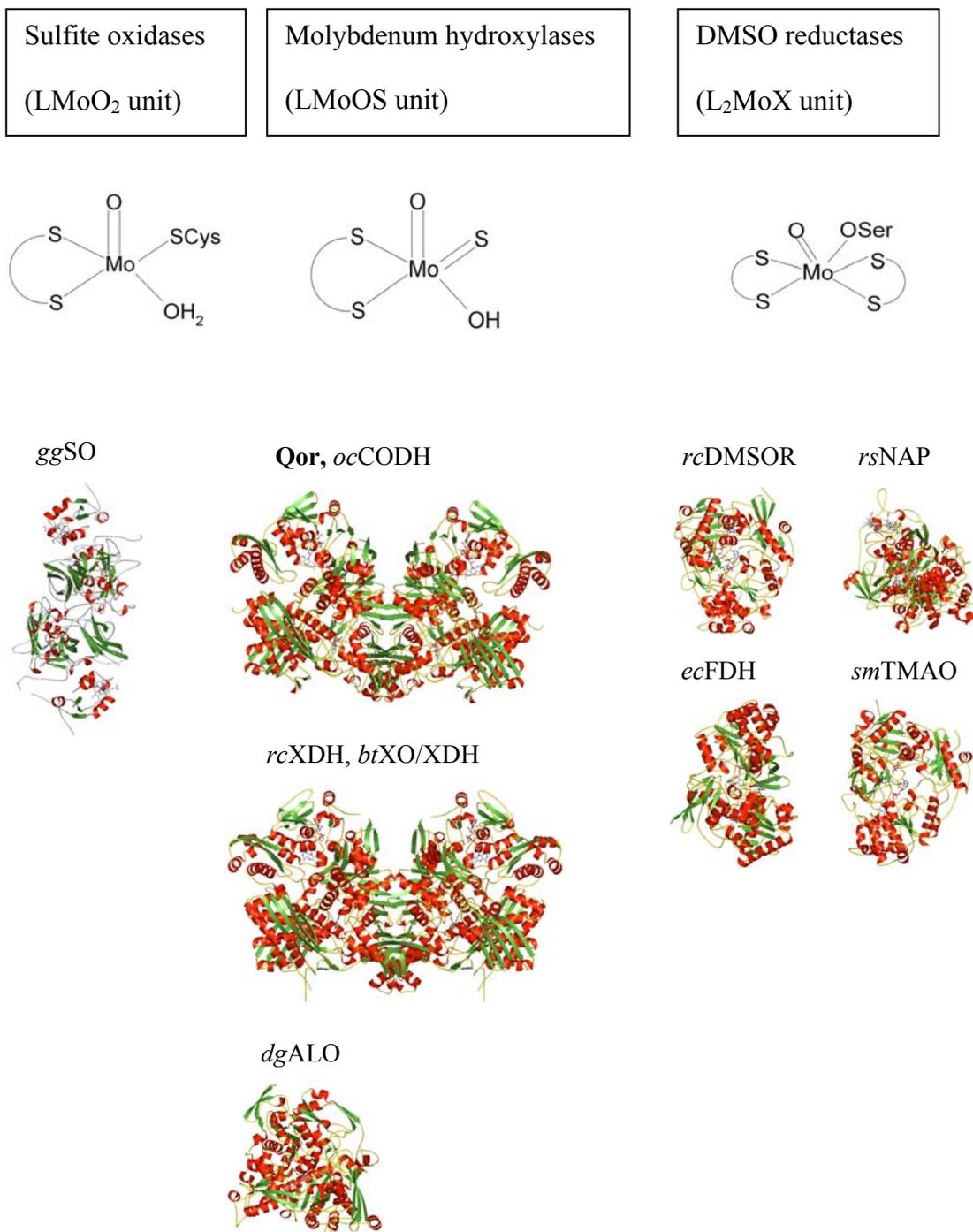


Figure 1. The mononuclear molybdenum enzyme family. From left to right: *Gallus gallus* sulfite oxidase (*ggSO*; Kisker et al., 1997a), *Pseudomonas putida* 86 quinoline 2-Oxidoreductase (**Qor**; Bonin et al., 2004), *Oligotropha carboxidovorans* CO dehydrogenase (*ocCODH*; Dobbek et al., 1999), *Rhodobacter capsulatus* xanthine dehydrogenase (*rcXDH*; Truglio et al., 2002), *Bos taurus* xanthine oxidase/dehydrogenase (*btXO/XDH*; Enroth et al., 2000), *Desulfovibrio gigas* aldehyde oxidoreductase (*dgALO*; Romão et al., 1995), *Rhodobacter capsulatus* dimethyl sulfoxide reductase (*rcDMSOR*; Schneider et al., 1996), *Rhodobacter sphaeroides* nitrate reductase (*rsNAP*; Arnoux et al., 2003), *Escherichia coli* formate dehydrogenase (*ecFDH*; Boyington et al., 1997), *Shewanella massilia* trimethylamine N-oxide reductase (*smTMAO*; Czjzek et al., 1998).

3.1.3 Molybdenum Hydroxylases

Several members of the molybdenum hydroxylase family are at present well characterized (Hille, 1996) (Table 1).

Enzymes	source	subunits	cofactor
Xanthine oxidase/dehydrogenase	<i>Bos taurus</i>	α_2	MPT
Xanthine dehydrogenase	<i>Rhodobacter capsulatus</i>	$(\alpha\beta)_2$	MPT
Aldehyde oxidase	<i>Homo sapiens</i>	α_2	MPT
Aldehyde oxidoreductase	<i>Desulvibrio gigas</i> <i>Desulvibrio desulfurincans</i>	α_2	MCD
CO-dehydrogenase	<i>Oligotropha carboxidovorans</i> <i>Hydrogenophaga pseudoflava</i>	$(\alpha\beta\gamma)_2$	MCD
Quinoline 2-oxidoreductase	<i>Pseudomonas putida</i> 86	$(\alpha\beta\gamma)_2$	MCD
Isoquinoline 1-oxidoreductase	<i>Pseudomonas diminuta</i>	$(\alpha\beta)_2$	MCD
Quinoline 4-carboxylate 2-oxidoreductase	<i>Agrobacterium</i> sp. 1B	$(\alpha\beta\gamma)_2$	MCD
Quinaldine 4-oxidoreductase	<i>Arthrobacter</i> sp.	$(\alpha\beta\gamma)_2$	MCD
Quinaldic acid 4-oxidoreductase	<i>Pseudomonas</i> sp. AK-2	$(\alpha\beta\gamma)_2$	MCD
Nicotinate dehydrogenase	<i>Bacillus niacini</i>	$(\alpha\beta\gamma)_2$	MCD
6-hydroxy nicotinate dehydrogenase	<i>Bacillus niacini</i>	$\alpha\beta\gamma$	MCD
Nicotine dehydrogenase	<i>Arthrobacter oxidans</i>	$\alpha\beta\gamma$	MCD
(2 <i>R</i>)-hydroxy carboxylate oxidoreductase	<i>Proteus vulgaris</i>	$(\alpha\beta\gamma)_2$	MCD

Table 1. Selected representatives of the molybdenum hydroxylase enzyme family. Structurally characterized members of the family highlighted in yellow.

Members of the molybdenum hydroxylase family generally catalyze hydroxylation reactions of the type shown in Scheme 1.



Scheme 1. Hydroxylation reaction catalyzed by the molybdenum hydroxylases.

INTRODUCTION

In this reaction reducing equivalents are generated and the water is used as source of oxygen that is incorporated into an aromatic/aliphatic substrate. The overall mechanism is composed of reductive and oxidative half-reactions (Hille, 1996). In the reductive half-reaction the two electrons are transferred to the molybdenum ion that changes its oxidation state from Mo(+6) to Mo(+4) with the concomitant hydroxylation of the substrate. In the following oxidative half-reactions electrons are transferred via other cofactors to the final electron acceptor regenerating the active center.

Enzymes of this family are organized either as α_2 homodimers, such as *dgALO* with the redox-active cofactors placed within a single subunit, or as multisubunit enzymes (e.g. *rcXDH* ($\alpha\beta$)₂ heterotetramer, *Qor* ($\alpha\beta\gamma$)₂ heterohexamer) (Hille, 1996). In addition to Moco, the molybdenum hydroxylases also contain two distinct [2Fe-2S] clusters discernible by their characteristic EPR signals and a FAD moiety. *dgALO* is functional without the flavoprotein subunit/domain usually found within this enzyme family. The xanthine dehydrogenase from *Veilonella atypica* contains on the other hand only one [2Fe-2S] cluster (Gremer and Meyer, 1996).

The classification of a molybdoenzyme as a molybdenum hydroxylase family member is based on their sequence homology and when not available according to the cyanide reactivity. The inactivation of the molybdenum hydroxylases by cyanide derives from the formation of a thiocyanate anion by the reaction of the molybdenum ligand sulfido with CN (Coughlan et al., 1980). The position of the sulfido-ligand has direct implications for the mechanism of substrate conversion and the still enigmatic functions of the sulfido- and oxo-ligands in the catalytic cycle (Young, 1997). The high resolution crystal structure of *dgALO* resulfurated at high pH with very high sulfide concentration revealed apical sulfido- and equatorial oxo- and water ligands at the active center (Huber et al., 1996). This configuration has subsequently been used to interpret the electron density of the active sites of *btXO/XDH* (Enroth et al., 2000) and *rcXDH* (Truglio et al., 2002). In contrast to the former structures, an active site geometry with an equatorial sulfido-ligand has been identified in the true atomic resolution structure of *ocCODH* (Dobbek et al., 2002). *ocCODH* is an atypical molybdenum hydroxylase in which the sulfido-ligand bridges the two metals forming a [CuSMo(=O)OH] cluster. *Qor*, as well as the other characterized molybdenum hydroxylases, contains a terminal sulfido-ligand at the molybdenum ion (Bonin et al., 2004).

3.1.4 *Pseudomonas putida* 86 Quinoline 2-Oxidoreductase

The aerobic soil bacterium *Pseudomonas putida* 86 is a chemoorganoheterotrophic organism using quinoline (2,3-benzopyridine) as sole source of carbon and energy. *P. putida* 86 was first isolated from the soil of a coal tar refining factory (Rütgerswerke, Castrop-Rauxel, Germany) and is a Gram-negative bacterium having rod-shaped cells with multitrichous flagella (Figure 6). Quinoline is a hepatocarcinogen in mice and rats and several quinoline derivatives have been proven to be mutagens (Suzuki et al., 1998). For bacteria the ability to biomineralize aromatic compounds may allow the exploitation of these microorganisms in bioremediation of polluted areas.

The “8-hydroxycoumarin pathway” of quinoline degradation has already been described in the 1980ies, but many steps are not yet fully understood (Figure 2). In the first step, quinoline 2-oxidoreductase (Qor) catalyzes the hydroxylation to 1*H*-2-oxoquinoline with concomitant reduction of a suitable electron acceptor (Bläse et al., 1996; Figure 2).

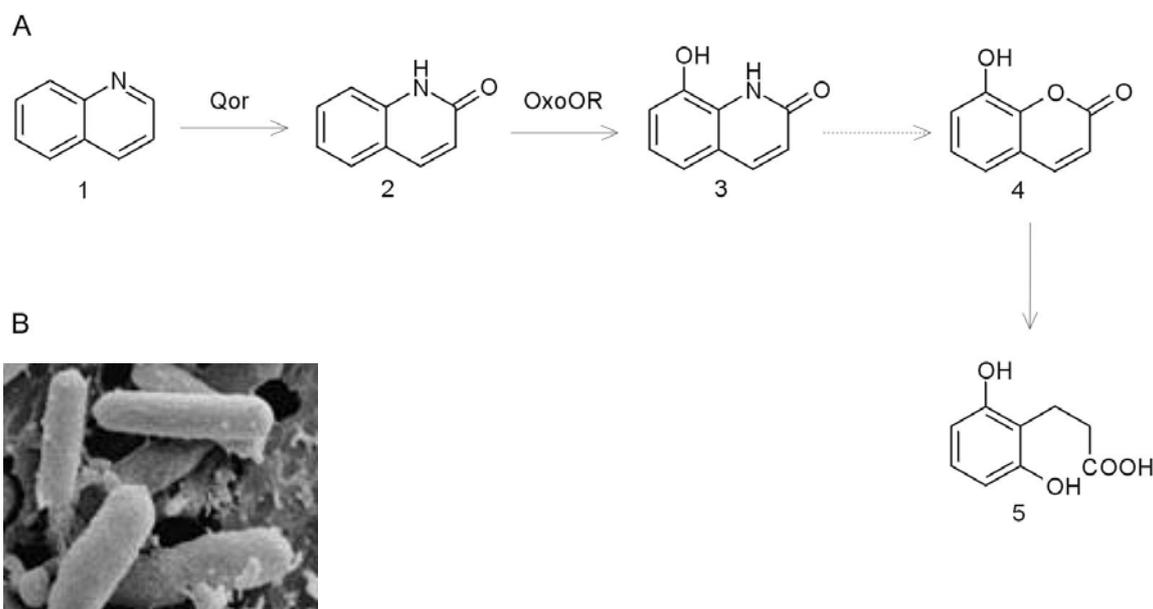


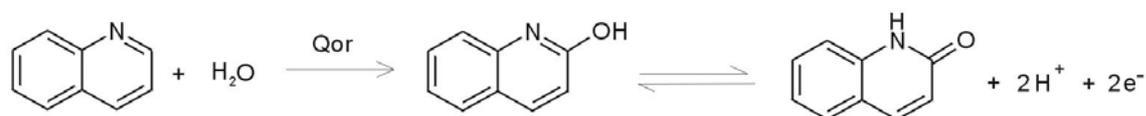
Figure 2. Pathway of quinoline degradation. **A.** The “8-hydroxycoumarin pathway” of quinoline degradation. (1) Quinoline (2) 1*H*-2-oxoquinoline (3) 1*H*-8-hydroxy-2-oxoquinoline (4) 8-hydroxycoumarin (5) 3-(2,3-dihydroxyphenyl)propionic acid. Qor, OxoOR: 1*H*-2-oxoquinoline 8-monooxygenase. **B.** *Pseudomonas putida* image obtained by scanning electron microscopy.

In the next step the enzyme 1*H*-2-oxoquinoline 8-monooxygenase (OxoOR), catalyzes the formation of 1*H*-8-hydroxy-2-oxoquinoline. The subsequent steps of the degradation pathway

are speculative, since the further proposed enzymes have not yet been detected, precluding a detailed elucidation of the “8-hydroxycoumarin pathway” (Fetzner et al., 1998).

3.1.5 Molecular Properties of Qor

Qor is the key enzyme involved in the bacterial degradation of quinoline. The substrate becomes hydroxylated at the C2-atom of the heterocyclic nitrogen-containing ring generating the prevailing tautomer 1*H*-2-oxoquinoline (Scheme 2).



Scheme 2. Catalytic reaction of Qor.

Qor belongs to the molybdenum hydroxylase enzyme family (Hille, 1996). The native enzyme (molecular mass 300 KDa) consists of a large (L) 85 KDa, a medium (M) 30 KDa and a small (S) 20 KDa subunit, forming a dimer of heterotrimers (LMS)₂. It contains four redox active centers that constitute an internal electron transport chain, consisting of a Moco, two distinct [2Fe-2S] clusters and one FAD moiety (Bläse et al., 1996). In the Qor system, the electron acceptor has not yet been identified and only artificial electron acceptors have been used in the oxidation of quinoline. However, the probable electron acceptor is a membrane-bound cytochrome, as seen for the CODHs. In the CODH system, the electron accepting flavoproteins lie on one side of the dimer. Each forms an electron-transfer chain from the active centers to the membrane-bound cytochrome (Dobbek, 2000).

3.1.6 The Molybdenum Cofactor

The molybdenum cofactor (Moco) consists of a mononuclear tungsten or molybdenum ion coordinated to a molybdopterin ligand (Schindelin et al., 1997). In molybdenum hydroxylases and sulfite oxidases, it is present in a 1:1 stoichiometry to the protein. Its simplest structure is molybdopterin, such as found in eukaryotic enzymes (MPT; Figure 3B). In another common form, it exists as dinucleotide of guanine, adenine, cytidine or hypoxanthine as in Mo-containing prokaryotic enzymes (Hille et al., 1999). The first three-dimensional structure of an enzyme containing Mo-molybdopterin cytosine dinucleotide with the molybdenum ion (Mo-MCD) was *Desulfovibrio gigas* aldehyde oxidoreductase (Romão et al., 1995).

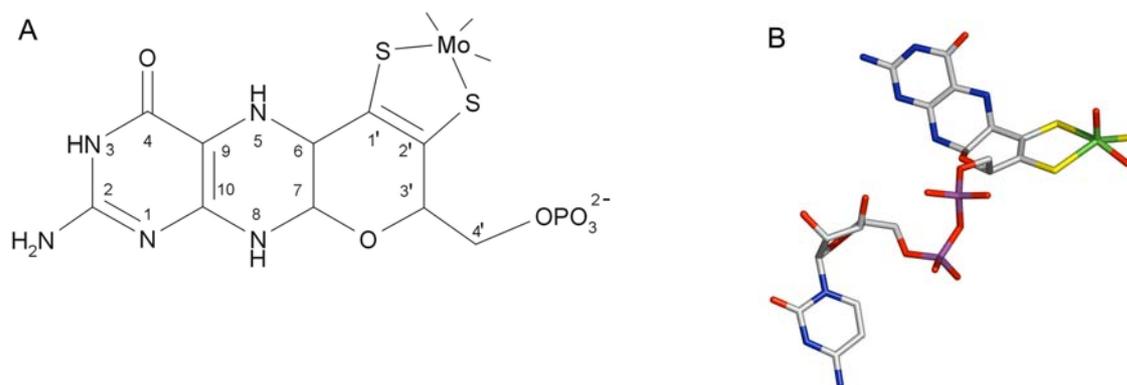


Figure 3. The molybdenum cofactor Moco. **A.** Schematic representation of the Mo-bound molybdopterin cofactor (MPT) as observed in two previous crystal structures (Enroth et al., 2000; Truglio et al., 2002). **B.** Structure of the Mo-molybdopterin cytosine dinucleotide (Mo-MCD) from the crystal structure of quinoline 2-oxidoreductase from *Pseudomonas putida* 86 (Bonin et al., 2004).

The Mo-MCD cofactor can be subdivided into four parts, the cytosine dinucleotide, the D-ribofuranose, the pyrophosphate and the tricyclic ring (Figure 3). The pyrimidine base is in *anti* conformation. The D-ribofuranose sugar is twisted with the C2" *endo* and the C3" *exo/endo* configuration (seen for *dgALO/Qor*), the pyrophosphate assumes different conformations in different structures (Rebelo et al., 2001; Bonin et al., 2004). The tricyclic ring was first identified in the tungsten-containing enzyme aldehyde oxidoreductase from *Pyrococcus furiosus*, and consists of a bicyclic pterin fused to the monocyclic pyran ring (Chan et al., 1995). The tricyclic ring is not planar with the pyrane ring adopting a half chair

conformation. In addition, the pterin ring is twisted at the C6 and C7 carbons. The three chiral centers C6, C7 and C3' are all in the *R* conformation, as was seen in *dgALO* (Romão et al., 1995). Most likely, an important role is played by the dithiolate portion of the pterin. These sulfurs are the site of metal coordination. Several functions have been attributed to this pyranopterin-ene-1,2-dithiolate system. Part of the cofactor might anchor the molybdenum, be involved in mediating the electron flow to and from the molybdenum, or modulate the redox potential of the molybdenum center via the dithiolene moiety (Romão et al., 1997).

3.1.7 The Iron-sulfur Clusters

The Fe-S clusters that are used as single electron carriers may well have been some of the earliest cofactors in enzymatic catalysis, reflecting naturally occurring chemical transformations in the prebiotic world. These clusters are ubiquitous in living organisms and contain sites with one to eight iron atoms, sometimes with multiple occurrence in the same protein like in nitrogenases (Beinert et al., 1997; Rees and Howard, 2003). In the course of evolution, other important uses of the Fe-S clusters related to electron transfer have developed. Fe-S centers have been recognized as sites for redox and nonredox catalysis, stabilization of DNA and iron/dioxygen/superoxide ion sensing (Rees and Howard, 2003). In proteins these centers occur as isolated clusters (e.g. in the electron-transferring ferredoxins), as part of multi-prosthetic group complexes (Ni-, Mo-, V-, Fe-heme), or as part of flavoproteins (Kaim and Schwederski, 1994).

The most common types of Fe-S clusters comprise [2Fe-2S], [3Fe-4S] and [4Fe-4S] cores with cysteinyl residues serving as fourth ligand to each iron atom (Figure 4).

INTRODUCTION

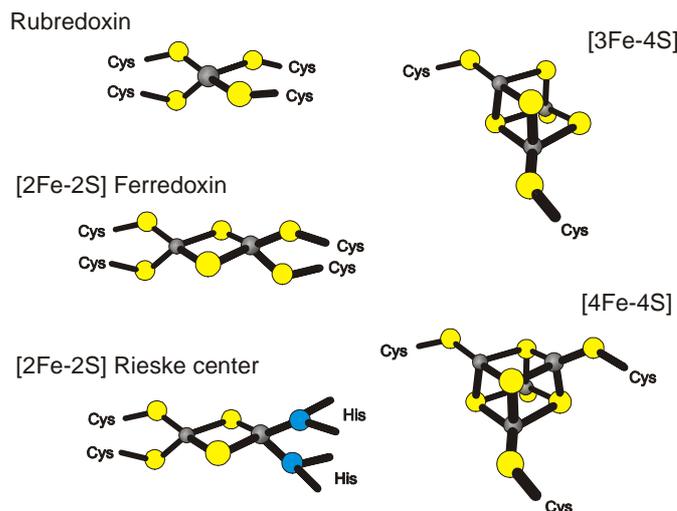


Figure 4. Structures of the most common types of iron-sulfur centers. Iron atoms are shown in grey, histidine nitrogens in blue, inorganic and cysteinyl sulfur in yellow.

The basic architecture of the Fe-S clusters is the [2Fe-2S] rhomb, with Fe-S bond distances averaging between 2.2 and 2.3 Å and with Fe-S-Fe and S-Fe-S bond angles that are slightly distorted from planarity (72-76° and 104-106°, respectively; Figure 5). The iron sites are generally tetrahedral and most commonly coordinated by sulfurs from inorganic sulfide and from cysteine thiol groups, although other amino acid side chains have been observed to coordinate iron (Rees and Howard, 2003). The different protein environment and, thus, the resulting electrostatic and structural asymmetry make the two iron centers not equivalent. The redox potentials are -405 to -380 mV for plant ferredoxins, -280 to -170 mV for putidaredoxin and others, -100 to 0 mV for Rieske proteins. During electron transfer the cluster undergoes a redox transition from a Fe(+3)/Fe(+3) state to a Fe(+3)/Fe(+2) state with charge localization (Kaim and Schwederski, 1994).



Figure 5. The canonical [2Fe-2S] cluster. Both iron ions are coordinated tetrahedrally with sulfur ligands.

The [4Fe-4S] clusters participate in many redox reactions such as photosynthesis, respiration and N₂ fixation acting as electron-transfer centers at redox potential of -700 mV. The charge of the oxidized cluster is -2 with two pairs of iron dimers present, both pairs showing approximately the same isomer shift (Mössbauer spectroscopy) corresponding to an oxidation state of +2.5 for iron. In contrast to the [2Fe-2S] system, the higher degree of delocalization found in the [4Fe-4S] cluster can be attributed to the structurally determined orthogonality of metal orbitals that interact via superexchanging sulfide bridges. The electron delocalization as such is, therefore, less susceptible to perturbation from external asymmetries as imposed from the protein (Kaim and Schwederski, 1994).

The [3Fe-4S] clusters have been found in ferredoxins of microorganisms and as a component of the enzyme aconitase, where a non-redox reaction within the Calvin cycle is catalyzed. These clusters can be found in equilibrium with a labile [4Fe-4S] form and can be derived by removal of a non cysteine-coordinated iron atom (Kaim and Schwederski, 1994).

3.1.8 The FAD Moiety

The dinucleotide FAD consists of adenosine monophosphate (AMP) linked to flavin mononucleotide (FMN) by a pyrophosphate bond (Figure 6). The FMN moiety is composed of the isoalloxazine-flavin linked to a ribityl phosphate (Figure 6). The catalytic function of the FAD is concentrated in the isoalloxazine ring, while the AMP and the ribityl phosphate mainly anchor the moiety to the protein (Dym and Eisenberg, 2001). The flavin's redox capacity is based on its ability to take up two electrons from one substrate and release them to a substrate/coenzyme or to an electron acceptor in either a two- or two one-electron steps (Massey, 2000). Therefore, the isoalloxazine ring can exist as an oxidized flavoquinone, one-electron-reduced flavosemiquinone, or a two-electron reduced flavoquinone. This enables flavins to participate in oxidation reactions involving free radicals and in reactions involving

metal ions. This ability to auto-oxidize allows some enzymes to pass electrons directly to oxygen and provides a basis for the functioning of flavins in hydroxylation reactions (Massey, 2000).

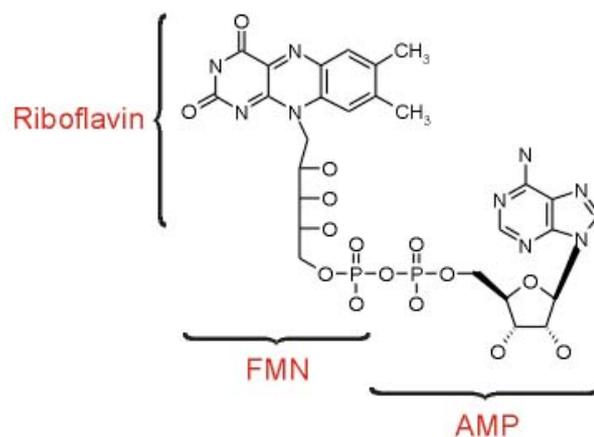


Figure 6. Chemical structure of the FAD moiety. The structure of flavin adenine dinucleotide (FAD), FAD is generated by the transfer of AMP to FMN.

Flavin moieties are usually tightly bound to proteins and cycle between the reduced and oxidized state while remaining attached to the protein. The flavin's redox potential depends on the nature of the active site in which the flavin resides and varies largely among different flavoproteins (-490 to 190 mV). Thus, flavoproteins can take part in a diversity of redox and electron-transfer reactions.

3.1.9 Purpose of this Study

The goal of this study was to determine the three-dimensional structure of the quinoline 2-oxidoreductase from *Pseudomonas putida* 86 by X-ray crystallography. The high resolution structural information allowed an unambiguously characterization of the molybdenum ligands. Insights into the substrate recognition and catalytic mechanism of Qor could be gained by comparison with *rc*XDH and *bt*XO/XDH, as the three enzymes catalyze the conversion of similar substrates. Moreover, the active site protein variants QorE743V and QorE743D were analyzed to assess the catalytic role of Glu743.

3.2 The NusA AR1-λN Complex, and NusB

3.2.1 Transcription in Prokaryotes and Eukaryotes

The first step in the flow of information from DNA to RNA is the synthesis of RNA from a DNA template, a process carried out by RNA polymerases (RNAP). Cells contain three major types of RNA: ribosomal RNA (rRNA), which constitutes two-third of the ribosomal mass, transfer RNAs (tRNA) that deliver amino acids to the ribosomes, and messenger RNA (mRNA), whose nucleotide sequences are translated into amino acids for protein synthesis. In addition, other small RNA species play various roles in processing newly transcribed RNA molecules (Voet et al., 1999).

Although the fundamental principles of transcription are similar in the kingdoms of life, eukaryotic transcription is distinguished by having multiple RNA polymerases and by relatively complicated promoter elements. In prokaryotes one type of RNAP synthesizes all of the cell's RNA with the exception of the RNA primers involved in DNA replication. The *E. coli* RNAP holoenzyme has been extensively studied and is composed of $\alpha_2\beta\beta'\sigma$ subunits with a molecular mass of approximately 449 KDa and initiates the RNA synthesis at specific sites. Once RNA synthesis has been initiated, the σ subunit dissociates from the core enzyme $\alpha_2\beta\beta'$, which carries out the actual polymerization process (Voet et al., 1999).

Eukaryotic nuclei contain three different types of RNAPs: RNAP I is located in the nucleoli and synthesizes the precursors of most rRNAs. RNAP II and RNAP III both occur in the nucleoplasm, where RNAP II synthesizes the mRNA precursors; RNAP III synthesizes precursors of 5S rRNA, tRNAs and different small RNAs. Eukaryotic RNAP have molecular masses that vary between 500 and 700 KDa with a complex subunit organization. The best characterized yeast RNAP II contains 12 subunits (Kornberg, 1999), where the two largest subunits contain regions of similarity to prokaryotic subunits (Figure 7).

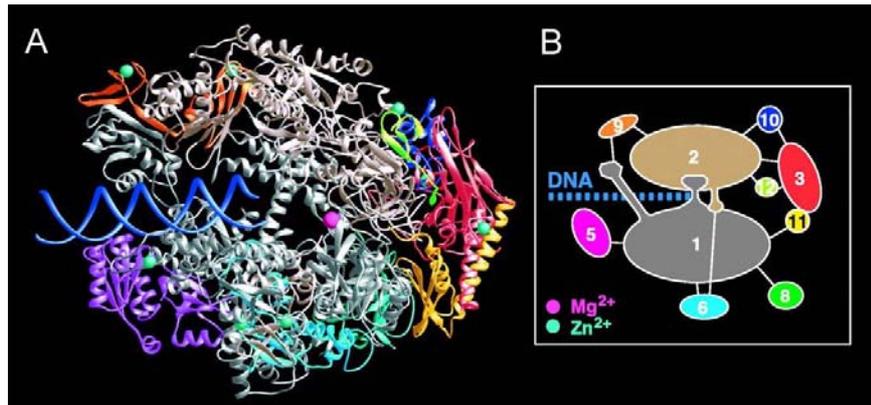


Figure 7. Yeast RNAP II. **A.** Architecture of yeast RNAP II. **B.** The ten subunits are shown as ribbon diagrams, right: interaction diagram (Cramer et al., 2000).

Many molecules are involved in the eukaryotic transcriptional control, including nucleosomes, polymerases, general transcription factors and the regulatory apparatus. Nucleosomes are keystones of the transcriptional control since they repress all genes except those, whose transcription is brought about by positive regulatory mechanism (Kornberg, 1999). Gene activation is predominantly achieved by the so-called activator proteins that bind regulatory DNA elements and stimulate transcription. Thus, the initiation of transcription proceeds in two stages: relief of repression by chromatin followed by the interaction of polymerase and accessory factors with the promoter. Activator proteins are presumed to trigger the first stage by the relief of the repression and through an additional protein factor interacting with RNAP II, termed ‘Mediator’. Moreover, they also appear to stimulate the subsequent elongation process. Therefore, in the eukaryotic transcriptional control two major themes prevail: first the relief by modifying enzymes and remodeling factors of nucleosome-induced repression; second the central role of Mediator in the transduction of regulatory information (Kornberg, 1999).

3.2.2 Prokaryotic Transcription: Regulation at Transcription Initiation

Transcription can be divided into three phases: Initiation, elongation and termination. RNAP binds to its initiation sites through promoter sequences, which are recognized by the corresponding σ factors. There are six to seven different σ subunits, each directing the enzyme to a unique set of promoters; σ^{70} is the predominant form and responsible for transcribing the majority of the genes in bacteria (Ptashne and Gann, 2002). In genes that are controlled by regulated recruitment, proteins such as *lac* and λ repressors bind to sites overlapping a promoter and thereby exclude RNAP. It has been proposed that repressors bound to sites flanking a promoter can form a tight loop that excludes RNAP (Ptashne and Gann, 2002).

The two proteins, *lac* repressor and the catabolite activator protein (CAP) dock with their DNA binding sites to the major groove of DNA using a similar structural motif, the so-called helix-turn-helix (HTH) motif. Each HTH bears the recognition α -helix that makes sequence-specific contacts with edges of base pairs (Ptashne and Gann, 2002; Figure 8).



Figure 8. Crystal structure of a CAP-DNA complex. The DNA is bent by 90 degrees (Schultz et al., 1991).

Even for proteins that use the same motif, e.g. the HTH motif, the details of DNA binding vary. CAP dramatically bends the DNA while other proteins like the λ repressor and Cro have virtually no effect on the helical structure of DNA (Ptashne and Gann, 2002).

Bacteria however, also use other ways to repress transcription. These methods include silencing, where proteins bind over extended regions of DNA and exclude RNAP, inhibition

of DNA-bound activators, and inhibition of DNA-bound RNAP. Examples of the above cited mechanisms are as follows: SopB is a membrane-bound protein that bears specific DNA-binding sites. When overexpressed, it silences plasmid genes located within 10 kb of its specific binding sites. CytR is a repressor that inhibits the activator CAP. The phage protein P4 binds at the *P-A2c* promoter and prevents RNAP escape (Ptashne and Gann, 2002).

3.2.3 Prokaryotic Transcription: Regulation After Initiation

After promoter clearance, RNAP moves along the DNA template and thereby elongates the nascent mRNA chain at the 3' end. The mRNA synthesis proceeds until the RNAP meets a terminator sequence. At this point the enzyme stops adding nucleotides to the growing mRNA chain, releases the complete product and dissociates from the DNA template. Terminators have been distinguished in *E. coli* according to whether RNAP requires any additional factors to terminate *in vitro* (Lewin, 2000). The two types of terminators are simple or intrinsic terminators (rho-independent) and the rho-dependent terminators.

Intrinsic terminators cause the core enzyme to terminate in the absence of the rho (ρ) factor. They have two evident structural features: a hairpin loop structure and a run of 6-8 U residues at the end of the unit. The hairpin usually contains G·C-rich inverted repeats near the base of the stem followed by four or more adenines (Lewin, 2000; Figure 9). The effect of the hairpin causes the RNAP to pause in mRNA synthesis, thereby creating an opportunity for termination to occur. The string of U residues in the right location is necessary to allow RNAP dissociation since the rU·dA mRNA-DNA hybrid is a weaker pairing than any other kind of RNA-DNA hybrid (Lewin, 2000).

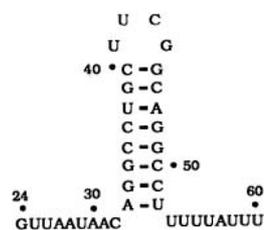


Figure 9. Variant of the λ R2 terminator. An intrinsic terminator shown as RNA sequence folded into the most stable secondary structure (Wilson and von Hippel, 1995)

INTRODUCTION

The sequences required for ρ -dependent termination are 50-90 bases long, rich in C bases and poor in G residues (Lewin, 2000). *E. coli* contains only few ρ -dependent terminators and most are found in phage genomes. The ρ factor is a hexameric mRNA/DNA helicase which contains a RNA-binding domain with an oligonucleotide/oligosaccharide binding fold at the N-terminus; a structure found in many single-stranded nucleic acid binding proteins (Richardson, 2003). The six RNA binding domains can be filled with a 60 nucleotides long mRNA that follow a zig-zag path from one subunit to the other (Figure 10). As ρ loads the nascent transcript, two regions (Q and R loops) determine RNA binding in the hole of the hexameric ring thereby stimulating ring closure. After appropriate loading, C-terminal domains are engaged in cycles of ATP binding and hydrolysis, translocating ρ along the mRNA until it reaches RNAP and unwinds the RNA/DNA heteroduplex at the transcription site (Skordalakes and Berger, 2003). Termination is completed by the release of ρ and RNAP from the nucleic acids.

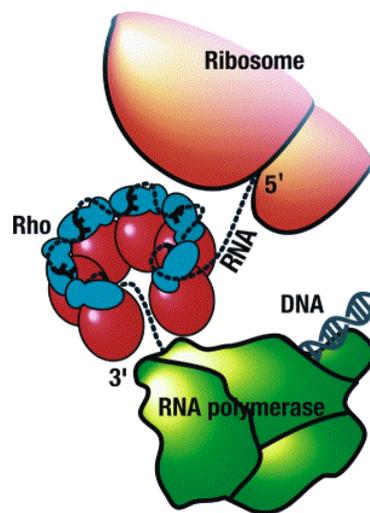


Figure 10. Rho-dependent termination. The diagram shows a nascent mRNA (black dashes) emerging from the exit site of RNAP and zig-zagging through the six RNA binding domains of ρ (Richardson, 2003). A ribosome is shown on the 5' side of the mRNA.

3.2.4 Using Termination to Regulate Gene Expression

Termination regulates gene expression downstream of a terminator, a process that includes a *cis*-acting element, which affects mRNA stability or translation. Antitermination and attenuation are the two well known ways of regulated termination. Attenuation controls the ability of RNAP to read through an intrinsic terminator located at the beginning of a

transcription unit. Attenuation requires simultaneous transcription and translation and therefore occurs only in prokaryotes. It is mediated by alternative secondary structures formed by the mRNA (Lewin, 2000). If one particular hairpin structure forms (region 3 and 4 in Figure 11), termination prevents RNAP from transcribing structural genes. But when the hairpin is prevented from forming, RNAP elongates through the terminator and genes are expressed. The changes in the mRNA secondary structure are determined by the position of the ribosome (Lewin, 2000; Figure 11). The best known example of attenuation is the *trp* operon, also attenuation is observed in five other operons in *E. coli* that are required for amino acid metabolism.

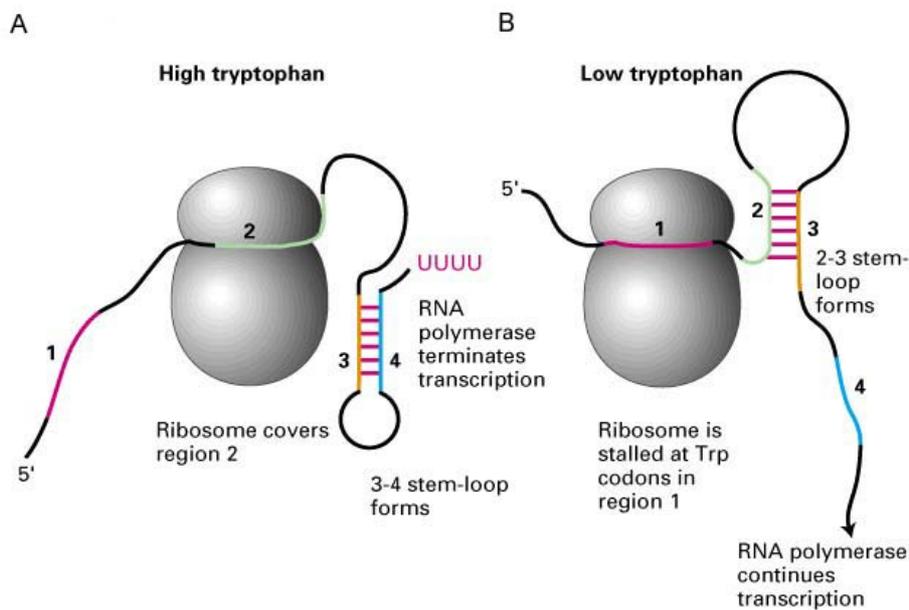


Figure 11. The *trp* operon. **A.** In the presence of *trp*-charged tRNA leader sequence is closely translated behind RNAP. **B.** When *trp*-charged tRNA is low the ribosome pauses before the end of the leader and allows to form an alternative stem-loop that prevents termination (Lewin, 2000).

Regulation of the *trp* operon in *B. subtilis* is mediated by binding of the MtrB protein to the leader of the transcript to promote formation of the terminator hairpin. MtrB is presumably activated by tryptophan and prevents the expression of *trp* genes by attenuation (Lewin, 2000). An alternative to control by regulatory proteins is control by small RNAs that are complementary to target mRNA regions. These regulatory RNAs thus induce the formation of a duplex RNA regions preventing translation by sequestering the initiation site. Regulatory RNAs that function by such means are called antisense RNAs (Lewin, 2000).

Antitermination was discovered in phage infections of *E. coli* and is used to switch between the different life cycles (Nudler and Gottesman, 2002). The lambda gene N codes for an antitermination protein (λ N) that is necessary for RNAP to read through the terminators located at the ends of the immediate early genes. Another phage protein λ Q is required later in phage infection. λ N and λ Q act on RNAP as it passes specific RNA sites: *nut* and *qut*, respectively. Binding of antitermination proteins, such as λ N between the promoter and the terminator allows a protein complex to form with Nus proteins and prevent termination. Antiterminators work at ρ -dependent and ρ -independent terminators (Ptashne and Gann, 2002). In eukaryotes RNAP I terminates when it approaches a polymerase-specific DNA binding protein. RNAP III, on the other hand, terminates at a series of U residues but does not require an upstream stem-loop to be present in the mRNA. RNAP II transcripts are terminated by cleavage near the polyadenylation site followed by the addition of the poly(A) tail. The cleaved 3' transcriptional product is rapidly degraded, like the un-polyadenylated transcripts (Lewin, 2000).

3.2.5 NusA, NusB, and the Family of Transcription Factors

Termination and antitermination are closely related and involve bacterial and phage proteins that interact with RNAP in response to specific sequences within certain transcription units. The phage protein λ N modifies RNAP in such a way that it ignores transcription termination signals located in the two early operons of the phage (Mogridge et al., 1995). This modification of RNAP requires a *nut* site situated between the promoter and the terminator of each operon. The *nut* site consists of a short conserved 12 nucleotide stretch (*boxA*) and is followed by a stem-loop structure (*boxB*) (Nodwell and Greenblatt, 1991). Sequence elements related to *boxA* are also found in bacterial operons and are required for binding bacterial proteins that are necessary for antitermination in both phage and bacteria. *boxB* is specific to the phage genome, and mutations in *boxB* abolish the ability of λ N to cause antitermination. The discovery of antitermination as a phage control mechanism led to the identification of further components of the transcription apparatus. The bacterial proteins that interact with λ N were identified by isolating *E. coli* mutants in which λ N is ineffective (Lewin, 2000). The host mutations that prevent λ N function identify the *nus* loci (N-ut^{ilization} substance). *nusA*, *nusB*, *nusC*, *nusD*, *nusE* and *nusG* code for proteins that form part of the transcription apparatus. The *nusC* mutations lie in the gene encoding the β subunit of RNA polymerase and

INTRODUCTION

the *nusD* mutations lie in the *rho* gene. *nusA*, *nusB* and *nusG* functions are concerned solely with the termination of transcription, while *nusE* codes for the ribosomal protein S10 (NusE) (Lewin, 2000). A closely related antitermination system occurs in the *rrn* (rRNA) operons of *E. coli*, and involves the same *nus* functions. The leader regions of the *rrn* operons contain *boxA* sequences; NusB-NusE dimers recognize these sequences and bind to RNAP as it elongates past *boxA*, and NusA joins subsequently. This complex formation changes the properties of RNAP in such a way that it can read through ρ -dependent terminators, which are present within the transcription unit. Factor NusA and the initiation factor sigma are mutually exclusive associates of the core enzyme. HK022, a λ competitor, produces an N-like protein, Nun, which operates with the same group of host factors at *nut* loci to stimulate transcription termination (Weisberg et al., 1999).

The 107 residue long unstructured λ N protein contains an N-terminal arginine-rich motif (ARM, residues: 1-22; Figure 12) and an activating region (residues 23-107; Figure 12). ARM binds as a bent α -helix to the *boxB* RNA hairpin in the *nut* site (Legault et al., 1998), while the activating region is required for the transcriptional antitermination (Mogridge et al., 1998a). The activating region can be further subdivided in an N-terminal sub-region containing the amino acids 34-47 responsible for binding NusA and a C-terminal sub-region (amino acids 73-107) that interact with RNAP (Figure 12).

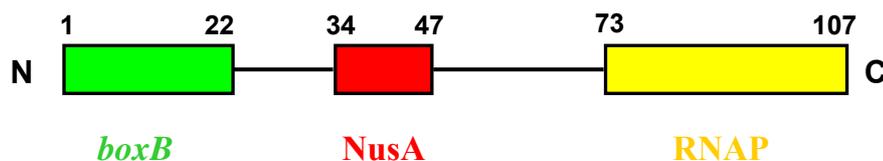


Figure 12. Functional segments of λ N.

The phage encoded λ N protein recognizes a signal sequence, *boxB*, and initiates the buildup of a multi-factor ribonucleoprotein complex on the surface of RNAP. NusA together with the *E. coli* proteins NusB, NusE (ribosomal protein S10), NusG, RNAP and the *nut* site on the phage RNA take part in the λ N-modified transcription complex (Greenblatt et al., 1993; Figure 13). The λ N-antitermination relies on intricate protein-protein and protein-RNA

networks defined by multiple weak protein-protein and protein-RNA interactions (Mogridge et al., 1995).

NusB and NusE form a dimer that specifically binds to RNA containing a *boxA* sequence. In *E. coli*, NusB alone can bind *boxA* but the presence of NusE significantly enhances the association by a still unknown mechanism. The interaction of NusB and *boxA* may be important, since it precludes the action of the inhibitor of the antitermination reaction (Patterson et al., 1994). The highly stable ribonucleoprotein complex that is then formed is able to suppress transcription termination over long distances and through multiple terminators (Mogridge et al., 1995).

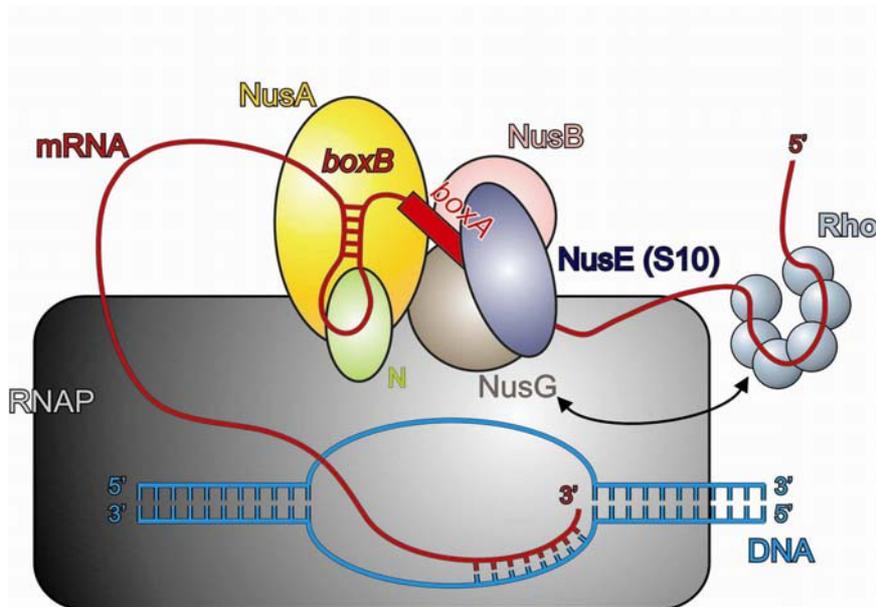


Figure 13. Model of the λ N-mediated antitermination complex. With the usage of specific terminators RNAP transcribes λ genes (Richardson and Greenblatt, 1996).

Within this complex NusA interacts with the λ N protein and both the C-terminal and N-terminal regions of NusA interact with the RNAP (Mah et al., 1999; Figure 13). NusA binds to RNAP shortly after the initiation of transcription and stimulates pausing and transcription termination at intrinsic terminators (Richardson and Greenblatt, 1996). Currently, the *in vivo* NusA: RNAP binding stoichiometry in the presence of λ N is an unresolved issue. It has been suggested that more than one molecule of NusA may accompany RNAP during elongation (Horwitz et al., 1987). The NusG factor has the opposite effect to the NusA action: it decreases transcriptional pausing and stimulates ρ -dependent termination. Both are required for stable modification of RNAP by λ N *in vivo* and *in vitro* (Richardson and Greenblatt, 1996).

The bacterial NusA structure contains the N-terminal α/β domain (NTD), the five-stranded β -barrel S1 domain and the tandemly duplicated K-homology regions (Worbs et al., 2001; Figure 14). The NTD portion of NusA (residues: 1-137) includes the RNAP-binding region. The S1 homology region and the two KH1/KH2 domains regions are the putative RNA-binding domains (Gibson et al., 1993; Bycroft et al., 1997; Worbs et al., 2001).

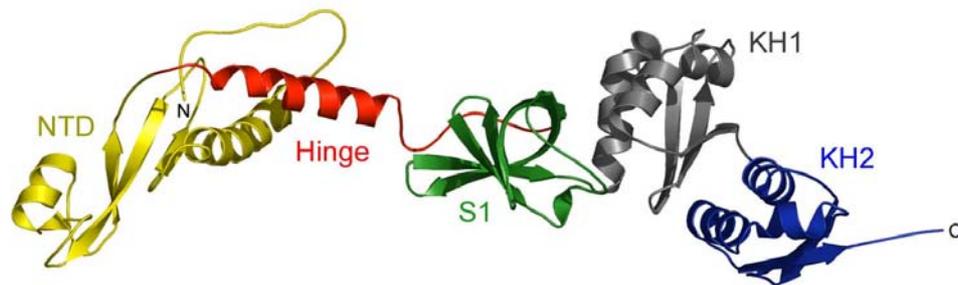


Figure 14. Domain arrangement in NusA from *T. maritima*. Displayed as secondary structure with colors according to the domains (Worbs et al., 2001).

NusA protein from *E. coli* and other γ -proteobacteria contain an additional C-terminal extension, which is made up of two, ~ 70 -residue dually repeated acidic domains. These modules are referred to acidic repeat (AR) 1 and 2. It has been suggested for this fold that it contains a modified version of the helix-hairpin-helix (HhH) motif (Mah et al., 2000). This motif has been found in a variety of DNA-binding and RNA-binding proteins, including many enzymes involved in DNA replication and repair, and is thought to mediate non-sequence-specific binding of proteins to nucleic acids. However, a different function for the HhH was postulated, specifically a participation in the complex formation between the C-terminal domain of the RNAP α subunit (α -CTD) and NusA. α -CTD and the λ N protein bind to the C-terminal region of NusA (Mah et al., 2000), which suggests that both α -CTD and λ N act in a similar way in controlling the binding of NusA to RNA. The interaction of α -CTD with NusA in the transcription complex allows NusA binding to the nascent transcript and determines its ability to stimulate pausing and termination by RNAP. The λ N protein on the other hand, reverses the effect of NusA on pausing and termination causing NusA to interact with *nut* site rather than to the RNA near the 3' end of the nascent transcript (Mah et al., 2000).

3.2.6 Purpose of this Study

- NusA AR1- λ N Complex

The NusA from *E. coli* contains, in contrast to the previously characterized *Thermotoga maritima* protein, dual acidic repeats AR1-AR2 at the C-terminus. The λ N-dependent antitermination is characterized by its interaction with this C-terminal extension in NusA, with RNAP and *boxB* RNA. Only the interaction of λ N with *boxB* RNA has thus far been structurally characterized. Therefore, a crystallographic study was initiated as to delineate the details of the λ N-NusA AR1-AR2 association. The crystal structure of a fragment of the λ N protein (residues 34-47) in complex with two molecules of the first acidic repeat AR1 was determined. By mutational studies it was possible to determine which of the two copies of AR1 is the physiologically relevant partner for the interaction with λ N.

- NusB from *Thermotoga maritima*

The aim of this study was the structural determination by X-ray crystallography, of the transcription factor NusB from the bacterium *Thermotoga maritima* (*tmNusB*). Two structures of the same protein were previously characterized: a monomeric *ecNusB* and a dimeric *mtNusB*. To determine the functional role of the dimer formation in NusB, *tmNusB* was crystallized and five different crystal forms were obtained. In three of them only monomers could be found, while two crystal forms harbored a dimeric NusB. Since the putative RNA-binding region is located at the N-terminus, a dimer formation is a possible way of silencing the protein. To verify this hypothesis solution studies together with RNA-binding studies were conducted.

3.3 Methods for Structural Determination

The necessity of determining the three-dimensional structure of macromolecules has led to the development of several techniques. The achievable resolution by any microscopic technique is limited by the applied wavelength. The radiation needed to analyze atomic distances lies within the spectral range of X-rays because their wavelength (e.g. $\text{CuK}\alpha$

radiation with a wavelength of 1.5148 Å) is comparable to atomic distances in the crystal lattice. Crystallographic methods rely on the analysis of the diffraction patterns that emerge from a sample that is targeted by an X-ray. While light or electron microscopy uses lenses to merge the waves diffracted by an object into an enlarged image, such lenses are not available for X-rays.

In the electron microscopy (EM) electrons scattered from the sample are recombined by the objective lens to form a magnified image and thereby retaining the phase information. This method can provide information about the surface feature of an object, picturing big assemblies like viruses, chaperones and the proteasome, but can not afford details at the atomic level. In a more recent method, the cryo-EM, it is possible to obtain a three-dimensional reconstruction of macromolecules at different states.

The ability of the X-ray crystallography to elucidate three-dimensional macromolecular structures at atomic level makes it the principal method for structural studies. This is reflected by the growing number of structures deposited in the Brookhaven Protein Data Bank (PDB). In the last update (13-July-2004) a total number of 26319 structures have been deposited within this archive, 22460 of them determined by X-ray diffraction techniques. To determine the structure of a macromolecule by X-ray however, single high quality diffracting crystals have to be obtained. This is a difficult challenge and often the rate-limiting step of the whole procedure.

Solution nuclear magnetic resonance (NMR) is a valuable tool in the determination of three-dimensional structures in solution. NMR detects chemical shifts of atomic nuclei having a nonzero spin in response to an applied magnetic field. This methodology has the advantage of giving useful information about the movement of individual molecular groups, or whole domains of proteins. The time-consuming procedure and the applicability to smaller proteins are the disadvantages of this technique. In X-ray crystallography, in contrast to NMR, there is no limit for the size of the macromolecule to be examined. This makes X-ray crystallography the method of choice for studying large macromolecules and their complexes at atomic level.

A common question that has been asked is whether the structural arrangement (or protein fold) displayed in a crystal is the same as in solution. There are three convincing arguments in favor of the fold similarity between solution and crystal-based structural data. (1) The protein in a crystal is also in an aqueous environment. (2) The crystals of enzymes are active (the enzyme activity is sensitive to structural changes). (3) Structures of several proteins solved by X-ray crystallography and NMR are identical within experimental variations. In few instances

INTRODUCTION

a difference might involve surface residues that in the crystal are implicated in intermolecular contacts and are therefore perturbed from their solution conformation.

4 Experimental Procedures

4.1 Protein Biochemistry

4.1.1 Crystal Structure of the Quinoline 2-Oxidoreductase

4.1.1.1 Cultivation of *P. putida* 86

P. putida 86 was cultured aerobically in 8 liter fermentors at 30°C in a quinoline minimal medium (Tshisuaka et al., 1993; Table 2). Fresh quinoline minimal medium was prepared by mixing the compounds in the amounts described in Table 2. The pH was adjusted to 7.3 by adding NaOH. After sterilization 0.5 ml/l of a 170 mM quinoline stock solution was added. When both quinoline and 2-oxo-1,2-dihydroquinoline were undetectable in the fermentation broth, an additional 0.5 ml/l quinoline stock solution was added. At an optical density (600 nm) of about 4, the cells were harvested by centrifugation at $14,000 \times g$ for 15 min at 4°C. The cell pellet was stored at -20°C until use.

P. putida 86-1 Δqor pUF1m743a and *P. putida* 86-1 Δqor pUF1m743b were cultured as described previously (Frerichs-Deeken et al., 2003). The construction of the plasmids for the Qor variants, the purification and the kinetic analysis were conducted in the group of Prof. Dr. S. Fetzner (Institut für Molekulare Mikrobiologie und Biotechnologie, Westfälische Wilhelms-Universität Münster).

Compound	(g/l)
Na ₂ HPO ₄	4.33
KH ₂ PO ₄	2.65
MgSO ₄ ·7H ₂ O	0.40
NaCl	1.00
FeSO ₄ ·7H ₂ O	0.10
Na ₂ MoO ₄ ·2H ₂ O	0.01

Table 2. Quinoline minimal medium according to Tshisuaka et al., 1993.

4.1.1.2 Purification of Qor and Qor Variants

Unless specifically mentioned, standard techniques and materials were used. Cells were suspended in 100 mM Tris-HCl buffer pH 8.5, containing 10 μ M PMSF and 0.05 μ l/ml Benzon nuclease (Merck, Darmstadt, Germany), and disrupted by sonication at 4°C. Cell debris was removed by centrifugation (48000 \times g, 45 min, 4°C). The supernatant was fractionated by a two step ammonium sulfate precipitation (20% (w/v) and 40% (w/v) saturation, respectively), the pellet of the 40% fraction was resuspended in buffer A (100 mM Tris-HCl pH 8.5, 400 mM ammonium sulfate). The resulting supernatant was loaded onto a Phenyl Sepharose column (Amersham-Pharmacia, Freiburg, Germany) and equilibrated with buffer A (flow rate 1.5 ml/min). The column was washed with buffer B (100 mM Tris-HCl pH 8.5, 200 mM ammonium sulfate) and a linear gradient from 0% buffer B to 100% buffer C (20 mM Tris-HCl pH 8.5, flow rate 1.0 ml/min, 150 ml) was applied. After an isocratic flow of 75 ml of buffer C the eluate was collected. The active pools were analyzed by native polyacrylamide gel electrophoresis and for Qor activity; peak fractions were combined, concentrated (Millipore, Eschborn, Germany) and loaded onto a DEAE-Sepharose column (BioRad, München, Germany), equilibrated with buffer D (200 mM Tris-HCl pH 8.0). The protein was eluted at a flow rate of 1.0 ml/min with a linear gradient from 300 to 900 mM Tris-HCl pH 8.0. Active fractions were pooled in three different batches and concentrated separately (Millipore, Eschborn, Germany). The concentrated protein was loaded onto a Superdex G200 size exclusion column (Amersham-Pharmacia, Freiburg, Germany) and chromatographed with 150 ml buffer E (50 mM Tris-HCl pH 8.0, 150 mM NaCl, at a flow rate 2.0 ml/min; Figure 15).

Qor variants were purified as described (Frerichs-Deeken et al., 2003). Both non-denaturing and SDS-PAGE were used to check the homogeneity of the purified Qor proteins.

4.1.1.3 Assay for Qor Activity, and Estimation of Protein Concentrations

The activities of Qor and Qor variants were determined spectrophotometrically by measuring the quinoline-dependent reduction of the artificial electron acceptor *p*-iodonitrotetrazolium violet (INT) to its formazan (Tshisuaka et al., 1993). The standard assay mixture consisted of 250 μ l test buffer (0.2 M Tris-HCl pH 8.0, with 1% Triton X-100), 750 μ l INT (5 mM in ddH₂O), 10 μ l quinoline (170 mM in isopropanol) and 10 μ l enzyme

solution. The reaction was started by the addition of quinoline. One unit of enzyme activity was defined as 1 μmol of formazan formed per minute at 25°C using the absorption coefficient of $19.3 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ at 503 nm.

Protein concentrations were estimated by the method of Bradford as modified by Zor and Selinger (Zor and Selinger, 1996) using bovine serum albumin as standard protein.

4.1.1.4 Polyacrylamide Gel Electrophoresis

SDS-PAGE (Laemmli, 1970) was performed using 12% acrylamide in the resolving gels. Non-denaturing PAGE was performed using the same protocol, but omitting the SDS in the resolving and stacking gels. Proteins were stained with Coomassie blue R-250 (0.1% (w/v) in 50% (w/v) aqueous TCA). For activity staining of Qor, polyacrylamide gels were immersed in the same buffer as used in the spectrophotometric assay, containing INT and quinoline.

4.1.2 Crystal Structure of the NusA AR1- λN Complex, and NusB

The cloning of the *nusA* and *nusB* genes and preliminary purification was done in the group of Dr. Gerald Richter (Technische Universität München, Garching, Germany). The *nusA* gene was cloned in two additional constructs, coding for amino acids 1-349 (lacking the C-terminal AR1 AR2 region) and 350-495 (encompassing solely the C-terminal region). The corresponding proteins were purified. Gel mobility shift assays, UV-induced crosslinking and mass spectrometric analysis experiments were performed in the group of Dr. M.C. Wahl (Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany). The synthesis of the λN -derived peptides was done by M. Krause at the Institut für Molekularbiologie und Tumorforschung, Marburg, Germany.

The purified NusA proteins were dialyzed against a buffer containing 10 mM Tris-HCl pH 7.0, 2 mM DTT and 100 mM NaCl prior to use. The buffer for *tmNusB* was exchanged by applying 3 ml aliquots on a 5 ml High Trap desalting column (Amersham-Pharmacia, Freiburg, Germany) and eluted with 10 mM Tris-HCl pH 7.0. No reducing agent has been used in the crystallization buffer, since *tmNusB* contain no cysteines.

The protein concentration was estimated by absorption spectroscopy at 276 nm using an extinction coefficient of $12650 \text{ M}^{-1} \text{ cm}^{-1}$ for *tmNusB*. For the NusA proteins the following

extinction coefficients (at 278 nm) were used: 30927 M⁻¹ cm⁻¹ for NusA, 22400 M⁻¹ cm⁻¹ for the N-terminal region and 8527 M⁻¹ cm⁻¹ for the AR1-AR2 region of NusA (Expasy, ProtParam Tool <http://www.expasy.org/cgi-bin/protparam>).

4.1.2.1 Isothermal Titration Calorimetry

The lyophilized λ N-derived peptides were solubilized prior to use in the dialyzing buffer. The NusA samples and the solubilized λ N peptides were dialyzed over night against a buffer containing 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 2 mM DTT. Protein and peptide concentrations were estimated by UV spectroscopy and amino acid analysis, respectively. The peptides were used at concentrations from 2.6 to 4.0 mM, the NusA proteins had concentrations ranging between 0.10 and 0.14 mM and all solutions were degassed prior to use. The measurements were carried out at 20°C using a MCS-ITC Instrument (MicroCal, Northampton, MA, USA) to record enthalpy and heat capacity changes (Wiseman et al., 1989). The λ N peptides were injected into the reaction cell of 1.37 ml containing NusA proteins in a series of 40 μ l injections at 400 s intervals. The data were analyzed with MicroCal Origin 5.0 (MicroCal, Northampton, MA, USA) to calculate the enthalpies and entropies of binding (ΔH_a and ΔS_a , respectively), the equilibrium dissociation constants (K_d) and the stoichiometries (n). Plain buffer was injected into protein solutions as a control. The heat releases from the control experiments were subtracted from the experimental data before evaluation.

4.1.2.2 Circular Dichroism Measurements

All measurements were done with a J-715 spectropolarimeter (JASCO, Tokyo, Japan) at a protein concentration of 0.2 mg/ml. The protein solution was dialyzed over night against a buffer containing 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 2 mM (*D,L*)-DTT (Sigma, Deisenhofen, Germany). Racemic DTT was used to avoid interferences in the measured spectra. For higher accuracy the protein concentration was determined by UV spectroscopy. For interpretation of the CD-spectra the program CDNN (Bohm et al., 1992) was used.

4.1.2.3 Identification of the Oligomeric State in Solution

In order to investigate the oligomeric state of *tmNusB* in solution, analytical gel filtration experiments and crosslinking experiments with glutaraldehyde were conducted.

Analytical gel filtration chromatography was performed on a SMARTTM FPLC system with a Superdex 75 PC 3.2/3.0 column (Amersham-Pharmacia, Freiburg). The column was calibrated with globular protein size standards (Fluka, Neu-Ulm; BioRad, München, Germany). The protein standards used in the calibration were as follows: bovine γ -globulin (158 KDa), chicken ovalbumin (44 KDa), cyclophilin H (20 KDa), aprotinin (6.5 KDa) and pyridoxal phosphate (1.35 KDa).

tmNusB samples in concentrations between 0.1 and 0.4 mM in the respective buffers (listed below) were incubated at 50°C for 15 min and then applied in 50 μ l aliquots to a size exclusion column. The column was run at a flow rate of 40 μ l/min in the following buffers:

- 20 mM Tris-HCl pH 7.0, 150 mM NaCl
- PBS (1mM KH₂PO₄, 10 mM Na₂HPO₄, 135 mM NaCl, 2.7 mM KCl pH 7.4)
- 20 mM Tris-HCl pH 7.0, 500 mM NaCl
- 20 mM Tris-HCl pH 7.0, 10% (v/v) ethanol

The void volume (V_0) was determined with blue dextran (2 MDa; Fluka, Neu-Ulm, Germany), the total volume of the liquid phase (V_t) with water. The apparent M_r values were extracted from plots of $\log(M_r)$ against K_{av} ($(V_e - V_0)/(V_t - V_0)$, in which V_e is the analyte elution volume).

For the glutaraldehyde crosslinking *tmNusB* at a concentration of 20 μ M in 20 mM Tris-HCl pH 7.0, 150 mM NaCl or in PBS, was mixed with glutaraldehyde (final concentration 1 mM). At defined times aliquots of the reactions were mixed with an equal volume of SDS loading buffer, briefly heated to 95°C and frozen in nitrogen stream. Samples were analyzed on a 15% SDS-PAGE.

4.2 Crystallography

4.2.1 Phasing

4.2.1.1 The Phase Problem

The electron density, which describes the surface feature and overall shape of all molecules in the unit cell, is a three-dimensional periodic function. It can be described as a Fourier series over all structure factors $F(h,k,l)$ implying that it specifies a value ρ for electron density at every position x,y,z of the unit cell (eq. 1).

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h, k, l)| e^{i\alpha(h, k, l)} e^{-2\pi i(hx + ky + lz)} \quad (\text{eq. 1})$$

While the structure factor amplitude $|F(h,k,l)|$ per reflection (h,k,l) can be derived from measured intensity, the $\alpha(h,k,l)$ phase angle per reflection (h,k,l) is not available through the diffraction experiment. There are a number of ways for solving this phase problem and the available methods include among others:

- Patterson Search or Molecular Replacement (MR)
- Multi-wavelength Anomalous Dispersion (MAD)
- Single/Multiple Isomorphous Replacement (SIR/MIR)
- Direct Methods

For the NusA AR1- λ N complex the method used to solve the phase problem was the MAD method. The availability of a three-dimensional structure of homologous proteins in Qor and *tmNusB* allowed solving the crystal structures by the Patterson search method.

4.2.1.2 The Patterson Function

The Patterson function $P(u,v,w)$ is the Fourier transform of $F^2(h,k,l)$ and can therefore always be calculated from the recorded diffraction intensities (eq. 2)

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l |F^2(h, k, l)| e^{-2\pi i(hu + kv + lw)} \quad (\text{eq. 2})$$

The coordinates (u,v,w) locate a point in the Patterson map in the same way that coordinates (x,y,z) locate a point in an electron density map. Thus, the Patterson function gives a map that contains peaks at the distance vectors between atoms in the structure. The peak height in the Patterson map is proportional to the product of the heights of the peaks in the electron density map.

4.2.1.3 Patterson Search or Molecular Replacement (MR)

The Patterson search method can be used when a model for a reasonably large fraction of the structure in the crystal is available. Thus, a model that is fairly complete and shares at least 30% sequence identity with the unknown structure can be used to calculate the starting set of phases. To carry out molecular replacement, the model structure must be placed in the correct orientation and position in the unit cell of the new protein. For the positioning of the model, rotation and translation functions are calculated as correlations functions between observed and calculated Patterson maps. Once the model is correctly positioned, the phase angles can be calculated from this model and combined with the observed amplitudes to give an initial electron density map of the structure of interest. This model will then have to be carefully analyzed and, through several refinement and model rebuilding steps, modified to account for the differences between the model and the target molecule, in order to maximize the agreement with the experimental data.

4.2.1.4 Multi-wavelength Anomalous Dispersion (MAD)

Scattering is the result of the interaction between electromagnetic waves, the X-rays, and electrons. When an electromagnetic wave collides with a system of electrons it induces an oscillation at the same frequency as the incident wave. The oscillating electrons behave as radiation scatterers emitting radiation at the same frequency as the incident radiation but with a phase difference of π . For heavy atoms the energy of the absorbed X-ray photon promotes the transition from a ground state orbital to an excited state, and the corresponding electronic acceleration leads to the anomalous scattering. The characteristic energy for such a transition is the absorption edge. When the X-ray wavelength is near the heavy atom absorbing edge a fraction of the radiation is absorbed by the heavy atom and re-emitted with altered phase. The shift in amplitude and phase is the anomalous scattering. The largest changes in the anomalous scattering factors of atoms occur around characteristic absorption edges since their values are directly related to the atom's atomic absorption coefficient. This is convenient since the majority of heavy atoms which bind to proteins have absorption edges lying in the energy range typically used for protein diffraction work (5-20 keV).

A resonant modulation of X-ray scattering accompanies the dispersion of X-ray absorption; these anomalous scattering increments to the normal scattering include an imaginary part, f'' , that is proportional to the X-ray absorption spectrum and a real part, f' , which is related to f'' . Due to the dispersion and absorption effects the Friedel's law is not anymore valid ($hkl \neq -h-k-l$). The non-equivalence of the Friedel pairs (reflections having identical structure factor amplitudes and phase angle but reverse signs) can be used to search for the anomalous scatterers' positions. Once the heavy atom positions have been identified, the vector corresponding to the difference of anomalous scattering from the two wavelengths can be calculated. Therefore, to conduct a MAD experiment, variable-wavelength synchrotron radiation has to be used to obtain phase information.

4.2.2 Crystal Structure of the Quinoline 2-Oxidoreductase

4.2.2.1 Crystal Growth and Data Collection

Prior to crystallization purified Qor was concentrated up to 10 mg/ml by ultra-filtration (Ultrafree, Millipore, Eschborn) in 50 mM Tris-HCl pH 8.0, 150 mM NaCl. Crystallization trials were carried out using standard material and solutions, homemade incomplete factorial solutions and sparse-matrix screens (Hampton Research, Laguna Hills, USA) at 20°C and 4°C. All crystallization experiments were performed with the sitting drop vapor diffusion method with drop volumes from 2 to 4 μ l and 500 μ l reservoirs. Small crystals were obtained after two weeks at 4°C in 1.5 M ammonium sulfate pH 8.5. A fine screen of the pH and of the precipitant concentrations led to a further improvement of the crystal quality.

The diffraction quality of the crystals was tested on an in-house image plate detector (MAR Research) coupled to a RU-200 rotating anode X-ray generator (Rigaku, Tokyo, Japan) producing CuK α radiation with a wavelength of 1.5148 Å. For data collection using synchrotron radiation, the crystals were harvested in the cryoprotectant buffer (0.1 M Tris-HCl pH 8.4, 1.4 M ammonium sulfate, 30% glycerol) and flash cooled in a nitrogen stream at 100 K (Oxford Cryosystems).

The crystals belonged to space group C2, with unit cell constants of $a = 278.32$ Å, $b = 72.10$ Å, $c = 202.65$ Å and two molecules were found in the asymmetric unit and a Matthews coefficient of $V_m = 2.86$ Å³/Da, corresponding to a solvent content of 56.6% was calculated (Matthews, 1968).

4.2.2.2 Structure Determination

The availability of the three-dimensional structure of the homologous protein *oc*CODH (PDB accession code 1N5W) made possible the determination of the Qor crystal structure by the Patterson search method (section 4.2.1.3) using the program AMoRe (Navaza, 1994). Rigid body and positional refinements of the model were performed using CNS (Brunger et al., 1998). Density modification by two-fold averaging was applied as implemented in RAVE (Jones, 1992). A single round of electron density averaging between two monomers was sufficient to improve the initial $2F_o - F_c$ electron density map and allowed the stepwise building and repositioning of side chains in MAIN (Turk, 1996). Refinement proceeded with

standard protocols (positional- and B-factor refinement) in CNS (Brunger et al., 1998) using a non-crystallographic symmetry (NCS) restraint. The last refinement steps were carried out without NCS restraints.

4.2.3 Crystal Structure of the NusA AR1- λ N Complex, and NusB

4.2.3.1 Crystal Growth and Data Collection

NusA AR1-AR2 was used at a concentration of about 8 mg/ml in 10 mM Tris-HCl pH 7.0, 2 mM DTT and mixed with a ten-fold molar excess of the synthetic peptide λ N(34-47) in ddH₂O. *tm*NusB at a concentration of about 5 mg/ml in 10 mM Tris-HCl pH 7.0, 100 mM NaCl was used for initial crystallization trials. All crystallization trials were performed with the sitting drop vapor diffusion method using drop volumes of 2-4 μ l and 500 μ l reservoirs of the screening solutions. Crystallization conditions were screened over a broad range of solutions (commercial sparse-matrix screens (Hampton Research, Laguna Hills, USA) and homemade incomplete factorials) at 20°C.

After refinement of pH, precipitant concentration, and temperature, crystals of the *tm*NusB and of the NusA- λ N complex were obtained. The five crystal forms of *tm*NusB and NusA- λ N crystals were frozen in a nitrogen gas stream at 100K (Oxford Cryosystems, Oxford, UK) after transfer into perfluoropolyether. The NusA- λ N crystals were soaked for one day in 10 mM potassium KAu(CN)₂ before transfer to the cryoprotection solution. Native and anomalous data were collected at the beamline BW6 (DESY, Hamburg). A peak and a remote wavelength near the Au absorption edge were selected on the basis of the fluorescence spectrum of a NusA- λ N crystal. Indexing, integration, and reduction of the diffraction data were carried out with DENZO/SCALEPACK (Otwinowski and Minor, 1997).

4.2.3.2 Structure Determination

The NusA AR1- λ N complex was solved by MAD (section 4.2.1.4). Heavy atom sites for the NusA- λ N crystal were determined with SHELXD (Schneider and Sheldrick, 2002) using the anomalous differences in the peak wavelength data. Programs from the CCP4 suite (Bailey, 1994) were used for phase calculations (MLPHARE) and density modification (DM).

ARP/wARP (Morris et al., 2003) was employed to generate a first atomic model based on the experimental $2F_o - F_c$ electron density map.

The structure of *tmNusB* in the P3₁21 form with the highest resolution could be solved by the Patterson search method (section 4.1.2.3) using MOLREP (CCP4; Bailey, 1994) with a monomeric poly-alanine model of *mtNusB* (PDB accession code 1EYV). The refined *tmNusB* structure including side chains but excluding water oxygens was subsequently employed to solve the other four structures by molecular replacement.

Further manual model building was conducted with the program MAIN (Turk, 1996). The models were refined by standard procedures with CNS (Brunger et al., 1998). The final round included a TLS-refinement step with REFMAC5 (CCP4; Bailey, 1994) with different anisotropic temperature factor corrections for space groups, which contained two molecules per asymmetric unit, and for the two AR1 molecules and the λ N peptide.

4.2.3.3 Molecular Modeling Studies

Protein sequences were aligned using BLAST (Altschul et al., 1997) and homology modeling was performed using the Homology package of Insight II (Accelrys, San Diego, USA) with the sequence of AR1 (residues 353-419) as template. Relaxation of all protein atoms except atoms of the secondary structure elements was allowed during minimization. Energy minimization was performed using the Discover module of Insight II for 9990 cycles employing a steepest descent gradient until convergence was reached, as judged by root mean square of the energy gradient (average derivative $< 0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$).

4.2.4 Structural Analysis and Graphical Representation

The geometric quality assessment on the refined model was done using PROCHECK (Laskowski et al., 1993), and superposition of coordinate files was done with the program TOP3D (CCP4; Bailey, 1994).

Illustrations of the structures were created with MOLSCRIPT (Esnouf, 1997), RASTER3D (Merritt and Bacon, 1997) and PyMOL (DeLano, 2002). Surface representations were prepared with GRASP (Nicholls et al., 1991). The sequence alignment was performed using ClustalW version 1.81 (Aiyar, 2000) and displayed with ALSRIPT (Barton, 1993).

5 Results and Discussion

5.1 Crystal Structure of the Quinoline 2-Oxidoreductase

5.1.1 Protein Biochemistry

For the isolation of the enzyme, the bacteria were cultured in an 8 liter fermentor (section 4.1.1.1). The amount of cells obtained was 29 g (wet weight).

With the purification scheme (section 4.1.1.2), from 29 g of cells 24.5 mg of highly active protein was obtained. The protein eluted as a single peak from the gel filtration column and the SDS-PAGE revealed an approximately equal amount of the three subunits with the expected molecular masses (Figure 15). The availability of a relatively large amount of protein with a high degree of purity and molecular homogeneity is an essential prerequisite for crystallization.

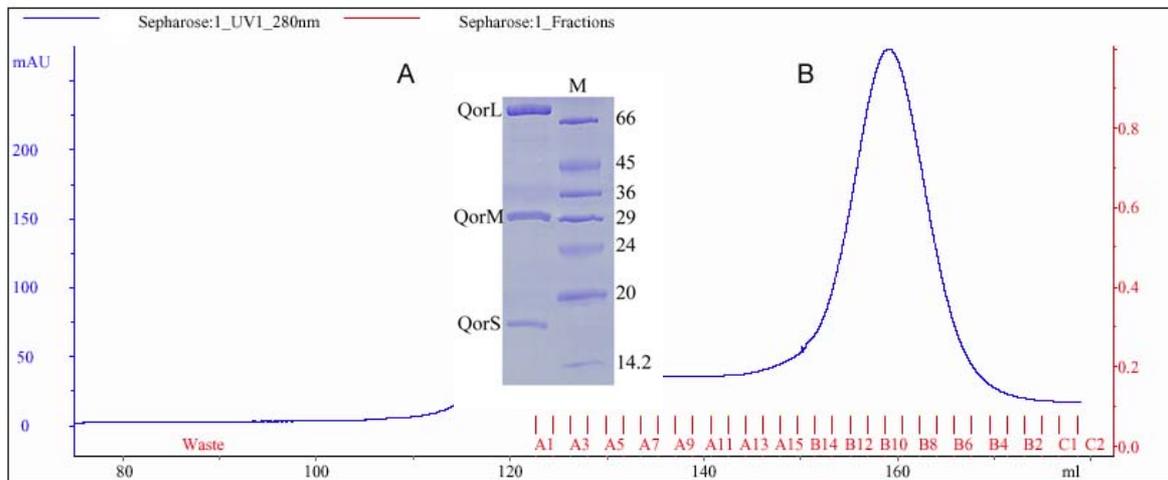


Figure 15. Qor purification. **A.** SDS-PAGE from the concentrated fractions (left). The three subunits have the following molecular masses: QorL (85 KDa), QorM (30 KDa), QorS (20 KDa). **B.** Qor elutes as single peak from gel filtration column Superdex S-200.

5.1.2 Crystal Growth and Data Collection

The initial crystallization trials were carried out using commercial and homemade incomplete factorial solutions at 20°C and 4°C. At 20°C crystals were obtained as small needles and could not be improved. At 4°C however, the first crystals obtained in 0.1 M Tris-HCl pH 8.5, 1.5 M ammonium sulfate, 12% glycerol and had already a three-dimensional appearance. After refinement of the chemical parameters, the optimized condition was found to be: 0.1 M Tris-HCl pH 8.4, 1.4 M ammonium sulfate, 12% glycerol with protein to reservoir ratio 3:1. The optimized growing time of the crystals was two weeks at 4°C, reaching the dimensions of roughly $400 \times 250 \times 50 \mu\text{m}^3$ (Figure 16).

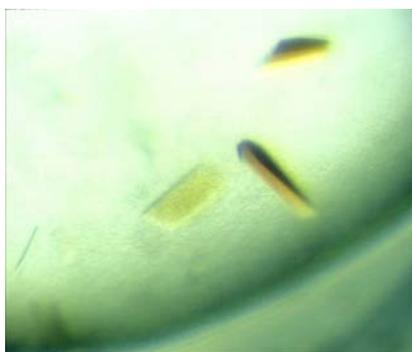


Figure 16. Crystals of Qor. Grown from 0.1 M Tris-HCl pH 8.4, 1.4 M ammonium sulfate, 12% glycerol at 4°C

In searching for the cryoprotectant solution the crystal mother solution was supplemented with increasing amount of glycerol. An amount of 30% glycerol made flash cooling possible. Upon cooling in the nitrogen stream (Oxford Cryosystems) the crystals diffracted to 1.8 Å using synchrotron radiation (Figure 17).

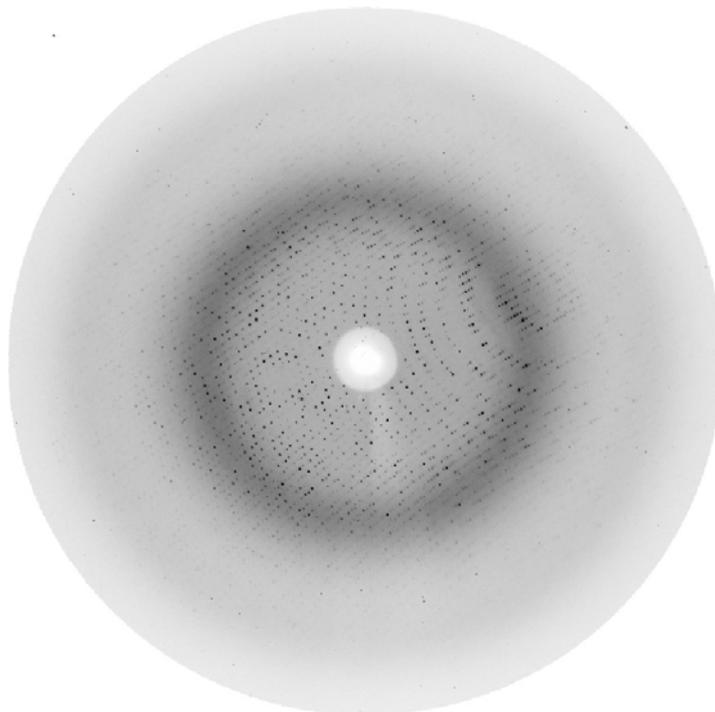


Figure 17. Diffraction image of Qor crystals. The image was recorded on a MarCCD detector on beamline BW6 at DESY, Hamburg with a rotation of 0.25° during exposure and a mosaicity of approximately 0.21° . The detector edge corresponds to a limiting resolution of 1.76 \AA at an X-ray wavelength of 1.05 \AA .

The crystal belonged to the space group C2 with cell dimensions of $a = 278.32 \text{ \AA}$, $b = 72.10 \text{ \AA}$, $c = 202.65 \text{ \AA}$. Two molecules comprised the asymmetric unit and a Matthews coefficient of $V_m = 2.86 \text{ \AA}^3/\text{Da}$ corresponded to a rather high solvent content of 56.6%.

A high resolution (1.8 \AA) native data set was collected from a single crystal using a Mar Research CCD detector at beamline BW6 at DESY (Hamburg, Germany; statistics are depicted in Table 3). The data set collection at the wavelength of 1.05 \AA started with an oscillation range of 0.25° for the first 285 images, was further reduced to 0.20° to decrease the number of partial spots (Figure 17).

Indexing, integration and reduction of the diffraction data were carried out with XDS (Table 3). The structure of Qor was solved by the Patterson search method using the homologous protein, *oc*CODH (PDB accession code 1N5W). The search model consisted in a monomeric protein, with non-conserved residues replaced by alanine, comprising the three subunits and the cofactors.

5.1.3 Patterson Search and Model Refinement

Data in the resolution range between 15 and 3.5 Å were used for the rotation and translation functions computed with AMoRe. Using a Patterson integration range of 30 Å, three rotation peaks were obtained with two top peaks having a correlation coefficient (CC) of 9.4% and 9.0%, respectively. Using the first top peak of the rotation function, the translation function gave a top solution with a CC of 13.5% and an R-factor of 56.1%. In the next translation function, the number of molecules to search for was increased to two, and the second peak had a CC of 21.9% and an R-factor of 53.9%. The model comprising the two monomers was then subjected to rigid-body refinement to give a CC of 43.1 and an initial R-factor of 49.6%. The two monomers were reoriented using PDBSET.

Rigid body and a positional refinement of the model were performed using CNS. To facilitate model building, density modification by two-fold averaging was applied. A single round of electron density averaging between two monomers was sufficient to lead to an initial $2F_o-F_c$ electron density map that allowed the stepwise building or repositioning of side chains. Refinement proceeded with standard protocols (positional- and B-factor refinement) using a non-crystallographic symmetry (NCS) restraint. The last refinement steps were carried out without NCS restraints. The final model had an R_{work} of 18.6% (R_{free} of 20.7%) and good stereochemistry (refinement statistics in Table 3).

The complete main chains of both monomers and the vast majority of side chains, except for some surface-exposed flexible residues, were completely enveloped in the final $2F_o-F_c$ electron density map when contoured at 1σ . The final model contained all residues of the protein, except the first six N-terminal amino acids of the small subunit and the three and two C-terminal amino acids of the medium and large subunits, respectively (Table 3). In addition 2535 water molecules, 7 sulphate ions, and 9 glycerol molecules were included in the model.

RESULTS AND DISCUSSION

Data Collection	
Space group	C2
Unit cell parameters	
a, b, c (Å)	278.32, 72.10, 202.65
α (°)	90
β (°)	127.98
γ (°)	90
Resolution (Å)	1.80
Total / Unique observations	853819 / 290508
Redundancy	2.94
Completeness (%) ^a	99.0 (99.1)
Mean I/ σ ^{a,b}	8.66 (3.56)
R _{sym} (%) ^{a,c}	7.6 (28.9)
Refinement	
Resolution (Å)	20.0 - 1.80
R _{work} ^d	0.186
R _{free} ^d	0.207
Contents per a.u.	
Protein molecules	2
Protein atoms	18482
Water oxygens	2535
Sulfate ions	7
Glycerol molecules	9
Mean B factor (Å ²)	
Protein	24.05
Water	38.18
Sulfate	73.76
Glycerol	47.30
R.m.s.d. ^e from ideal geometry	
Bond length (Å)	0.006
Bond angle (°)	1.38
R.m.s.d. ^e B-factors (Å)	
Main chain bonds	1.04
Main chain angles	1.52
Side chain bonds	1.79
Side chain angles	2.48
Ramachandran plot	
Preferred	0.992
Add. allowed	0.007
Disallowed	0.001
Cruickshanks DPI	0.1025

^a Numbers in parentheses represent statistics in the highest resolution shell (1.9-1.8 Å).

^b Mean I/ σ is the mean signal to noise ratio, where I is the intensity of the reflection and σ is the error in the measurement.

^c $R_{sym} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$ where I is the integrated intensity of reflection h having i observations and $\langle I_h \rangle$ is the mean intensity of reflection h over multiple recording.

^d $R_{work} = \sum | |F_o| - |F_c| | / \sum |F_o|$ where F_o and F_c are the observed and calculated structure factors. R_{free} is calculated for 5 % randomly chosen reflections.

^e r.m.s.d. - root-mean-square deviation

Table 3. Data collection and refinement statistics for Qor.

5.1.4 Description of the Structure

5.1.4.1 Overall Structure

Two monomers of Qor were found in the asymmetric unit, which build up two independent physiological dimers by the crystallographic two-fold axis. The dimer has overall dimensions of $149 \times 111 \times 78 \text{ \AA}^3$ with an accessible surface area of 77240 \AA^2 (Figure 18). The dimer interface is formed by a head-to-head arrangement of the two large subunits. The distance of roughly 53 \AA (Figure 18) between the two molybdenum ions impairs possible cross-reaction between the two independent monomeric catalytic units. A similar arrangement of subunits has been found for all other molybdo-iron-sulfur flavoproteins displaying a butterfly-shaped dimer with two independent catalytic units.

The two monomers are basically identical with an r.m.s.d. of 0.22 \AA for C_α atoms. Each monomer contains three redox components, one Moco, two [2Fe-2S] clusters, and the FAD moiety. The cofactors are lined up in such a way that their respective distances are shorter than 15 \AA (Figure 18), as frequently found for redox enzymes (Page et al., 1999). The [2Fe-2S] clusters form the internal electron transfer pathway from Moco to FAD (Figure 18).

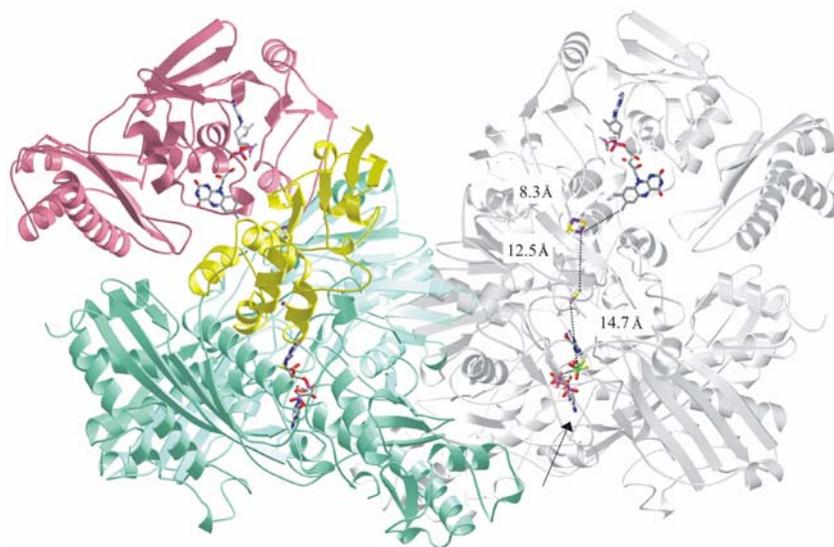


Figure 18. Ribbon plot representation of the Qor dimer and arrangement of its cofactors. The left monomer is colored according to the different subunits. The iron-sulfur containing subunit QorS is shown in yellow, the FAD-containing subunit QorM in pink and the Moco-containing subunit QorL in light blue. Cofactors are displayed as colored balls-and-sticks: carbon: grey, oxygen: red, nitrogen: blue, sulfur: yellow, phosphorus: magenta, iron: violet and molybdenum: green. The right monomer is displayed in gray. The shortest distances between the cofactors are shown as dashed lines and labeled with Å-values. Distances are taken from the molybdenum ion to the nearest iron in the type I [2Fe-2S] cluster, and from this to the nearest iron of the type II [2Fe-2S] cluster, and from this last iron atom to the C7 atom of FAD.

5.1.4.2 Structure of the Three Subunits

The S subunit or iron-sulfur protein

The iron-sulfur protein or S subunit (QorS, 162 residues, yellow colored in Figure 18) is located between the M and the L subunit. QorS has dimensions of $49 \times 39 \times 44 \text{ \AA}^3$ and can be further subdivided in two domains, each of them carrying one [2Fe-2S] cluster (Figure 19). According to EPR measurements the clusters were termed type I and type II cluster (Tshisuaka et al., 1993; Canne et al., 1997). The N-terminal domain (residues 7-82) is similar to plant-type [2Fe-2S]-cluster ferredoxins (Sticht and Rosch, 1998) and harbors the type II [2Fe-2S] cluster, which is localized next to the flavoprotein at approximately 8.3 \AA to the closest atom (C7) of the FAD. The iron-sulfur binding motif is $\text{CX}_2\text{CX}_{31}\text{CX}_1\text{C}$.

The C-terminal domain (residues 83-168) contains the type I [2Fe-2S] cluster which is buried approximately 11 \AA inside the protein. The [2Fe-2S] cluster is located within the loop

region of a four-helix bundle with pseudo two-fold symmetry as first shown for *dgALO* (Romão et al., 1995) and the binding motif is of the type $CX_4CX_2CX_{11}C$. The QorS structure has an additional C-terminal α -helix, which is positioned opposite to the loop region enclosing the cluster and runs parallel to the first helix of the bundle (Figure 19). The cluster is located at 14.7 Å from the molybdenum ion and at 12.5 Å from the nearest iron of the type II cluster, mediating the electron transfer from the Moco to the FAD (Figure 18).

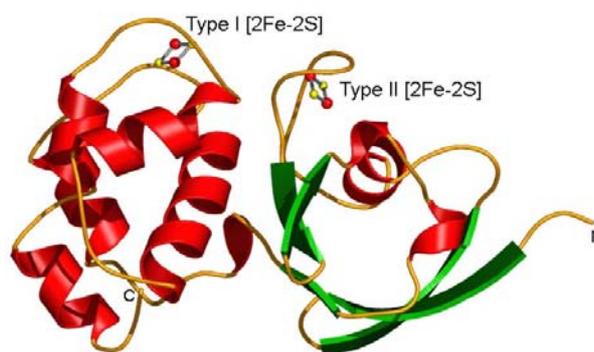


Figure 19. Ribbon plot representation of the iron-sulfur protein (QorS). The N-terminal domain contains the type II [2Fe-2S] cluster; the C-terminal domain harbors the type I [2Fe-2S] cluster. QorS is colored according to the secondary structure elements: α -helices red, β -sheets: green.

The sequence of QorS is homologous to the iron sulfur proteins of molybdenum hydroxylases, with a sequence identity of 50% to *hpCODH*¹, 49% to *ocCODH*², 41% to *ddALO*³, 39% to *dgALO*⁴, 34% to *rcXDH*⁵ and 31% to *btXO/XDH*⁶.

^{1,2} Carbon monoxide dehydrogenases from *Hydrogenophaga pseudoflava* (*hpCODH*) (Hänzelmann et al., 2000) and *Oligotropha carboxidovorans* (*ocCODH*) (Dobbek et al., 1999; Dobbek et al., 2002).

^{3,4} Aldehyde oxidoreductases from *Desulfovibrio desulfuricans* (*ddALO*) (Rebelo et al., 2000) and *Desulfovibrio gigas* (*dgALO*) (Romão et al., 1995; Rebelo et al., 2001).

^{5,6} Xanthine dehydrogenase from *Rhodobacter capsulatus* (*rcXDH*) (Truglio et al., 2002) and xanthine oxidase/dehydrogenase from *Bos taurus* milk (*btXO/XDH*) (Enroth et al., 2000).

The M subunit or flavoprotein

The M subunit (QorM, 285 residues, colored pink in Figure 18) is the flavoprotein of Qor with dimensions of $58 \times 52 \times 44 \text{ \AA}^3$. QorM binds FAD and can be divided into three domains (Figure 20). The N-terminal domain (residues 1-53) is composed of a three-stranded parallel β -sheet flanked by two α -helices (Figure 20). This domain contains the first FAD binding motif, $^{31}\text{AGGQS}^{35}$, termed as the glycine motif (Schulz, 1992). Although present in all members of the protein family, this binding motif is not strictly conserved. There is a single amino acid substitution (a histidine in place of glutamate) in *oc*CODH (Dobbek et al., 1999) and *hp*CODH (Hänzelmann et al., 2000), whereas the *rc*XDH (Truglio et al., 2002) maintains the AGG sequence and the *bt*XO/XDH (Enroth et al., 2000) has only one glycine. The middle domain (residues 59-173) is composed of a five-stranded antiparallel β -sheet and six small α -helices, which surround the cofactor at the adenosine dinucleotide part of the FAD (Figure 20). The second glycine motif (Schulz, 1992), $^{110}\text{TLGG}^{113}$ in Qor, is located in the sixth helix and is well conserved in the molybdenum hydroxylase family (TIGG in *oc*CODH, *hp*CODH and *rc*XDH, and SLGG in *bt*XO/XDH). The C-terminal domain (174-285) is formed by a three-stranded antiparallel β -sheet attached to a bundle of three α -helices (Figure 20). The C-terminal domain is not involved in FAD binding.

The sequence identity of QorM to the corresponding FAD binding proteins is 33% to *oc*CODH, 30 % to *hp*CODH, 23 % to *rc*XDH and 17% to *bt*XO/XDH.

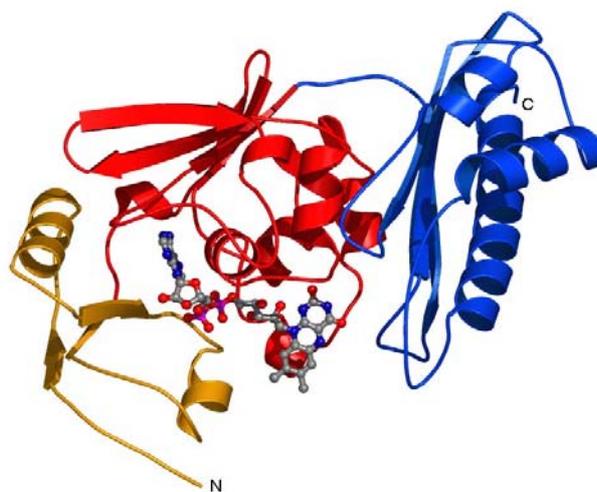


Figure 20. Ribbon plot representation of the flavoprotein (QorM). The N-terminal domain is colored yellow, the middle domain red and the C-terminal domain in blue.

The residues interacting with the FAD moiety are conserved among the molybdenum hydroxylases and can be seen as fingerprints in other families of FAD binding proteins (Dobbek et al., 1999). Using the program DALI (Holm and Sander, 1993), four proteins with similar fold to QorM were identified: *E. coli* UDP-N-acetylmuramate dehydrogenase (PDB ID code: 2MBR), cholesterol oxidase (PDB ID code: 1I19) from *Brevibacterium sterolicum*, *E. coli* D-lactate dehydrogenase (PDB ID code: 1F0X) and the vanillyl-alcohol oxidase from *Penicillium simplicissimum* (PDB ID code: 1QLT). All have been described as members of a novel family of structurally related oxidoreductases (Fraaije et al., 1998).

The four proteins show a comparable architecture for the FAD-binding domain and share the consensus glycine motifs ssGHs and shsG responsible for the binding of the adenosine-5'-biphosphate part of the FAD (Fraaije et al., 1998).

The isoalloxazine ring is the reactive part of FAD; it can accept two electrons and two protons or take up one proton and one hydride ion (Ghisla and Massey, 1989). In contrast to the adenosine-5'-biphosphate part, which is largely solvent accessible, the isoalloxazine ring is relatively well shielded within the QorM subunit, particularly by residue Tyr190 (Figure 21). This residue is located in an extended loop of the C-terminal domain. Residual density indicates that the loop can move so that the carbonyl oxygen of Gly188 would be at hydrogen bond distance to the N5 nitrogen of the isoalloxazine ring resulting in the shield of the C4a and N5 atoms by Tyr190 (Figure 21).

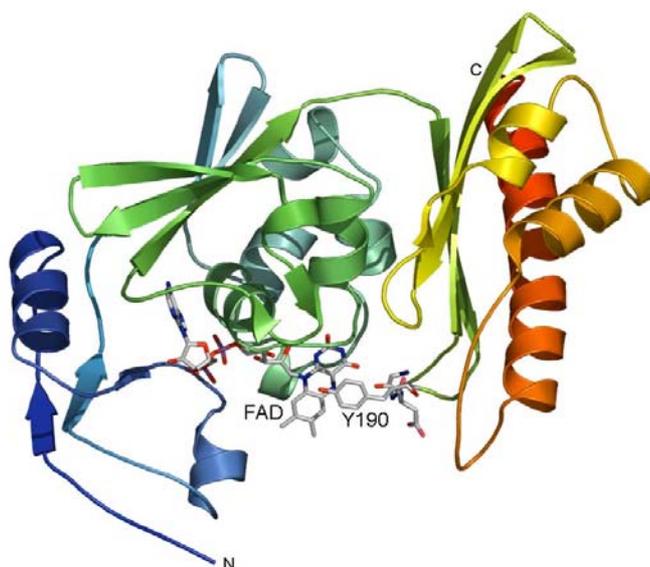


Figure 21. Shielding of the isoalloxazine ring of FAD from the solvent by Tyr190. Ribbon representation of the flavoprotein colored in rainbow with the N-terminus in blue and the C-terminus in red. The residues of the loop are labeled.

The L subunit or molybdoprotein

The L subunit (QorL, 786 residues, light blue in Figure 18) carries the active site with the Moco. The shape of QorL can be described as heart-like with overall dimensions of $91 \times 76 \times 62 \text{ \AA}^3$. The subunit can be subdivided in two domains running almost perpendicularly to each other (Figure 22). The N-terminal domain (residues 1-422) interacts with both QorS and the QorM subunits and its fold is dominated by four and five-stranded mixed β -sheets (Figure 22). This domain is mainly responsible for anchoring the Moco and contributes to the dimer interface by interacting with the QorL subunit of the other monomer. The C-terminal domain (residues 426-786) can be further subdivided in two subdomains. The first (residues 426-620) is characterized by a long central α -helix (residues 551-575) that runs into a three-stranded β -sheet. The second (residues 620-786) consists of three β -strands parallel to two α -helices with a third central α -helix running across the β -strands (Figure 22). This subdomain creates large parts of the dimer contact.

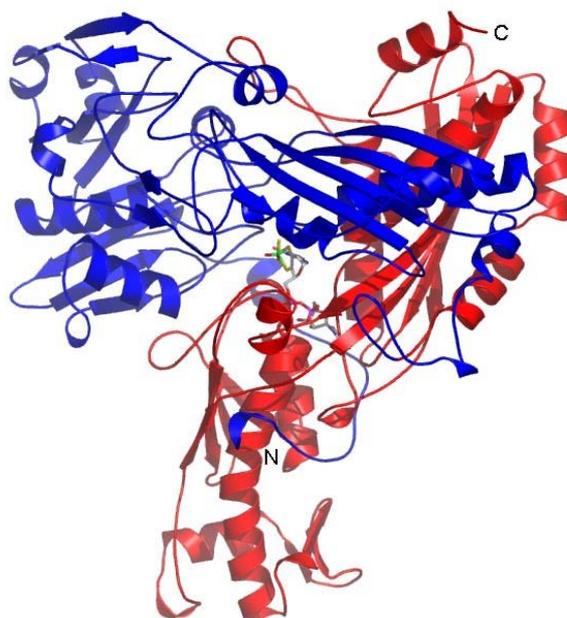


Figure 22. Ribbon representation of the molybdoprotein (QorL). The N-terminal domain is colored blue, the C-terminal domain is red.

The root mean square deviations for a pairwise comparison and 3D alignments of protein structures using Secondary Structure Matching (Krissinel and Henrick, 2003) have been calculated running the QorL structure against the whole PDB archive. The lowest r.m.s.d. for

C_{α} atoms values were obtained for the optimal superposition to *oc*CODH and *hp*CODH structures (1.58 and 1.59Å, respectively), followed by *dd*ALO and *dg*ALO (1.69 and 1.76Å, respectively) and the xanthine dehydrogenases *rc*XDH and *bt*XO/XDH (1.78Å and 1.97Å, respectively).

A multiple sequence alignment was performed using the protein sequence of QorL. The highest similarity in terms of sequence found for the *Arthrobacter nicotinovorans* nicotine dehydrogenase (Grether-Beck et al., 1994) followed by CODHs. QorL appears to be less similar to *dg*ALO sequence (Figure 23).

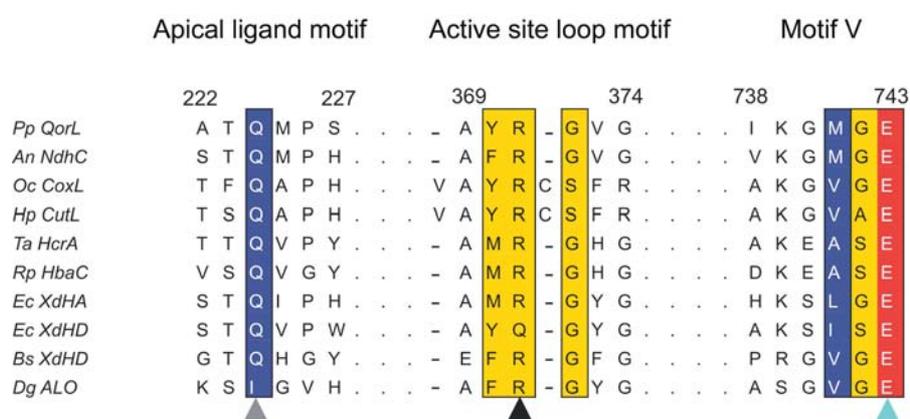


Figure 23. Alignment of ten sequences of bacterial molybdenum hydroxylases. Numbering corresponds to the sequence of QorL. The background of identical residues is red, that of highly conserved amino acids yellow, and that of residues with similar physico-chemical properties blue. Residues (Gln224, Arg371 and Glu743) involved in interactions with the Moco, belonging to the apical ligand motif, the active site loop and the Motif V (Hänzelmann et al., 2000) are indicated by grey, black and light blue arrows, respectively.

5.1.4.3 Substrate Channel

The substrate quinoline can gain access to the active site through a channel that is about 17 Å deep and 9 Å wide (distances relative to the surrounding C_{β} atoms). The entrance is located at the N-terminal domain of QorL near the dimer interface and has been marked by an arrow in Figure 24. The channel is build up by residues belonging to both the N-terminal and C-terminal subdomains. Mostly aromatic residues (e.g. Phe542, Phe228, Trp331, and Tyr370) create a hydrophobic environment for two-ring system of the substrate. Threonine residues

(Thr90, Thr231, Thr469, Thr481, and Thr482) situated near the entrance of the channel may contribute to the substrate entrance/product release.

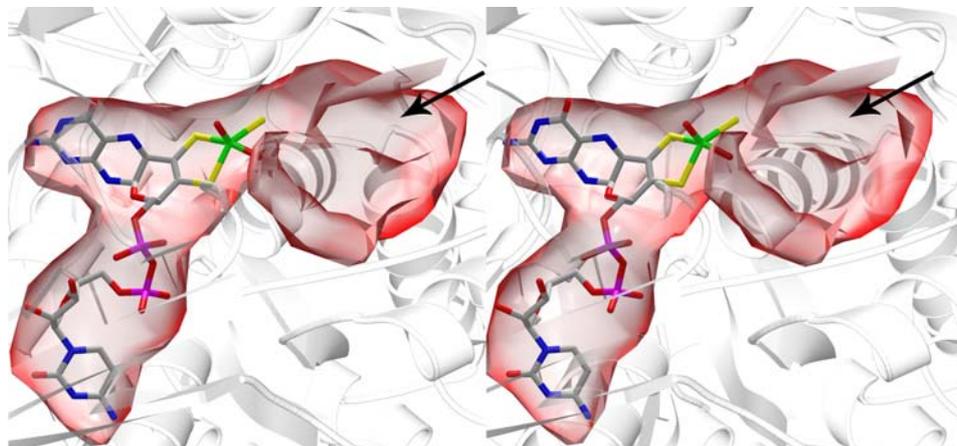


Figure 24. Stereo view of the electrostatic surface of the active site channel. The entrance of the channel is depicted by a black arrow. The channel surface was created with a probe of 1.4 Å radius. It is colored according to the electrostatic potential and rendered transparent. The side chains of residues Trp331 and Tyr545 were omitted from the calculation of the molecular surface for the purpose of a better view into the binding pocket. The Moco is shown in a ball-and-stick representation and colored like in Figure 18. The protein is represented in ribbons and rendered transparent.

The channel of CODHs (Dobbek et al., 1999; Hänzelmann et al., 2000) has a diameter of 6 Å consistent with the small dimensions of the respective substrate carbon monoxide. In the case of *dgALO* it was suggested that residues Phe425, Phe494, Leu497 and Leu626 might change their conformation to permit the different aldehyde substrates to gain access to the binding pocket (Romão et al., 1995). The structurally corresponding residues in Qor are small amino acids, such as Ala259, Thr332, Leu334, and Gly470, making the channel readily permeable to larger molecules like quinoline. In *rcXDH* and *btXO/XDH* the corresponding residues have similar sizes with the exception of His875 in *btXO/XDH* (L334 in Qor). Residue Gly470 of Qor lies on a coil region 4.6 Å apart from the corresponding residue (phenylalanine) in *rcXDH* and *btXO/XDH*, which results in a wider channel for Qor. TEI-6720, a potent inhibitor of *btXO/XDH*, binds in the channel leading to the active site and is stabilized by several conserved both neutral and hydrophobic residues (Okamoto et al., 2003). The structurally corresponding residues in Qor are threonine residues with the exception of Trp331 (Leu873 in *btXO/XDH*) and Val373 (Phe914 in *btXO/XDH*).

5.1.4.4 Qor active site. Moco and the Ligands around the Molybdenum Ion

The Moco is composed of a molybdenum ion with an oxidation state +VI in the air oxidized form and a molybdopterin cytosine dinucleotide. The ligands around the molybdenum ion can be described as adopting a distorted square pyramidal geometry (Figure 25). The molybdenum ion is coordinated by two ene-dithiolate sulfurs which are positioned in the equatorial plane together with an oxo- and a sulfido-ligand. The apical ligand was modeled as an oxo-group with a distance of 1.69 Å to the molybdenum ion.

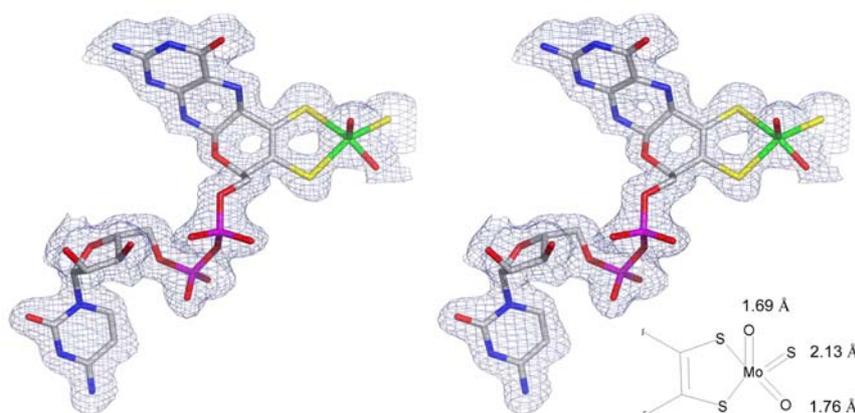


Figure 25. Stereo view of the Moco. The distances from the molybdenum ion to the two oxo- and the sulfido-ligands are the average between the values observed in the two monomers. The $2F_o-F_c$ electron density map contoured at the 1σ level is colored blue.

Still controversial for molybdenum hydroxylases is the nature of the apical ligand. In the two crystal structures of molybdenum hydroxylases, in which independent evidence could be gained for the position of the Mo-ligands (Qor and *oc*CODH), the sulfido-ligand was found in the equatorial position and an oxo-ligand was found in the apical position. In contrast, the re-sulfurated *dg*ALO structure (Huber et al., 1996) reveals an equatorial sulfido-ligand and an apical oxo-ligand.

To avoid model bias the three molybdenum ligands were treated from the beginning on as 8-electron atoms, corresponding to a fully occupied oxo- or a 50% occupied sulfido-ligand. Using this configuration the B value of the refined structure indicated a clearly lower B value for one of the equatorial atoms (average B_{SR1} : 19.7) compared to the other two oxo-ligand (B_{OM1} : 26.4, B_{OM2} : 29.7). In addition to the lower B value positive electron density around the equatorial ligand (difference Fourier map F_o-F_c at a level of 2.5σ remained visible) indicated

a lack of electrons in treating this ligand as an oxygen atom. Furthermore the bond length of this equatorial ligand refined to 2.13 Å (average distance in the two monomers, Figure 25), which is similar to the typical values of molybdenum sulfur double bonds of 2.18-2.19 Å (Thapper et al., 1999). After the equatorial ligand had been defined as a fully occupied sulfido-ligand the B values refined to 28.6 and no further residual density in the final difference Fourier calculations above a σ -level of 2.5 was detectable. Initially, the presence of a sulfido-ligand, for Qor was identified by cyanide inactivation and sulphide reactivation of Qor and by EPR spectroscopy of cyanide-inactivated Qor (Tshisuaka et al., 1993). New evidences in terms of a comparison between the *btXO/XDH* and sulfite oxidase by EPR spectroscopy was recently published (Peariso et al., 2003). According to the authors the oxo-ligand occupies the apical position in *btXO/XDH* during the enzymatic turnover, and moreover, the large proton hyperfine coupling detected in the EPR signal in the Mo^{V} state rapid type I could only be provided by a sulfhydryl proton positioned in the equatorial plane (Peariso et al., 2003). These results agree with the structural analysis of Qor and provide evidence for a conserved position of the sulfido-ligand in the family of molybdenum hydroxylases. The atypical apical position of the sulfido-ligand in *dgALO* may be an artifact of the drastic resulfuration conditions at molar concentrations of sulfide at very high pH or it may indicate a genuine structural mechanistic variation of bacterial aldehyde oxidoreductases.

5.1.4.5 Qor Active Site and Catalytic Pocket

Several residues are responsible for the stabilization of the cofactor in the active site. The residues which are involved in a direct interaction with the molybdenum-site (distance closer than 3.3 Å) are represented in Figure 26. Starting from the first coordination sphere the apical oxo-group (OM1) is in hydrogen-bonding distance to the $\text{N}\epsilon 2$ of Gln224 (distance of 2.9 Å) and to the backbone nitrogen of Gly256 (distance of 3.1 Å), both are highly conserved interactions. The equatorial oxo-ligand is in hydrogen-bonding distance to the backbone nitrogen of Ala546 (distance: 2.3 Å). Another highly conserved residue in Qor is Phe255, whose amide nitrogen is at hydrogen-bonding distance to the carbonyl atom of the pterin moiety (Figure 26). The rings of Phe255 and of the pterin are coplanar so that the pterin is stabilized by a π - π interaction.

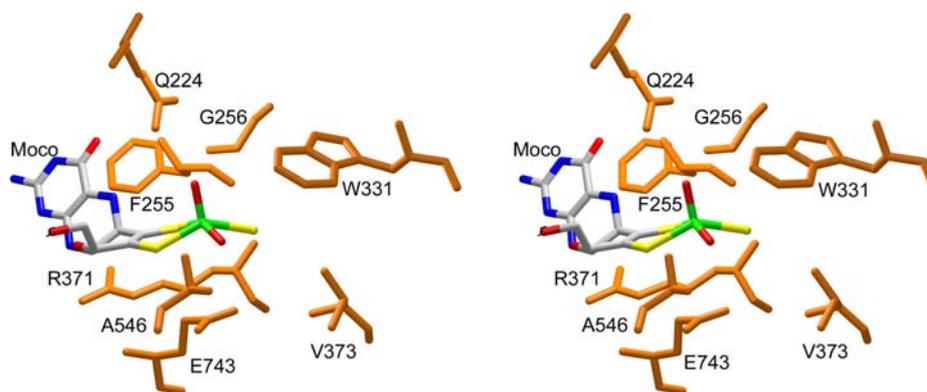


Figure 26. Stereo view of the active site residues. The Moco neighboring residues are colored in orange and labeled.

The residue Glu743 occupies the same position as in the CODHs and in *rc*XDH (Figure 27) but further away from the molybdenum ion (distance of 4.1 Å for Qor, 3.1 Å for *oc*CODH and 2.7 Å for *rc*XDH).

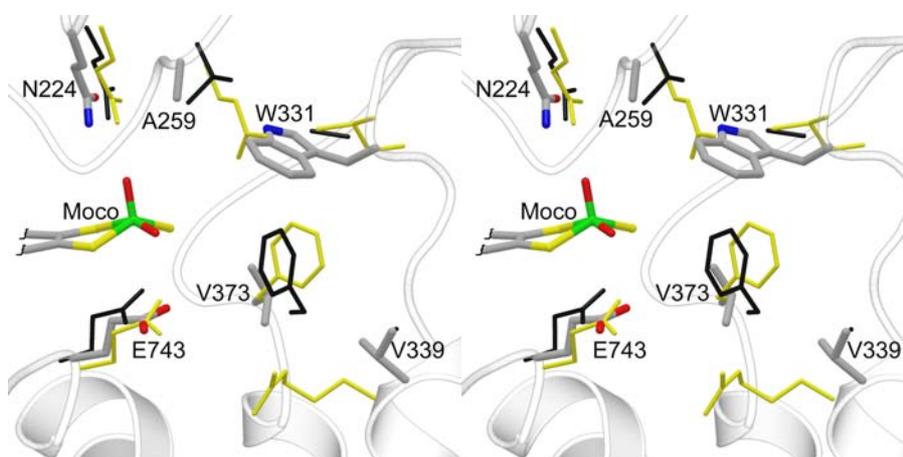


Figure 27. Stereo view of the active site superposition of Qor, *oc*CODH and *rc*XDH. Residues of Qor are colored as in Figure 18, of *oc*CODH black, and of *rc*XDH in yellow. Labels correspond to Qor.

The side chain of Arg371 adopts the same conformation in all structures, as its side chain extends towards the pyrazine ring of the Moco (distances of 3.7 Å and 3.2 Å between the guanido group and the N8 atom of the pyrazine ring). The backbone amide of Arg371 is at hydrogen-bonding distance to the sulfido-ligand (distance of 3.1 Å). This residue is conserved in the molybdenum hydroxylase family (black arrow in Figure 23) and is part of the active

site loop ³⁶⁹AYR-GVG³⁷⁴ in Qor. Among the molybdenum hydroxylases the consensus active site loop motif .AaR.sas (a: aromatic (Phe, Tyr, Trp), s: small (Gly, Ala, Ser) and . all amino acids) seems to vary according to the specific substrate (Hänzelmann et al., 2000). Qor and nicotine dehydrogenase have a very similar motif A(Y/F)R-GVG and can be grouped in one category, while the CODHs with their unique VAYRCSFR amino acid composition belong to a different one (Figure 23). The 4-hydroxybenzoyl-CoA reductases (Boll et al., 2001) display for the active loop the sequence AMR-GHG (*Ta HcrA* and *Rp HbaC* in Figure 23). *dgALO*, *ddALO* and aldehyde oxidoreductases can be grouped as containing the motif AFR-G(F/Y)G (*dgALO* in Figure 23). These sequences can, therefore, be considered as a key element and might be involved in determining the specificity of these enzymes toward their cognate substrate (Hänzelmann et al., 2000).

Interestingly the three amino acid residues Gln224, Arg371 and Glu743 maintain the same conformations in all the structures with the exception of the *dgALO/ddALO* structures where no equivalent of the glutamine residue has been found (Figure 23).

Although not directly participating in catalysis, the apical oxo-ligand may play an important role in the stabilization of intermediate states of the catalytic cycle by increasing the Mo=O bond strength by the so-called “spectator-oxo effect” (Rappe and Goddard, 1980). The preservation of the structural environment of the molybdenum ion suggests functional importance for the residues involved. It is remarkable that only *dgALO* and *ddALO* lack the glutamine residue (Figures 23 and 28A). The specific role of the highly conserved glutamine (Figures 23 and 28B) appears to be the stabilization of the oxo-ligand during biogenesis and catalysis to prevent its exchange against water or a sulfido-ligand by the formation of a hydrogen-bond. As *dgALO* lacks this stabilizing interaction for the apical oxo-ligand a replacement of this ligand against a sulfido-ligand may be allowed in the presence of high concentrations of sulfide.

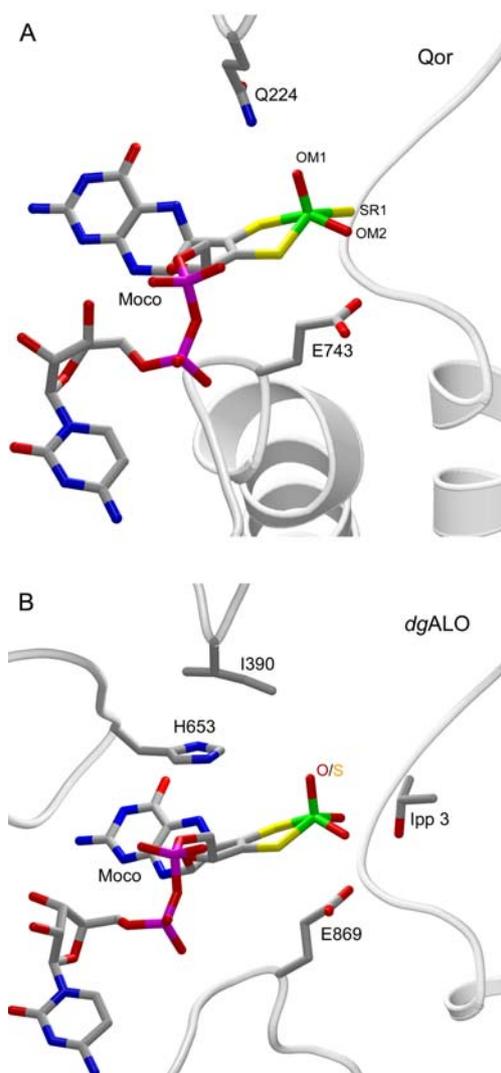


Figure 28. Comparison of the active sites of Qor and *dgALO*. **A.** Qor active site with the coordination of the apical oxo-ligand by the conserved Gln224. **B.** *dgALO* active site with isopropanol (Ipp 3) molecule in the putative substrate pocket. The O/S label indicates the oxo- to sulfido- substitution of the apical ligand during resulfuration (Huber et al., 1996). No stabilizing interactions are detected for the apical-ligand.

5.1.4.6 Putative Substrate Binding Mode

The refined structure of Qor displayed positive density in the active site of one monomer, into which either a glycerol molecule or an aromatic two-ring system could be modeled (Figure 29A).

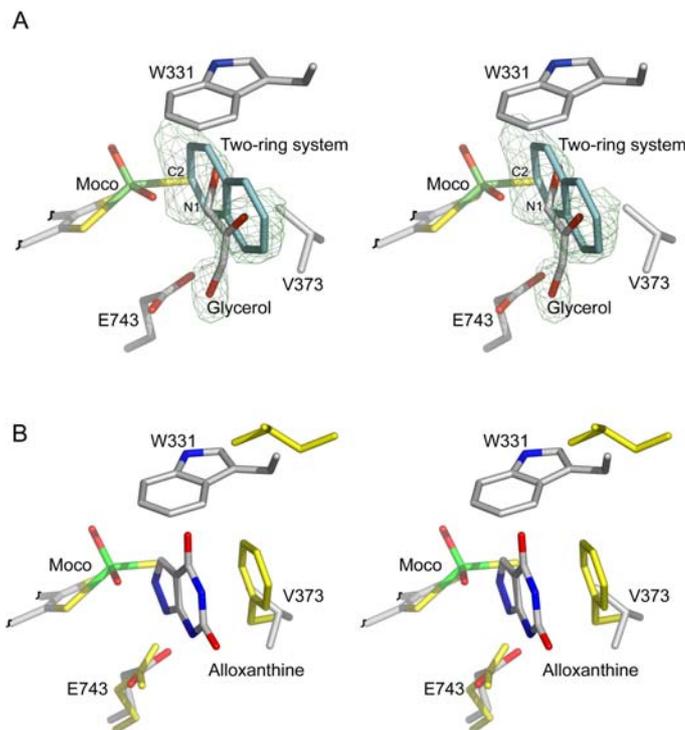


Figure 29. Stereo view of the observed positive density in the active site of Qor. **A.** The F_o-F_c map contoured at 3σ is colored green. Color code for side chains and glycerol as used in Figure 28. The modeled two-ring system is colored in light blue. Possible interacting residues with the two-ring system are labeled. **B.** Alloxanthine inhibited *rc*XDH is superimposed to the Qor active site.

The presence of glycerol is not surprising since 12% of glycerol was used as a precipitating agent. The two-ring system might represent the product of the reaction (2-oxo-1,2-dihydroquinoline) or a similar compound, originating from the enzyme purification. Other crystal structures of molybdenum hydroxylases showed substrate or product analogues bound in the active site as alloxanthine in *rc*XDH (Truglio et al., 2002), salicylate in *btXO/XDH* (Enroth et al., 2000), and isopropanol in *dg*ALO (Huber et al., 1996). Alloxanthine has been modeled into the active site of Qor by superposition with *rc*XDH structure (Figure 29B).

The active pockets of the two enzymes have different hydrophilic properties. Whereas, the two ligands in Qor are surrounded by Val373 and Trp331 (Figure 29A), alloxanthine in *rc*XDH is at hydrogen-bonding distance to Glu232 (Ala259 in Qor) and Arg310 (Val339 in Qor; Figures 28 and 30B). From the modeled two-ring system some important information can be gathered about the functional role of residues involved in substrate binding and catalytic turnover. The catalytic Glu743 is at 4.1 Å distance to the molybdenum ion. As seen in the case of *oc*CODH (Dobbek et al., 2002) a likely role of a deprotonated glutamate is to stabilize the +VI state of the molybdenum ion.

In Qor, Glu743 (Oε2 atom) is at hydrogen-bonding distance to the nitrogen atom N1 of the modeled two-ring system (Figure 29A). Based on the structure it would be reasonable that Glu743 contributes to substrate orientation, as no other side chains are found at hydrogen-bonding distance to the N1 atom. The nucleophilic attack of the equatorial oxo-group on the C2 atom is enabled by the short distance between the two atoms (2.3 Å). Furthermore an interaction between the C2 atom of quinoline and the equatorial sulfido-ligand is likely, as they are 2.6 Å apart, enabling a hydride transfer between them.

All other interactions with the two-ring system are of hydrophobic nature with Trp331 and Val373 being well positioned to align the aromatic two-ring system in the active site (Figures 28 and 29A). Val373 is a unique feature of Qor, as other structurally characterized molybdenum hydroxylases contain an aromatic residue (F/Y) at this position (Figure 28), which serves as a stacking partner for an aromatic substrate. Trp331 has no counterparts in the other enzymes. It is a leucine in both *btXO/XDH* and *rcXDH* (Figure 28) and a phenylalanine in *dgALO*. Its interaction with the two-ring system can be described as edge-on type comparable to the interaction of *btXO/XDH* with the salicylate ring (Enroth et al., 2000).

5.1.4.7 Active Site Protein Variants

Production of Qor mutant proteins by *P. putida* 86-1 Δqor pUF1m743a and *P. putida* 86-1 Δqor pUF1m743b was verified by PAGE and Western blot analysis. During purification, the Qor variants showed the same elution behavior in the chromatographic steps as wild-type Qor. In non-denaturing PAGE, the electrophoretic mobility of the purified Qor variants corresponded to that of the wild-type protein. Both QorE743V and QorE743D showed an about two-fold increase in K_m *app* (for quinoline) compared to Qor, suggesting that the mutations only slightly affect substrate affinity. In QorE743V, k_{cat} was significantly reduced (Table 4), indicating the catalytic relevance of Glu743. In contrast, the turnover number of QorE743D was reduced only 155-fold compared to Qor, suggesting that the carboxyl group of aspartate can - to a certain extent - take over the catalytic role of the functional group of glutamate. Structurally this behavior can be explained by the increased distance from the negative charge of the Glu/Asp residue to the counterpart molybdenum ion without significantly altering the given orientation.

The mutagenesis studies confirm a catalytic role for Glu743, since its replacement by a valine or an aspartic acid severely affects the catalytic efficiency. However these studies do

not support a crucial role of Glu743 in substrate binding as QorE743V shows only small changes in $K_{m \text{ app}}$ compared to the wild-type enzyme.

Qor protein	$K_{m \text{ app}}$ [quinoline] [mM]	$k_{\text{cat app}}$ [quinoline] [s ⁻¹]
Qor [recombinant]	0.12 ^a	85.4 ^a
QorE743D	0.25	5.6×10^{-1}
QorE743V	0.26	2.2×10^{-5}

^a (Frerichs-Deeken et al., 2003)

Table 4. Kinetic parameters of recombinant Qor and Qor variants.

5.1.4.8 Conclusions and Future Perspectives

In the present work the equatorial position for the sulfido-ligand was unambiguously identified, which is in contrast to the previous assignment to the apical position in *dgALO* and *rcXDH*. Recently, the 1.9 Å crystal structure of the *btXO/XDH* in complex with a “slow” substrate was published (Okamoto et al., 2004). In contrast to previous assignments, the sulfido-ligand was found at equatorial position in the square-pyramidal metal coordination sphere and in agreement with the sulfido assignment in Qor and *ocCODH*. The geometry of the molybdenum ligands with attributing the equatorial location to the sulfido-ligand puts it into a better position for hydride transfer from the substrate. This conclusion is consistent with the currently available spectroscopical and structural data. Thus, a sulfido-ligand occupying the equatorial position is a conserved feature in the molybdenum hydroxylase family.

Future work on Qor will include the attempt to obtain a three-dimensional structure in complex with substrate and inhibitors to achieve a better understanding of the catalytical mechanism of this molybdenum hydroxylase family member. Furthermore, studies of substrate-binding and -conversion by Qor will help to predict the feasibility to biologically oxidize other toxic organic compounds.

5.2 Crystal Structure of the NusA AR1- λ N Complex

5.2.1 Protein Biochemistry

5.2.1.1 NusA- λ N Binding Reactions

Isothermal titration calorimetry measurements were performed in order to estimate the binding between different NusA constructs and λ N or λ N-derived peptides (section 4.1.2.1). The chemical reaction created by each injection of the λ N peptide into the reaction cell, which contained the NusA protein, released a certain amount of heat (Figure 30). This was used to derive the changes in enthalpy and entropy, the magnitude of binding and the stoichiometry of the reaction. In the case of NusA λ N-binding could be detected for the full-length NusA and for the C-terminal region AR1-AR2. NusA lacking the C-terminal domain did not bind λ N. Furthermore, the ITC studies indicated for the NusA constructs binding to the peptide substrate in a micromolar range and in an equimolar stoichiometry (Figure 30).

In further experiments, different mutants of λ N were tested against NusA AR1-AR2. The λ N residues were systematically mutated to alanine to determine their specific involvement in AR1-AR2 binding. The obtained results indicated the participation of four side-chain residues, whose depletion completely abolished the binding, in AR1- λ N complex formation.

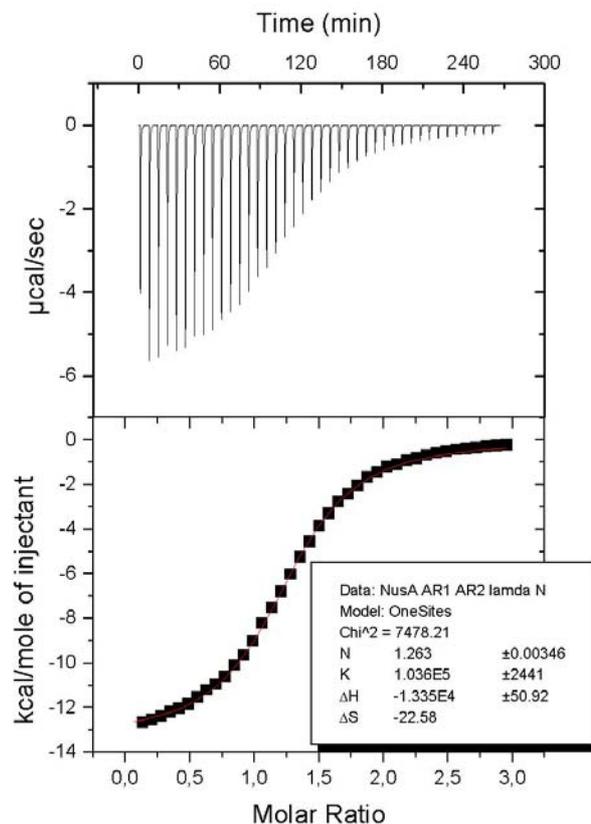


Figure 30. Binding of the λ N-derived peptide λ N(34-40) by NusA AR1-AR2. The negative peak indicates an exothermic reaction.

5.2.1.2 Analysis by Circular Dichroism Spectroscopy

CD spectroscopy has been used to determine the helical content of NusA AR1-AR2 (section 4.1.2.2). The spectra feature indicates a mixture of helical, sheet, and random coil structure (Finkelstein et al., 1977; Figure 31). The helical content was extrapolated from the CD spectrum using the program CDNN (Bohm et al., 1992). Analysis of the spectrum using the CDNN method suggested a helical content of 51% with a β -sheet content of 12% and 15% of β -turn (Table 5).

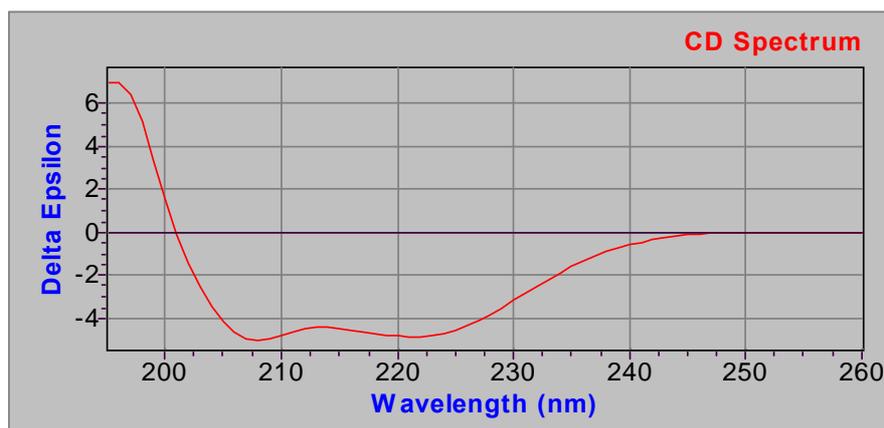


Figure 31. Circular dichroism spectra of NusA AR1-AR2. The protein sample was used in a concentration of 0.2 mg/ml in 50 mM Tris-HCl pH 7.0, 150 mM NaCl, 2 mM (*D,L*)-DTT.

	210-260 nm
Helix	51%
Beta sheet	12%
Beta turn	15%
Random coil	22%

Table 5. Secondary structure elements calculated for the NusA AR1-AR2 CD-spectrum.

5.2.2 Crystallography

5.2.2.1 Crystal Growth and Data Collection

For crystallization, the concentration of the NusA AR1-AR2 preparation was adjusted to about 8 mg/ml and the solution was mixed with a ten-fold molar excess of synthetic peptide λ N(34-47). Crystallization conditions were initially probed with a number of commercial and homemade incomplete factorial solutions at 20°C. After one year two crystals appeared in 2.0 M ammonium sulfate at pH 5.8. After refinement of pH, precipitant concentration, and temperature the optimized reservoir condition was found to be 0.1 M MES (pH 5.6) and 1.8 M ammonium sulfate. At 30°C the crystal growth time was reduced to two months. The crystals belong to the space group $P4_32_12$ with unit cell constants $a = b = 69.8 \text{ \AA}$ and $c = 67 \text{ \AA}$.

(Figure 32). There were two molecules per asymmetric unit of the unit cell and the Matthews coefficient of $V_m = 2.8 \text{ \AA}^3/\text{Da}$ corresponds to a solvent content of 55.2%.



Figure 32. Crystal of $(\text{NusA AR1})_2\text{-}\lambda\text{N}$. Crystals were grown from 0.1 M MES (pH 5.6) and 1.8 M ammonium sulfate at 30°C.

Crystals were soaked for one day in mother liquor supplemented with 10 mM potassium bis(cyano)aurate and could be flash cooled after transfer into perfluoropolyether. Multi-wavelength anomalous data from a single crystal were collected at beamline BW6 of DESY (Hamburg; Table 6) in 0.5° rotation frames. A preliminary X-ray fluorescence spectrum at the Au edge was used to select the monochromator settings at the peak ($\lambda_1 = 0.95 \text{ \AA}$) and at a remote ($\lambda_0 = 1.05 \text{ \AA}$). Bijvoet pairs to 2.3 \AA resolution were collected in a continuous 55° sweep for each wavelength, in order to optimize the measurement of the Bijvoet differences. The diffraction images were evaluated, merged, scaled and reduced with the programs DENZO/SCALEPACK.

5.2.2.2 Structure Determination

The dataset collected at the wavelength λ_1 was scaled to the data measured at λ_0 . The gold site was determined by analyzing the strong anomalous differences in the λ_1 data. The heavy atom position was refined and anomalous experimental phases at 2.3 \AA were calculated yielding to a final figure of merit of 0.85 after iterative solvent flattening, considering a solvent content of 50%. The program ARP/wARP was employed for the automatic model building based on the experimental electron density map that clearly showed secondary structure elements. Further manual model building was conducted in MAIN and the refinement was performed by standard procedures. The final round included a TLS-

RESULTS AND DISCUSSION

refinement step with using different anisotropic temperature factor corrections for the two AR1 molecules and the λ N peptide. Refinement converged at a $R_{\text{work}}/R_{\text{free}}$ -factor of 21.5%/24.0% for all data within 50 and 1.9 Å resolution. The overall stereochemistry of the structure was excellent, with no outliers in the Ramachandran plot (Table 6). The mean residual coordinate error estimated from Cruickshank's diffraction data precision indicator (DPI) amounted to 0.17 Å.

RESULTS AND DISCUSSION

<i>Data collection</i>	Remote	Peak
Space group	P4 ₃ 2 ₁ 2	
Unit cell (Å)		
a, b	69.8	
c	67.0	
Wavelength (Å)	1.05	0.95
Resolution (Å)	50.0-1.85	50.0-2.32
Unique reflections	26671	13192
Redundancy	2.4	2.6
Completeness (%)	98.5 (99.4)	97.8 (96.9)
I/σ(I)	17.8 (1.3)	23.8 (8.4)
R_{sym}^a (%)	3.6 (67.6)	3.4 (11.8)
Phasing		
Resolution (Å)	20.0-2.0	20.0-2.32
Heavy atom sites	1	
R_{Cullis}^b		
Isomorphous	0.79	
Anomalous	0.82	0.54
Phasing power^c		
Isomorphous	0.36	
Anomalous	1.45	3.09
FOM^d		
Before DM ^e	0.43	
After DM	0.85	
Refinement		
Resolution (Å)	50.0-1.90	
Model atoms		
Protein	1133	
Water oxygens	148	
Au ⁺	1	
R_{work}^f (%)	21.5 (25.6)	
R_{free}^f (%)	24.0 (36.6)	
R.M.S.D.		
Bond lengths (Å)	0.008	
Bond angles (°)	1.12	
Bonded B-factors (Å ²)		
Main chain	2.1	
Side chain	2.4	
Wilson B-factor (Å²)	30.5	
Model B-factors (Å²)		
Protein	35.2	
Water	56.1	
Au ⁺	43.1	
Cruickshanks DPI^g (Å)	0.17	
φ/ψ (%)		
Preferred	100	
Add. allowed	0	
Disallowed	0	

Table 6. Crystallographic data for the (NusA AR1)₂-λN complex.

Values for the last 0.1 Å in parentheses

- ^a $R_{\text{sym}}(\text{I}) = (\sum_{\text{hkl}} \sum_i [|I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle |] / \sum_{\text{hkl}} \sum_i [I_i(\text{hkl})]); I_i(\text{hkl})$ – intensity of the i^{th} measurement of hkl; $\langle I(\text{hkl}) \rangle$ – average value of hkl for all i measurements
- ^b $R_{\text{Cullis}} = \sum_{\text{hkl}} [| |F_{\text{PH}} \pm F_{\text{P}}| - |F_{\text{H,calc}}| |] / \sum_{\text{hkl}} [|F_{\text{PH}} \pm F_{\text{P}}|]; (F_{\text{PH}} + F_{\text{P}})$ if signs are opposite, $(F_{\text{PH}} - F_{\text{P}})$ if equal
- ^c Phasing Power = $(\sum_n [|F_{\text{h}}|^2] / \sum_n [|E|^2])^{1/2}$; $\sum_n |E|^2$ = lack of closure error = $\sum_n [|F_{\text{PH}}| (\text{obs}) - |F_{\text{PH}}| (\text{calc})]^2$
- ^d FOM = Figure of merit = $[|F(\text{hkl})_{\text{best}}|] / [F(\text{hkl}) |]; \mathbf{F}(\text{hkl})_{\text{best}} = \sum_{\alpha} [P(\alpha) \mathbf{F}_{\text{hkl}}(\alpha)] / \sum_{\alpha} [P(\alpha)]$
- ^e DM, density modification (solvent flattening)
- ^f $R_{\text{work}} = \sum_{\text{hkl}} [| |F_{\text{obs}}| - k |F_{\text{calc}}| |] / \sum_{\text{hkl}} [|F_{\text{obs}}|]; R_{\text{free}} = \sum_{\text{hkl} \in \text{T}} [| |F_{\text{obs}}| - k |F_{\text{calc}}| |] / \sum_{\text{hkl} \in \text{T}} [|F_{\text{obs}}|]; \text{hkl} \in \text{T}$ – test set.
- ^g $\text{DPI} = \text{sqrt}(N_{\text{atom}} / (N_{\text{refl}} - N_{\text{param}})) R_{\text{work}} D_{\text{max}} \text{compl}^{-1/3}$; N_{atom} – number of the atoms included in the refinement; N_{refl} – number of the reflections included in the refinement; D_{max} – maximum resolution of reflections included in the refinement; compl – completeness of the observed data

5.2.3 Description of the Structure

5.2.3.1 Crystal Content and Overall Structure

For crystallization a C-terminal NusA fragment comprising residues Ala350-Ala495, which covered the entire dual acidic repeat region AR1-AR2, and a synthetic peptide spanning residues 34-47 of λN have been used. Crystals were obtained only after prolonged incubation times. The structure of NusA- λN was solved by a two-wavelength multiple anomalous diffraction strategy after derivatization with $\text{KAu}(\text{CN})_2$ (Table 6). The experimental electron density map was straight forwardly interpreted in terms of two copies of AR1, termed AR1^A and AR1^B, in complex with a single λN fragment (Figure 33). The AR2 domain was not seen in the electron density. The AR1^A molecule encompasses residues Ala352 to Gln419, AR1^B consists of residues Ala352 to Gln421 (Figure 33). The ordered portion of the λN peptide spans residues Asn34 to Leu40 (Figure 33). The final $2F_o - F_c$ map covered the entire backbone regions and virtually all side chains of the refined model (Figure 33B).

only AR1 crystallized in complex with the λ N fragment, while AR2 remained in the supernatant fraction. Since their connection can be cleaved, the two C-terminal NusA domains are most likely linked by a flexible element. In addition, part of the peptide (residues Asn41-Arg47) must be disordered, because it is present, but invisible in the electron density.

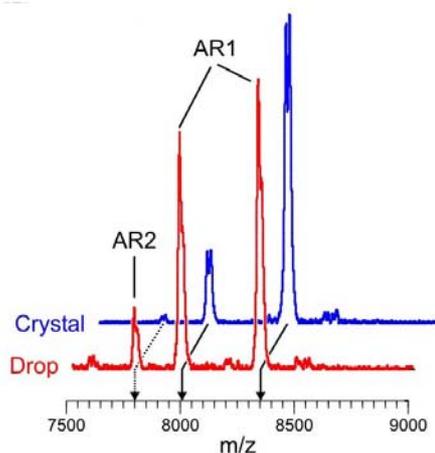


Figure 34. Mass spectrometric analysis on NusA AR1-AR2. Recorded mass spectra from the drop and the crystalline sample of the dual acidic repeat region.

5.2.3.2 Structure of the NusA Acidic C-Terminal Repeats

While the crystal structure of NusA covering its N-terminal domain (NTD) and the S1/KH region has been determined (Worbs et al., 2001), no experimental structure is available so far for the C-terminal repeats. The two crystallographically independent AR1 molecules of the current structure display an almost identical, all-helical fold (r.m.s.d. of 0.64 Å for all common C $_{\alpha}$ atoms). Five α -helices (α 1- α 5) are arranged as two perpendicularly packed helix-hairpin-helix (HhH) motifs (Figure 33C). The first HhH motif is made up of helices α 1 (residues 354 to 363) and α 2 (367 to 375), the second is built from helices α 4 (389 to 392) and α 5 (400 to 418). Helix α 3 serves as a connector (Figure 33C). The axes of the helices within the two pairs, α 1- α 2 and α 4- α 5, are almost parallel, which distinguishes the HhH motif from the approximately perpendicular arrangement of helices in helix-turn-helix motifs (Shao and Grishin, 2000). Duplication of a HhH motif to form a (HhH) $_2$ domain is also known from other proteins (Shao and Grishin, 2000). The structure of AR1 is consistent with previous motif predictions (Mah et al., 2000) and recent NMR assignments for the NusA C-terminal extension (Eisenmann et al., 2004).

A widespread function of HhH motifs is the mediation of sequence-independent interactions with DNA and RNA but HhH-type protein-protein interaction motifs are known as well (Shao and Grishin, 2000). Structural similarity searches identified both protein and nucleic acid interaction domains as closely related partners to AR1. The Secondary Structure Matching (SSM) server (Krissinel and Henrick, 2003) assigned the AR1 structure to the sterile alpha motif (SAM) fold. The SAM domain is an important protein-protein interaction module found in diverse protein families (Thanos et al., 1999). In contrast, the DALI server (Holm and Sander, 1993) identified regions in *Thermus filiformis* DNA ligase (PDB ID code: 1DGS; r.m.s.d. of superimposed C_{α} atoms 2.3 Å), human Ephb2 receptor SAM domain (1B4F; r.m.s.d. 2.9 Å), human DNA polymerase β (1BPY; r.m.s.d. 3.3 Å), *E. coli* RuvA (1HJP; r.m.s.d. 2.4 Å) and the *E. coli* 3-methyladenine DNA glycosylase I (1LMZ; r.m.s.d. 2.6 Å) as the most closely related motifs to NusA AR1. In all these proteins the HhH portions are known or expected to interact with nucleic acids (Sawaya et al., 1997; Nishino et al., 1998; Lee et al., 2000). The hairpin loops of these proteins, which are contacting the nucleic acids, are rich in positively charged residues and usually contain a consensus GhG (*h* – hydrophobic amino acid) fingerprint (Thanos et al., 1999; Shao and Grishin, 2000). In contrast, the surface of AR1 is decisively negatively charged (Figure 35) and the protein does not exhibit patches of positively charged side chains. Therefore, it is unlikely that AR1 constitutes a nucleic acid binding module besides its known function as a protein-binding domain.

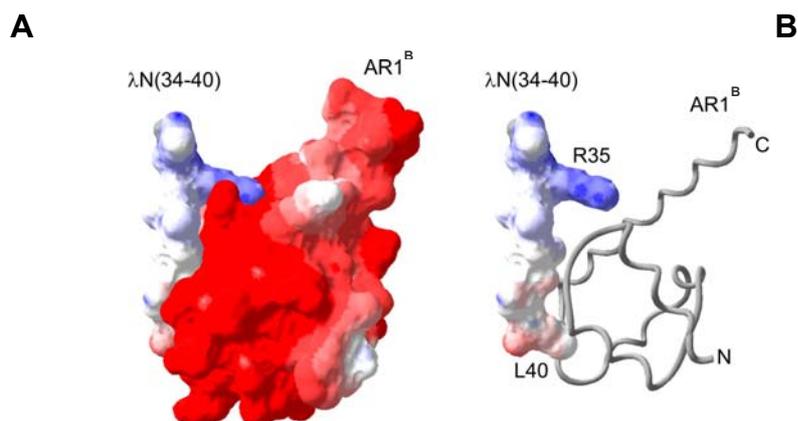


Figure 35. Electrostatic surface potentials of the binding partners. Negative potential is in red, positive potential in blue. **A.** The electrostatic surface potential was calculated separately for λ N(34-40) and NusA AR1^B. A charge complementarity at the level of the residue Arg35 can clearly be discerned. **B.** Surface representation of the λ N(34-40) peptide with NusA-AR1^B displayed as a grey ribbon.

Based on the crystal structure of AR1, a three-dimensional homology model of AR2 using the Insight II program package was devised. Consistent with the high sequence similarity (Figure 33A), the AR2 model closely resembled the AR1 fold with an r.m.s.d. of 1.33 Å between all common C_α atoms. It maintained a proper hydrophobic core during energy minimization and displayed conserved residues at similar positions to AR1. The surface electrostatic features of the AR2 model matched those of AR1. Consistently, the secondary structure content determined from a CD spectrum of the AR1-AR2 repeat (47% helix), largely corresponded to the helical content predicted from the crystal structure of AR1 and the model of AR2 (51% helix). Therefore, AR2 presumably harbors a very similar fold as AR1.

5.2.3.3 Structure of the (NusA AR1)₂- λ N Peptide Complex

In an asymmetric unit of the present crystal structure, two AR1 molecules are associated through an antiparallel apposition of the N-termini with their α 5 helices (Figure 33C). This arrangement leaves the two α 1- α 2 loops of the domains facing each other and running in opposite directions. The λ N peptide (residues Asn34 to Leu40) is sandwiched between these loops (Figure 33C). Peptide residues Ile37-Ser39 build up a short parallel β -sheet like arrangement with residues Asp364 to Asp366 of one AR1 domain and engage in a short antiparallel β -sheet like association with the same residues of the other AR1 molecule (Figure 33C). Thus, the same regions of the two NusA fragments are engaged in the contacts to the

peptide, but these interactions aim at opposite surfaces of λ N. Upon complex formation about 700 Å² of combined surface area are covered by the AR1^A- λ N contact, while about 850 Å² are buried in the AR1^B- λ N complex.

The peptide adopts an extended conformation in the complex. Thus, it exposes a large number of chemical functionalities to the environment explaining how a rather short stretch of λ N suffices to build up a specific complex with NusA AR1. The conformation of the present peptide is quite different from that of the first 22 residues of λ N in complex with a *boxB* RNA hairpin. Although protein λ N is unfolded in isolation, its N-terminal region assumes an α -helical structure upon interaction with *boxB* RNA (Legault et al., 1998). Van Gilst and coworkers have shown that NusA interacts with an unfolded region of λ N and that the conformation of this region is independent of *boxB* RNA binding consistent with the extended peptide conformation of the present complex structure (Van Gilst et al., 1997).

The present structure indicates that AR1 comprises a major λ N binding site of NusA in agreement with other biochemical studies. Mah et al. demonstrated that a NusA fragment containing AR1 is necessary for binding to λ N, while the C-terminal 80 amino acids encompassing AR2 are dispensable (Mah et al., 1999). However, it could not be excluded that AR2 does engage in additional contacts to λ N, which were not reflected in the assays.

In order to test whether the observed AR1- λ N(34-40) interaction is stable in solution, electrophoretic gel mobility shift assays were conducted. A peptide corresponding to λ N(residues 1-40) efficiently bound to synthetic *boxB* RNA (Figure 36A). NusA AR1-AR2 alone did not interact with the RNA, even at very high concentrations, consistent with the acidic nature of the protein (Figure 36). However, NusA AR1-AR2 was capable of supershifting the λ N(1-40)-*boxB* complex (Figure 36A). These results indicate that region 41-47 of λ N is dispensable for rudimentary binding to NusA AR1-AR2.

RESULTS AND DISCUSSION

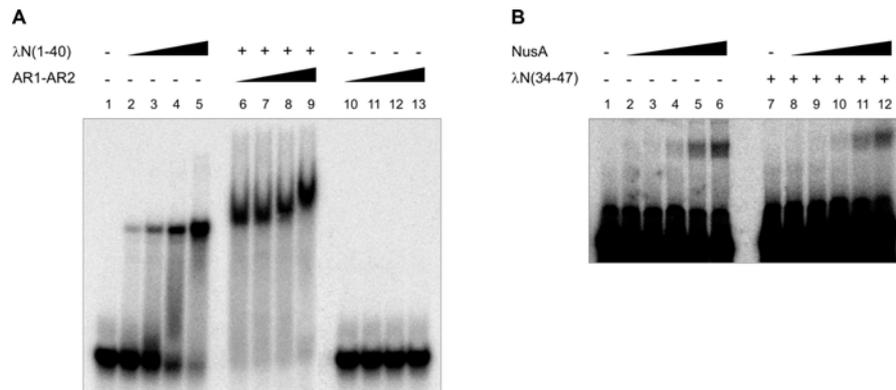


Figure 36. Electrophoretic gel mobility shift analysis on NusA AR1-AR2- λ N-*nut* RNA. **A.** Titration of radiolabeled *boxB* RNA with increasing concentrations (0-10 μ M) λ N(1-40) (lines 1-5). Titration of a λ N(1-40)-*boxB* complex with increasing concentrations (0.25-10 μ M) of NusA AR1-AR2. High concentrations of NusA AR1-AR2 protein can slightly supershift the λ N(1-40)-*boxB* complex (lines 6-9). Titration of radiolabeled *boxB* RNA with increasing amounts (0-10 μ M) of NusA AR1-AR2. No RNA binding for the AR1-AR2 region was detectable in this assay (lines 10-13). **B.** Titration of *nut* site RNA NusA with increasing concentration of full length NusA (0-20 μ M) in the absence of λ N peptide (lines 1-6) or in the presence of 100 μ M λ N(34-47) (lines 7-12). Binding of NusA to the RNA is weak irrespective of the presence of the peptide.

5.2.3.4 λ N Interaction is Stoichiometric in Solution

Quantitative Western blot analysis of antitermination complexes purified from *E. coli* extracts have suggested that two NusA molecules may be in contact with λ N during antitermination (Horwitz et al., 1987). These results could explain the binding of two AR1 modules to a single λ N peptide seen in the present crystal structure. Alternatively, only one of the observed AR1- λ N interaction modes may be physiologically relevant. In order to distinguish between these possibilities, the interactions between various NusA and λ N constructs in solution were studied by isothermal titration calorimetry (ITC). The amount of heat released after injections of peptides into solutions of NusA was used to derive the dissociation constants, enthalpies and entropies of the interactions, and the stoichiometries between the binding partners (Table 7; Figure 37).

RESULTS AND DISCUSSION

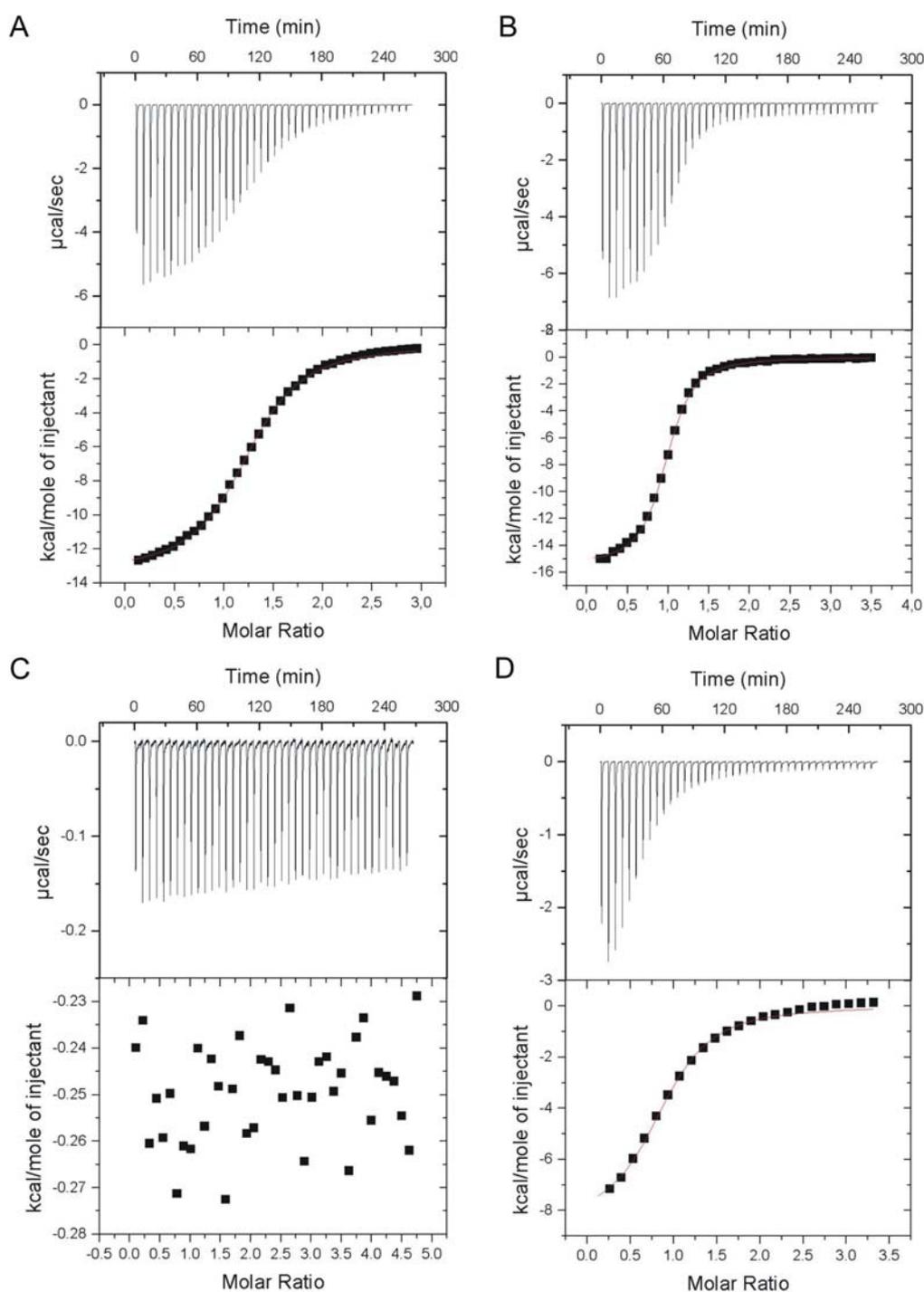


Figure 37. Probing of the NusA- λ N interactions by ITC. Panels display the reactions of NusA AR1-AR2 with peptides (A) λ N(31-43), (B) λ N(34-47), (C) mutant peptide L38A; no binding; and (D) of full length NusA with λ N(31-43). Top panels: Raw data plotted as power (μ cal/sec) versus time (min). The negative peaks indicate an exothermic reaction. The area under each peak represents the heat released following an injection of the peptide into the solution of NusA. Lower panels: Binding isotherms obtained by plotting peak areas against the molar ratio of λ N peptide to NusA AR1-AR2. The red lines represent the best-fit curves obtained from least squares regression analysis assuming a one-site binding model.

For these binding studies, a fragment spanning residues Lys31-Lys43 was chosen (Table 7). It expands the region, which in the crystal structure is in contact with AR1, by three residues on either side, in order to reduce end effects of the free N- and C-termini. First, it had to be verified whether the λ N peptides interacted within the C-terminal region of NusA in solution. The binding affinity of λ N(31-43) to the NusA AR1-AR2 region ($K_d = 9.6 \mu\text{M}$) was comparable to its affinity for full length NusA ($K_d = 7.9 \mu\text{M}$). In contrast, a NusA fragment covering only the NTD and the S1/KH region, but lacking AR1-AR2, completely lost its ability to bind the λ N peptide. Therefore, the AR1-AR2 region is exclusively responsible for binding λ N.

Because full-length NusA binds the peptide equally well or slightly better than AR1-AR2, the λ N binding site on the C-terminal repeats must be directly accessible in the framework of the full-length NusA protein. ITC results reveal that the enthalpic contribution to peptide binding is less favorable for full-length NusA than for the AR1-AR2 fragment (-8544 cal/mol vs. -13350 cal/mol, respectively), but the difference is made up by a more favorable entropic term (-5.8 cal/mol per K vs. -22.58 cal/mol per K, respectively). Thus, in full-length NusA, the AR1-AR2 module may be fixed by intramolecular contacts so that a loss in entropy upon interaction with the peptide is reduced as compared with free AR1-AR2. These results support the notion that the AR1-AR2 region interacts with the remainder of NusA, as suggested for the autoinhibition of RNA binding (Mah et al., 2000).

Approximately equimolar interaction stoichiometries were observed for various λ N peptides and NusA constructs, indicating that one NusA monomer binds one peptide substrate (Table 7; Figure 37). This finding contradicts the idea that the present crystal structure reflects an *in vivo* situation, in which two NusA molecules bind by means of AR1 to the same region of λ N. The NusA: λ N binding stoichiometry in antitermination complexes is an unresolved issue. In agreement with the here presented results, other studies have shown that NusA binds λ N in a 1:1 complex *in vitro* (Mogridge et al., 1995; Van Gilst and von Hippel, 1997; Xia et al., 2003). Furthermore, NusA binds stoichiometrically to RNAP in the absence of λ N both *in vitro* (Greenblatt and Li, 1981; Gill et al., 1991) and *in vivo* (Horwitz et al., 1987). Therefore, the 2:1:1 interaction of NusA: λ N:RNAP may result from the purification of the complexes under low salt conditions (Horwitz et al., 1987).

λ N Peptide	Sequence	K_d (μ M)	ΔH_a (cal/mol)	ΔS_a (cal/(mol per K))	Stoichiometry
λ N(34-47)	NRPILSLNRKPKSR	3.55 ± 0.11	-9452 ± 24.01	-27.77	0.965 ± 0.003
λ N(31-43)	KPVNRPILSLNRK	9.65 ± 0.23	-13350 ± 50.92	-22.58	1.263 ± 0.003
N34A	KPVARPILSLNRK	13.48 ± 0.23	-12700 ± 53.09	-21.21	0.862 ± 0.002
R35A	KPVNAPILSLNRK	Below DL	Below DL	Below DL	Below DL
P36A	KPVNRAILSLNRK	12.86 ± 0.24	-15340 ± 56.77	-29.96	1.087 ± 0.003
I37A	KPVNRPALSLNRK	Below DL	Below DL	Below DL	Below DL
L38A	KPVNRPIASLNRK	Below DL	Below DL	Below DL	Below DL
S39A	KPVNRPILALNRK	41.24 ± 0.46	-10650 ± 41.32	-16.25	1.229 ± 0.003
L40A	KPVNRPILSANRK	Below DL	Below DL	Below DL	Below DL

Below DL, Below Detection Limit

Table 7. Interactions of NusA AR1-AR2 with λ N peptides.

5.2.3.5 Identification of the Biologically Relevant Complex by Alanine Scanning

In the next series of experiments, mutated λ N(31-43) peptides were tested for binding to NusA AR1-AR2 (Table 7; Figure 37). Significantly, λ N residues, which upon mutation to alanine displayed reduced affinity, were found to preferentially or exclusively engage with domain AR1^B of the crystal structure (cyan in Figure 33C).

In detail, Asn34 is not in direct contact with either of the AR1 modules (Figure 38), and, consequently, binding of N34A is similar to that of the wild-type peptide (Table 7). The Arg35 side chain, alternatively, is deeply inserted into a crevice of AR1^B. It forms a salt bridge to O δ 2 of Asp364^B (superscript identifies the AR1 molecule) and hydrogen bonds to the carbonyl oxygen of Arg409^B (Figure 38). In addition, a water molecule mediates an interaction to the carbonyl atom of Tyr362^B. A second water molecule connects N ω 2 and the carboxyl function of Glu408^B. The intimate contacts to AR1^B are responsible for the complete loss of binding upon deletion of the Arg35 side chain (Table 7). Mutation on the following peptide residue, Pro36, to alanine shows only a marginal reduction in binding affinity (Table 7). Pro36 stacks on the aromatic ring of Phe369 from AR1^A (Figure 38). If this interaction was important, a more severe loss of affinity would have been expected, arguing against the significance of the λ N-AR1^A interaction. A proline at position 36 may be preferred slightly over other residues to facilitate a peptide conformation, in which its own carbonyl oxygen is in hydrogen bonding contact to the side chain of Arg409^B and at the same time the preceding

residue, Arg35, can occupy its binding pocket on AR1^B. This suggestion is corroborated by a more favorable entropy of interaction for the wild-type peptide compared to the P36A mutant (Table 7). Ile37 engages in very similar hydrophobic contacts to both AR1 molecules, so the loss of affinity of the I37A mutant cannot discriminate the relevant interaction. Replacement of Leu38 with an alanine results in complete loss of binding to NusA AR1-AR2 (Table 7). One side of residue Leu38 favorably interacts with Ile365 and Phe369 of AR1^B (Figure 38). At the same time, Leu38 points to the same side of the peptide as the Pro36 and Leu40 side chains. It thus seems to stabilize the extended peptide conformation through hydrophobic interactions with the latter two residues. Ser39 displays only a single specific contact, a hydrogen bond to Asp366 O δ 1 of AR1^B (Figure 38). The dissociation constant of the S39A mutant is about four-fold higher than that of the wild-type peptide (Table 7), again identifying AR1^B as the relevant interaction partner. The side chain of Leu40 exclusively interacts with AR1^B. It is in hydrophobic contact with Leu398^B and Val372^B and stacks on the aromatic ring of Phe369^B (Figure 38). A smaller residue would not be able to fill this hydrophobic pocket. In agreement, the corresponding L40A mutant has lost its affinity to NusA AR1-AR2 entirely (Table 7).

Taken together, the AR1^B- λ N complex of the current crystal structure constitutes a specific and biologically important interaction, because four of the seven peptide residues tested are absolutely required for NusA binding (Table 7). This finding is consistent with the larger surface area buried by the AR1^B- λ N complex. Furthermore, the electrostatic surface potentials of AR1^B and the peptide are compatible with an intimate interaction (Figure 35). In particular, the electropositive patch around Arg35 of the peptide fits nicely into an electronegative pocket of AR1^B.

RESULTS AND DISCUSSION

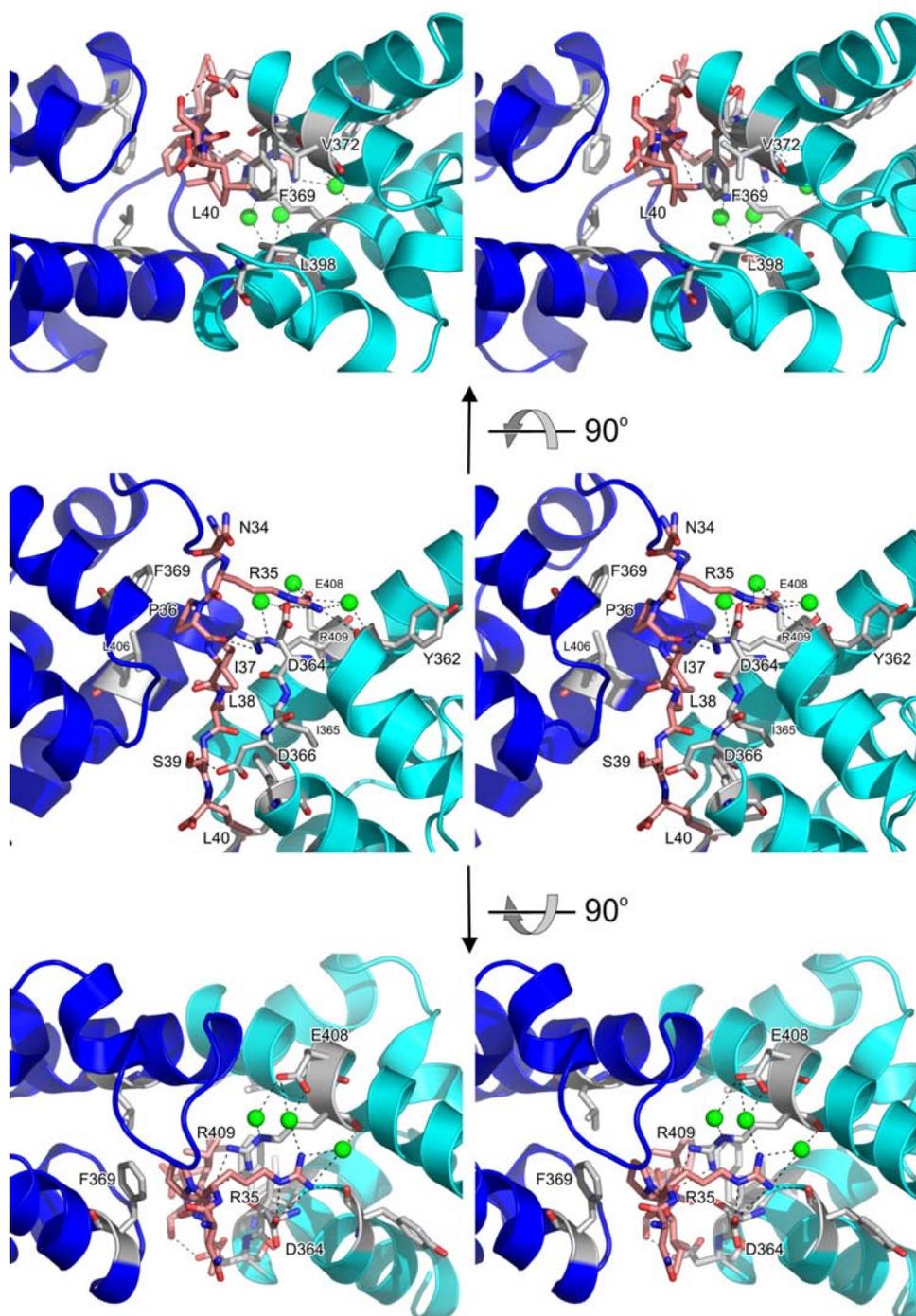


Figure 38. Three orthogonal stereo views detailing the AR1- λ N(34-40) interaction. The orientation of the central panel is the same as that of Figure 33C. Relative orientations of the top and bottom panels are indicated. The backbone ribbons of the AR1^A and AR1^B molecules are in blue and cyan, respectively. Residues interacting specifically with the peptide are shown as sticks and are color coded (carbon – grey, oxygen - red, nitrogen – blue). The λ N peptide is also drawn as a stick figure (carbon – pink). Three water molecules, which are mediating interactions, are shown as green spheres. Hydrogen bonds or salt bridges discussed in the text are indicated as dashed lines.

Single-residue mutations in the peptides, which completely abolished binding, support the biological relevance of the observed interactions despite the rather high dissociation constants. The NusA- λ N contacts occur within the framework of a large number of protein-protein and protein-RNA interactions (Mogridge et al., 1995; Mogridge et al., 1998b). In addition, all the interacting factors within the antitermination complexes are assembled on the surface of RNAP. Therefore, an enhancement of the individual interactions can be expected due to networking, proximity, and local concentration effects *in vivo*. Other protein-peptide interaction systems, e.g. poly-proline binding proteins and their ligands, show affinities *in vitro*, which can be even an order of magnitude weaker than the ones reported herein (e.g. Freund et al., 2002). Furthermore, specific single amino acid mutations in the ligand peptides can completely abolish binding. Thus, the C-terminal extension of NusA does not bind to any arbitrary peptide, but exhibits clear specificity for λ N.

5.2.3.6 AR2 Contributes to λ N Binding

Full-length λ N binds to full-length NusA with a K_d of 70 nM (Van Gilst and von Hippel, 1997), demonstrating additional contacts between NusA and λ N beyond those characterized herein. λ N(34-47), which was used for crystallization, corresponds to a mapped NusA binding site on λ N (Mogridge et al., 1998a). The seven amino acids, 41-47, which are contained in the present crystals, but do not engage in contacts to AR1, may harbor an additional binding site for AR2. In order to verify this notion, the affinity of λ N(34-47) towards AR1-AR2 was tested by ITC. Indeed, the dissociation constant of λ N(34-47) was three-fold lower than that of λ N(31-43) (3.5 μ M vs. 9.6 μ M) (Table 7). A straightforward interpretation for these results assumes that λ N residues 41-47 supply a binding site for AR2. A concomitant and independent binding of the two halves of λ N(34-47) to AR1 and AR2, respectively, would explain that loss of one of the two acidic repeats in NusA has only a minor effect on λ N binding (Mah et al., 1999). In support of this notion, the homology model of AR2 maintains an electronegative surface similar to that of AR1 and can be seen easily to engage in a complex with the positively charged region 41-47 of λ N. In addition, the sequence of λ N(34-40) (NRPILSL) resembles that of λ N(41-47) (NRKPKSR), suggesting that residues 41-47 of λ N may bind in a similarly extended conformation to AR2 as seen in the AR1- λ N(34-40) complex.

Previous results have indicated that the entire NusA interaction region of λ N is contained in residues 34-107 (Van Gilst and von Hippel, 1997; Mogridge et al., 1998a) and that residues 241-495 of NusA hold its entire λ N interaction region (Mah et al., 1999). The here presented results demonstrate that region 34-47 of λ N and full-length NusA interact with a K_d of 3.5 μ M. Therefore, additional interactions that confer the 70 nM dissociation constant must encompass NusA 241-495 and λ N 47-107. The *pun* mutations of λ N, which restore λ N function in a *nusA1* mutant background (Sparkowski and Das, 1992), map within and adjacent to the suspected AR2-interaction region of λ N (Friedman et al., 1981) (K45R, *punA134*; S50R, *punA150*; and I55M, *punA165*). *NusA1* is a point mutation in the core of the NusA S1 domain and may lead to a reduction in RNA affinity (Worbs et al., 2001). Assuming that tight NusA- λ N binding is the result of an interaction between AR1 and λ N(34-40) (this crystal structure) of a similar AR2- λ N(41-47) interaction and of additional, uncharacterized interactions between λ N(48-55) and AR2 yields a simple interpretation of the genetic data: the *pun* mutations could compensate for weakened S1 domain-RNA interactions in NusA1 with increased AR2- λ N(*pun*) binding affinity.

5.2.3.7 λ N Binding to NusA Does Not Enhance Its RNA Affinity

Full-length NusA is capable of efficient RNA binding only in the presence of α -CTD of RNAP, while NusA lacking the last eighty amino acids (i.e. AR2) can bind RNA in isolation and enhances transcriptional pausing and termination without the intervention of the RNAP α subunit (Mah et al., 2000). Thus, AR2 in the full-length protein most likely occludes the RNA binding domains, e.g. by directly folding back onto the S1/KH region. The activation of RNA binding may occur by alternative interaction of AR2 with α -CTD. The present studies infer that AR2, in addition, interacts with λ N. The dual role of AR2 as a main interaction site for the α -CTD and a binding site for λ N could be the main switch through which NusA is either directed to stimulate pausing and termination (α -CTD) or antitermination (λ N). An important question then is whether λ N and α -CTD bind to the same or different sites on AR2. If binding was to the same site, λ N would be expected to stimulate NusA-RNA interaction in a fashion similar to α -CTD (Mah et al., 2000). In order to test this possibility, gel mobility shift experiments using radiolabeled *nut* site RNA, full-length *E. coli* NusA and various λ N peptides were conducted. The λ N peptides were restricted to all or part of the NusA binding region because the strong *boxB* RNA affinity of the N-terminal part of λ N (Figure 36A)

would mask any RNA binding by NusA. Contrary to the findings with α -CTD (Mah et al., 2000), none of the λ N peptides tested stimulated *nut* site RNA binding by NusA (Figure 36B). Similar results were obtained when *boxA* RNA or *boxB* RNA oligomers were employed instead of the entire *nut* site. Complementary evidence comes from the work of others, in which the interaction of NusA and λ N was found to be independent of RNA binding (Van Gilst and von Hippel, 1997; Xia et al., 2003).

These results demonstrate that the interactions of λ N or α -CTD with AR2 have different physiological consequences, possibly because the two proteins interact with different surfaces of AR2. The aforementioned observation that full-length NusA binds equally well or better than the AR1-AR2 fragment to λ N is consistent with this notion. λ N can interact with the AR1-AR2 region without liberating AR2 from its interaction with the S1/KH portion; in contrast, α -CTD may interact with the same surface of AR2, which is used to block RNA binding pockets on NusA.

5.2.3.8 Summary and Conclusions

This research describes the crystal structure of a complex between the first C-terminal acidic repeat of *E. coli* NusA and a cognate peptide of protein N from phage λ . The structure of NusA AR1 and a homology model of AR2 clearly show that the two domains constitute duplicated HhH motifs. Solution studies demonstrate that these (HhH)₂ motifs do not interact with nucleic acids but rather constitute protein binding modules. The λ N fragment 34-40 was found in an extended conformation in the complex with AR1, different from the behavior of the N-terminal part of the protein in complex with *boxB* RNA. Solution studies further demonstrated that a second contact region for NusA AR2 exists in residues 41-47 of λ N and that the binding sites for λ N are accessible in isolated NusA. Binding constants measured for NusA- λ N complexes by ITC are in the micromolar range. Mutational studies clearly demonstrate the significance of these weak interactions. Although the present studies unequivocally demonstrate that λ N contacts NusA in the same region as α -CTD, λ N binding *per se* does not activate NusA's RNA interaction capacity, as does the RNAP domain. Therefore, it is likely that α -CTD and λ N interact with different surfaces of the NusA C-terminal repeat region. The NusA binding site on λ N (residues 34-47) directly adjoins the λ N interaction site for *boxB* RNA (residues 1-22). Although residues 34-47 alone do not stimulate

RESULTS AND DISCUSSION

RNA binding by NusA, it is likely that the close apposition of the *nut* site and NusA through the λ N scaffold will facilitate NusA-mRNA interactions.

5.3 Crystal Structure of the Antitermination Factor NusB

5.3.1 Crystallography

5.3.1.1 Crystal Growth and Data Collection

The initial crystallization trials on *tmNusB* were performed using commercial and home made screens. After refinement of pH and precipitant concentrations, optimized recipes for five crystal forms were obtained (Table 8). *tmNusB* could be crystallized under various conditions in a monoclinic ($P2_1$), an orthorhombic ($P2_12_12_1$), and three trigonal space groups (two different $P3_12_1$ forms, one $P3_22_1$ form) (Figure 39; Table 8).

All crystal forms were frozen in a nitrogen stream (100K; Oxford Cryosystems) after transfer into perfluoropolyether. Data were collected at beamline BW6 at DESY (Hamburg). Indexing, integration, and reduction of the diffraction data were carried out with DENZO/SCALEPACK (Otwinowski and Minor, 1997) and programs from the CCP4 suite (Bailey, 1994).

Crystal form	1	2	3	4	5
	pH 6.3	pH 7.4	pH 7.0	pH 6.6	pH 7.9
Crystallization buffer	0.2M AS ^a 30% PEG8000	10% isopropanol 20% PEG4000	0.2M CaCl ₂ 25% PEG4000	0.2M AA ^b 2.0M K/Na P _i	1.2M NaCit ^c
Space group	$P2_1$	$P2_12_12_1$	$P3_12_1$	$P3_12_1$	$P3_22_1$
Unit cell (Å, °)	a = 59.7, b = 34.4, c = 74.3; β = 94.2	a = 33.8, b = 59.6, c = 64.3	a = b = 46.0, c = 122.8	a = b = 60.2, c = 87.7	a = b = 60.0, c = 172.4

^a AS – ammonium sulfate

^b AA – ammonium acetate

^c Cit – citrate

Table 8. Crystallization conditions for *tmNusB*.

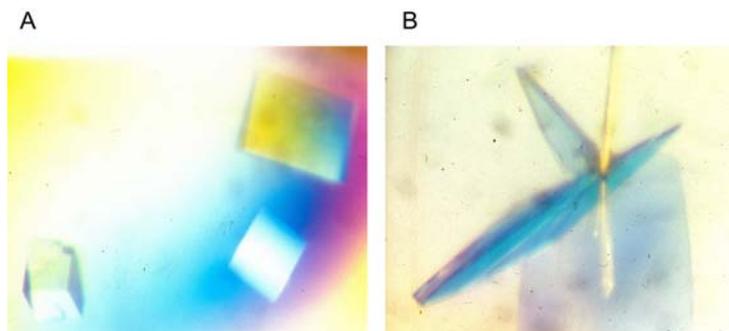


Figure 39. Crystals of *tmNusB*. **A.** Trigonal crystals ($P3_121$, crystal form 3). **B.** Orthorhombic crystal ($P2_12_12_1$, crystal form 2).

5.3.1.2 Structure Determination

The structure of *tmNusB* in the most highly resolved $P3_121$ form (form 4, Table 9) could be solved by the Patterson search method. The search model consisted of the monomeric poly-alanine homologous to *mtNusB* (Gopal et al., 2000; PDB accession code 1EYV). The best solution had a correlation of 36.0% and an R-factor of 54.5% in the resolution range between 20 and 3 Å. The refined *tmNusB* structure, including side chains but excluding water oxygens, was subsequently employed to solve the other four structures by Patterson search methods. In all cases, rigid body refinement produced initial $2F_o-F_c$ electron density maps, which allowed the stepwise building or repositioning of side chains. Refinement continued with standard protocols (positional-, simulated annealing- and B-factor refinement). During all refinement steps, all experimental data better than 20 Å resolution without intensity cutoffs were included and the R_{free} -factors were continuously monitored with 5% of the reflections (Table 9). For space groups, which contained two molecules per asymmetric unit, the final round included a TLS-refinement step with different anisotropic temperature factor corrections for the two molecules. In all structures, the main chains and the vast majority of side chains, except for some surface-exposed flexible residues, were completely enveloped in the final $2F_o-F_c$ electron density maps at the 1σ level. The final models contained all residues of the protein, some lacking the N-terminal methionine or the C-terminal leucine, and displayed good stereochemistry (Table 9) with low residual coordinate errors around 0.1 Å (Luzzati, 1952).

RESULTS AND DISCUSSION

Data collection					
Crystal form	1	2	3	4	5
Resolution (Å)	30.0 - 1.55	30.0 - 1.85	30.0 - 1.60	30.0 - 1.35	30.0 - 1.72
Reflections					
Unique	43301	11016	20371	41043	39263
Redundancy	3.8	2.5	3.5	5.4	4.7
Completeness (%)	97.9 (95.2)	94.6 (94.8)	98.4 (95.6)	99.9 (99.9)	99.6 (99.5)
I/σ(I)	28.7 (1.6)	20.6 (3.1)	21.4 (3.5)	38.1 (3.4)	37.6 (2.9)
R_{sym}	0.072 (0.616)	0.068 (0.253)	0.071 (0.323)	0.072 (0.358)	0.069 (0.373)
Refinement					
Resolution (Å)	20.0 - 1.55	20.0 - 1.85	20.0 - 1.60	20.0 - 1.35	20.0 - 1.72
Reflections (%)	40560/96.7	10471/94.9	20345/98.5	40933/99.7	39095/99.6
Test set (%)	5	5	5	5	5
R_{work}^d	0.235	0.224	0.187	0.199	0.217
R_{free}^d	0.251	0.253	0.223	0.214	0.249
Contents of a.u.^e					
Protein molecules	2	1	1	1	2
Protein atoms	2362	1175	1196	1188	2376
Water atoms	392	137	264	288	238
Ligands	2 sulfate ions	1 sulfate ion	1 calcium ion	-	2 citrate ions
Mean B-factors (Å²)					
Wilson	26.2	27.6	19.4	19.4	33.3
Protein	24.7	27.9	19.3	22.7	26.5
Water	45.8	49.8	35.2	39.3	56.8
Ligands	80.2	79.3	14.7	-	91.4
Ramachandran plot					
Preferred	0.992	0.984	0.992	0.992	1.000
Add. allowed	0.8	0.8	0.8	0.0	0.0
Disallowed	0.0	0.8	0.0	0.8	0.0
R.m.s.d.^f geometry					
Bond lengths (Å)	0.009	0.010	0.007	0.008	0.008
Bond angles (°)	1.18	1.02	1.40	1.27	1.08
R.m.s.d. B-factors (Å²)					
Main chain bonds	1.4	1.27	2.7	2.4	1.3
Main chain angles	2.2	2.42	3.3	3.3	2.2
Side chain bonds	1.6	3.40	5.3	4.6	1.5
Side chain angles	2.3	5.83	6.6	5.9	2.5

$$^d \quad R_{\text{work}} = \frac{\sum_{\text{hkl}} | | F_{\text{obs}} | - k | F_{\text{calc}} | | }{\sum_{\text{hkl}} | F_{\text{obs}} | }; \quad R_{\text{free}} = \frac{\sum_{\text{hkl} \in T} | | F_{\text{obs}} | - k | F_{\text{calc}} | | }{\sum_{\text{hkl} \in T} | F_{\text{obs}} | }; \quad \text{hkl} \in T - \text{test set}$$

^e a.u.- asymmetric unit

^f r.m.s.d. - root-mean-square deviation

Data in parentheses are for the final 0.1Å

Table 9. Data collection and refinement statistics for *tmNusB*.

5.3.1.3 Oligomerization State in Solution

tmNusB crystallized as monomer in three crystal forms (1, 2 and 3). In contrast, there are two molecules per asymmetric unit in the two trigonal crystal forms 4 and 5.

In order to investigate the quaternary structure of *tmNusB* in solution, crosslinking with glutaraldehyde and analytical gel filtration experiments were conducted. In a time course crosslinking experiment with glutaraldehyde the emergence of covalent *tmNusB* dimers occur after 5 min of incubation, reaching a maximum after about 45 min. Then, additional higher molecular weight bands were visible, which corresponded to presumably non-specific aggregates. In analytical gel filtration on the other hand, *tmNusB* at physiological salt conditions, eluted with an apparent molecular mass of 18 KDa consistent with the molecular mass of the monomer. This technique was particularly successful for the molecular mass estimation of *tmNusB* being a spherical molecule.

5.3.2 Description of the Structure

5.3.2.1 Overall Structure

In all crystal forms *tmNusB* adopts a globular, all-helical fold (seven α -helices, two 3_{10} -helices) with approximate overall dimensions of $21 \times 22 \times 24 \text{ \AA}^3$ (Figure 40). All eight crystallographically independent molecules observed (Table 9) can be superimposed with pairwise r.m.s.d.'s of around 0.7 \AA for all 142 C_α atoms and, thus, exhibit no major conformational differences. The structure of *tmNusB* is, therefore, rather independent of the crystalline environment and will presumably be similar in solution.

The *tmNusB* structure can be divided into two subdomains, both of which comprise three-helix bundles, which can be superimposed on one another with a root-mean-square deviation (r.m.s.d.) of 1.5 \AA (37 matching C_α positions). Bundle I is built from helices $\alpha 1$ (Arg5-His19), $\alpha 2$ (Leu27-Ile34) and $\alpha 3$ (Lys41-Glu56), and bundle II is constructed from helices $\alpha 5$ (Val79-Phe94), $\alpha 6$ (Ile100-Tyr114) and $\alpha 7$ (Glu117-His134). Helix $\alpha 4$ (Leu58-Ser66), the following loop (Arg67-Ser73), helix $3_{10}1$ (Leu74-Arg76) and another short loop (Leu77-Ser78) connect the two subdomains. The two helical bundles are stacked onto each other at approximately right angles, with helices $\alpha 1$ and $\alpha 3$ of bundle I lying on top of helices $\alpha 5$ and $\alpha 6$ of bundle II (Figure 40). Upon association, they cover $\sim 1560 \text{ \AA}^2$ of combined surface area. They interact

through a hydrophobic patch in the center and through multiple salt bridges and hydrogen bonds at the periphery of their contact surfaces.

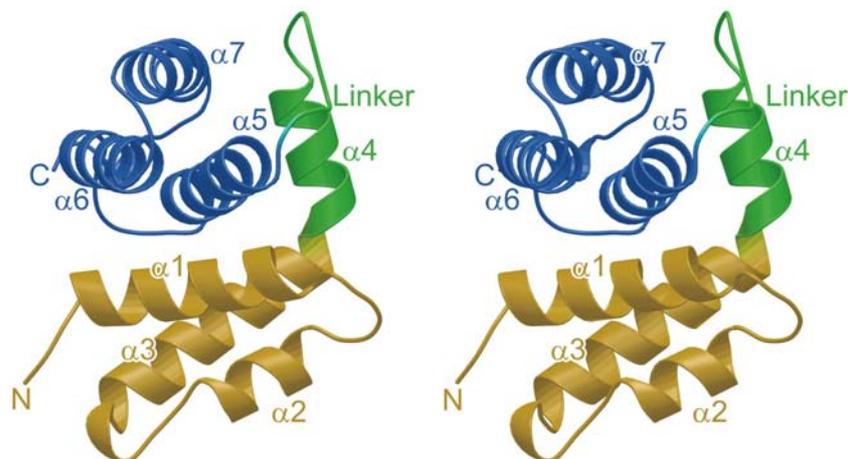


Figure 40. Close-up stereo ribbon plot of the *tmNusB* monomer. The two subdomains are shown in yellow (N-terminal three-helix bundle) and blue (C-terminal three-helix bundle), with the inserted linker in green. Secondary structure elements are labeled. The orientation is that of Figure 43 rotated 90° counterclockwise around the vertical axis.

5.3.2.2 Phylogenetic Comparisons

ecNusB and *mtNusB*, whose structures have been determined previously (Gopal et al., 2000; Altieri et al., 2000), contain respectively 40.9% and 41.7% identical residues with *tmNusB* and consistently maintain similar overall folds. *tmNusB* and *mtNusB* can be superimposed with an r.m.s.d. of 1.44 Å for 127 matching C α atoms (Figure 41). *ecNusB* fits on *tmNusB* with an r.m.s.d. of 1.80 Å for 88 matching residues (Figure 41).

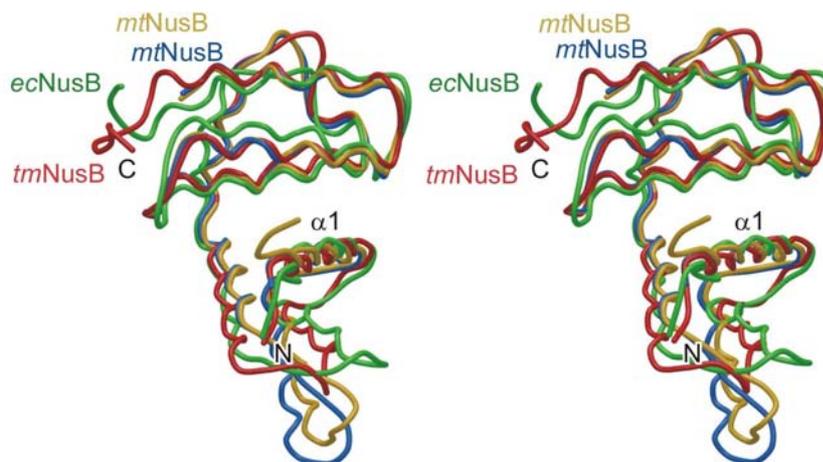


Figure 41. Superposition of known NusB structures. For clarity, only one *tmNusB* model is shown (red). Both crystallographically independent molecules of the *mtNusB* crystal structure (blue and gold; Gopal et al., 2000) and the NMR structure of *ecNusB* (green; Altieri et al., 2000) are displayed.

Local differences among these proteins are pronounced in the region around helices $\alpha 2$ and $\alpha 3$, which show relatively high sequence variation among NusB proteins (Figure 42). Helix $\alpha 2$ in *tmNusB* is shorter by one turn than the corresponding element in *mtNusB*, and the directions of the $\alpha 2$ helices relative to the bulk of the proteins deviate by $\sim 15^\circ$ from each other (Figure 42). Whereas in *tmNusB* and *ecNusB* helix $\alpha 2$ is in close contact with helix $\alpha 1$, this element is more closely associated with helix $\alpha 3$ in *mtNusB*. Overall, helix $\alpha 2$ and the following loop are further moved away from the remainder of the molecule in *mtNusB*, giving this protein a more elongated shape than *tmNusB*.

RESULTS AND DISCUSSION

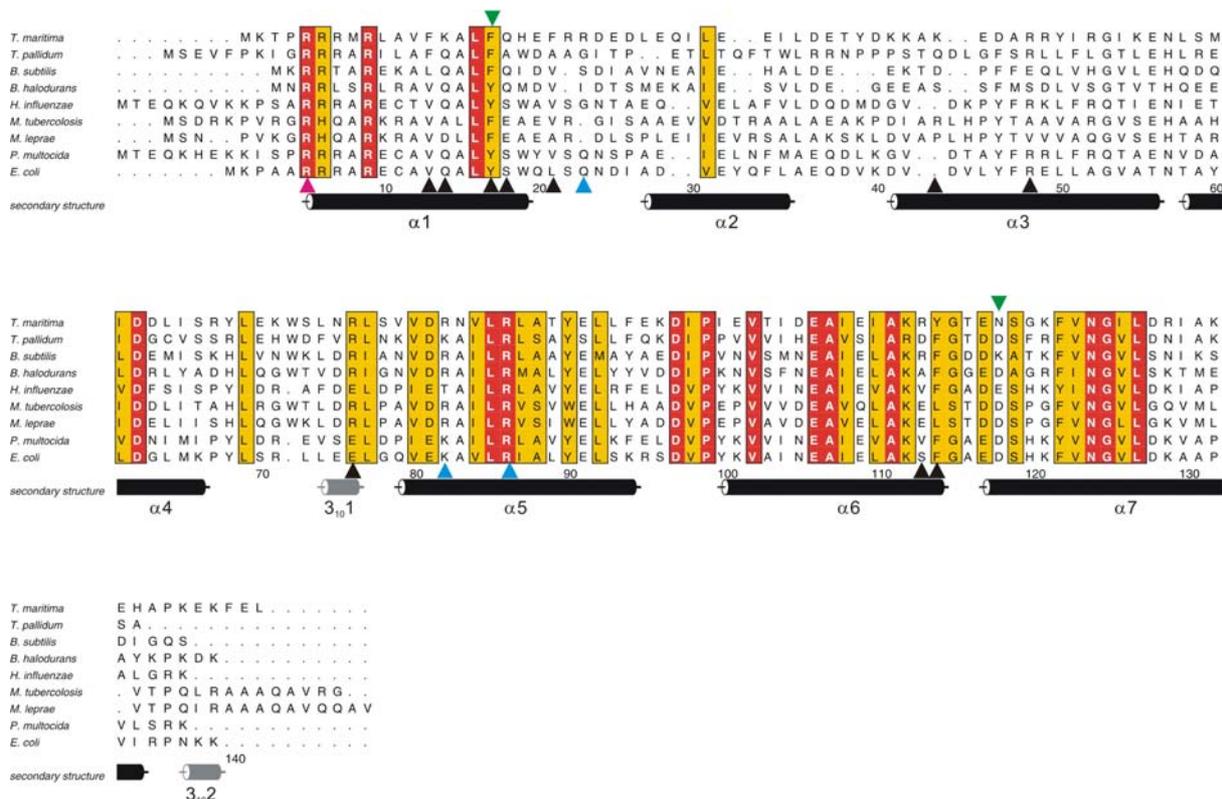


Figure 42. Alignment of nine bacterial NusB sequences. Numbering corresponds to *tmNusB*. The secondary structure elements indicated below the sequences refer to the situation seen in the present *tmNusB* crystal structures. α -helices are shown in black, and 3_{10} helices in gray. The background of highly conserved amino acids is red, and that of intermediately conserved amino acids is yellow. Black arrows denote residues involved in binding anions in the various *tmNusB* crystal forms. The magenta arrowhead indicates arginine, which in *mtNusB* is involved in binding a phosphate ion. Cyan arrowheads indicate residues forming a positive surface patch, which may constitute an RNA binding site (Altieri et al., 2000). Green arrowheads indicate mutations analyzed in the *E. coli* protein. Abbreviations: *T. maritima* - *Thermotoga maritima*; *T. pallidum* - *Treponema pallidum*; *B. subtilis* - *Bacillus subtilis*; *B. halodurans* - *Bacillus halodurans*; *H. influenzae* - *Haemophilus influenzae*; *M. tuberculosis* - *Mycobacterium tuberculosis*; *M. leprae* - *Mycobacterium leprae*; *P. multocida* - *Pasteurella multocida*; *E. coli* - *Escherichia coli*.

5.3.2.3 Quarternary Structures

In light of the previously observed *mtNusB* dimers (Gopal et al., 2000), the *tmNusB* crystal structures were analyzed for dimeric contacts. Important features that characterize dimerization motifs in proteins are: (1) a significant size of the dimer interface (2) the involvement of specific atomic interactions (3) the participation of conserved residues in the interchain contacts (4) the saturation of the interaction potential of the monomers, i.e. the formation of isologous contacts (Jones and Thornton, 1996). In crystal forms 1-3 (Table 9)

associations between neighboring molecules could not be discerned, which qualified as *bona fide* dimerization motifs (Figure 43).

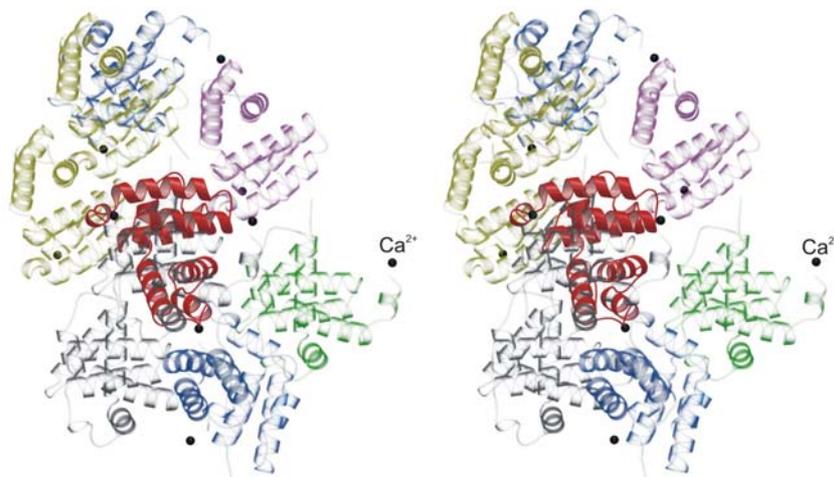


Figure 43. Stereo ribbon plot of a *tmNusB* monomer as seen in crystal form 3. Molecules surrounding a reference (solid red) are transparent and in various colors. Black spheres denote calcium ions, which in this crystal form mediate packing interactions, but are not seen in the other structures. The orientation of the reference molecule in this figure is the default setting for all following structure figures.

In contrast, two molecules in the form 4 crystals were related through a crystallographic two-fold axis in an isologous fashion, resembling the subunit arrangement of the *mtNusB* dimer (Gopal et al., 2000; Figure 44). Moreover, crystal form 5 maintained an identical and crystallographically independent apposition of two molecules.

Two contact regions sustain the crystallographic *tmNusB* dimers. First, the loop between helices α_6 and α_7 from one molecule lies alongside the loop preceding helix α_5 of the other, and *vice versa* (Figure 44). Second, the shape complementarity of these regions leads to van der Waals contacts, especially around the side chains of Trp72 and Thr116.

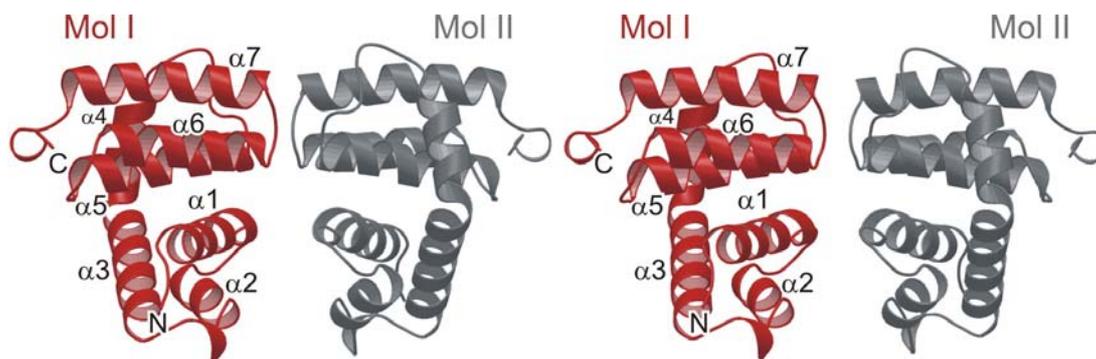


Figure 44. Stereo ribbon plot of a *tmNusB* dimeric arrangement as seen in crystal form 4. One monomer is shown in red and the other in gray. Secondary structure elements and the protein termini are labeled.

In addition, weak electrostatic interactions between Arg76 and Glu117 and between Lys71 and Asn118 can be discerned. Arg113 and Tyr114 from the C-terminus of helix $\alpha 6$ reach into the second contact site. Here, the N-terminal $\alpha 1$ helices of the two subunits run in antiparallel fashion through the body of the assembly, engaging in a number of side-chain to side-chain interactions (Figures 44 and 45). Phe13, Phe17 and Phe21 project from one side of helix $\alpha 1$ into the direction of the other monomer and partially stack on each other. The Phe21 ring also faces Tyr114 from the opposite monomer. In total, eight aromatic residues are arranged in a circular interaction network at the core of the dimer (Figure 45). The Arg113 side chain hydrogen bonds to the Tyr114 hydroxy group of its own monomer partially intervening into the intermolecular Phe21-Tyr114 association. Thus, it reinforces the Phe21-Tyr114 interaction through cation- π contacts to the two aromatic ring systems.

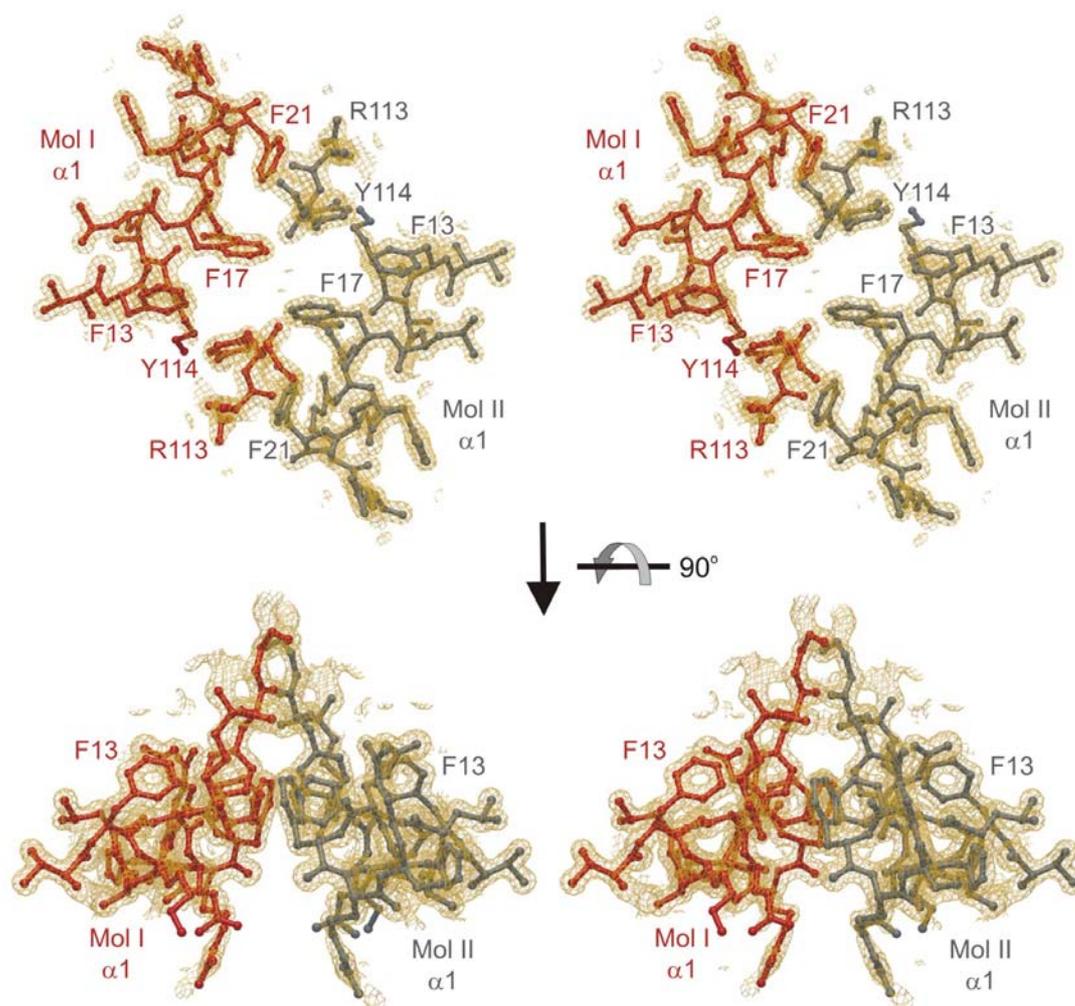


Figure 45. Stereo plot of the aromatic interactions in the dimer interface of crystal form 4. The orientation of the upper panel is that of Figure 43 rotated 90° into the plane of the paper around the vertical axis. Residues originating from one monomer are in red those from the other monomer are in gray. The aromatic residues participating in the interaction network are labeled. Arg113 may support the arrangement by additional cation- π interactions. A $2F_o-F_c$ electron density map for the final model contoured at the 2σ level is superimposed.

In addition, the electrostatic surface potential of *tmNusB* reveals a charge complementarity at the dimer interface, which should lead to an attraction between the monomers (Figure 46).

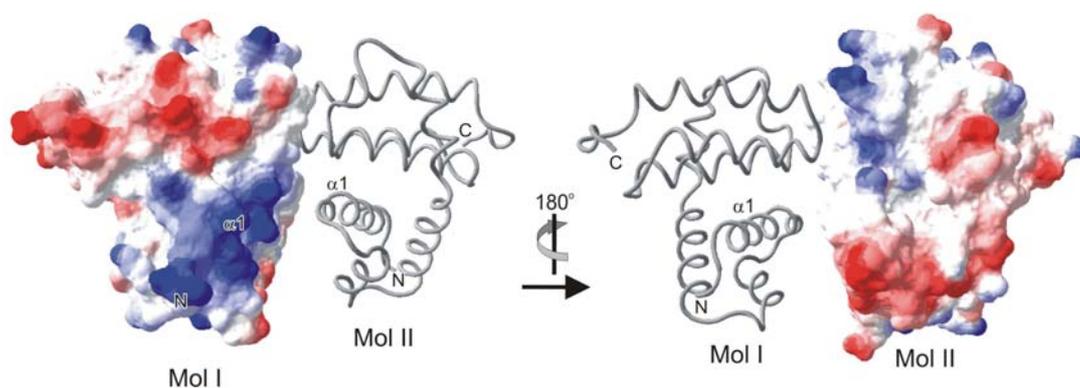


Figure 46. Electrostatic surface potential. The electrostatic potential was calculated for one monomer with the other displayed as a gray ribbon. Comparison of the 180° views illustrates a charge complementarity at the dimer interface. Negative potential is in red, positive potential in blue.

Notably, there is a gap of more than 30000 Å³ between the two subunits, mainly due to a deep cleft below the aromatic interaction motif (Figure 44). Voids between the subunits are filled with well-ordered water molecules (Figure 47), which mediate a considerable number of indirect interchain links. In the 1.35 Å structure of crystal form 4, there are ten single-water bridges and numerous additional connections through longer water chains.

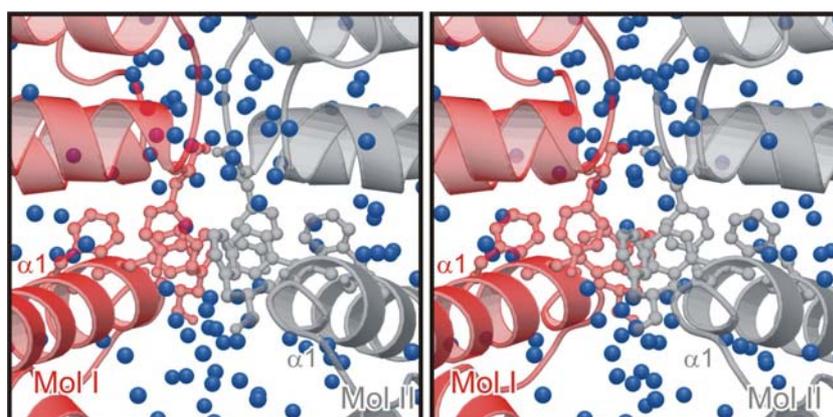


Figure 47. Close-up view of the dimer interface in default orientation. Blue spheres denote water molecules in the structure, mediating a large number of indirect interactions between the monomers. The side chains of the critical aromatic amino acids are drawn in ball-and-stick notation.

In addition, the *tmNusB* dimer arrangement buries only $\sim 1050 \text{ \AA}^2$ of accessible surface area. This number falls at the lower margin for typical dimeric proteins (Jones and Thornton, 1996). The comparably small contact area, the sparse number of direct interactions, the large number of water-mediated interactions and the charged residues surrounding the hydrophobic patches suggest that *tmNusB* dimers are facultative or unstable in solution (Jones and Thornton, 1996). This conclusion is consistent with the observation of both monomers and dimers in different crystal forms.

In *mtNusB*, the monomers interact more intimately, burying approximately double the accessible surface area (Gopal et al., 2000). The different positioning of helix $\alpha 2$ in *mtNusB* leads to additional contacts at the bottom of the dimers through the C-termini of these helices (Figure 48). However, there is a large channel ($\sim 5000 \text{ \AA}^3$) between the monomers. Similar to *tmNusB*, there are 16 direct water bridges between the monomers, as well as one salt bridge and seven hydrogen bonds. These contacts again surround a central stacking interaction, which in *mtNusB* only involves the two Phe22 residues of the $\alpha 1$ helices (Phe17 in *tmNusB*).

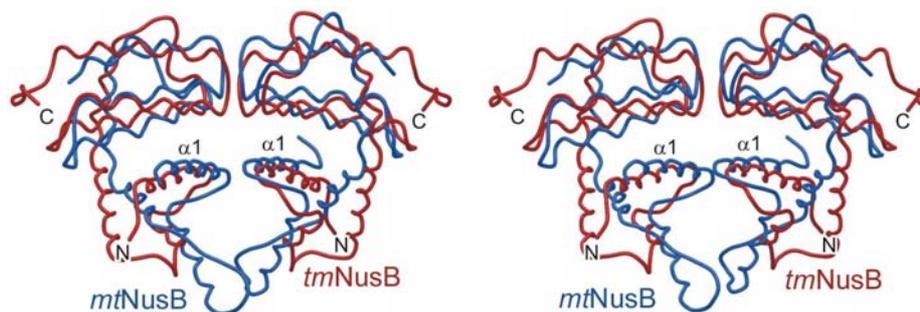


Figure 48. Superposition of the dimeric *mtNusB* structure on a *tmNusB*. The *tmNusB* dimer from the crystal form 4 is displayed in red, the *mtNusB* is blue. More extensive dimer contacts in the *mtNusB* molecules are apparent at the bottom, at the tip of helix $\alpha 2$. For orientation, N- and C-termini as well as helices $\alpha 1$ are labeled.

5.3.2.4 *tmNusB* is Monomeric in Solution

The similarity to the *mtNusB* dimers suggested that the dimeric associations in form 4 and 5 crystals could be physiologically relevant. Therefore, the potential of *tmNusB* to dimerize in solution was tested by glutaraldehyde crosslinking (Steiner et al., 2002). A continuous accumulation of covalent *tmNusB* dimers was observed throughout the reaction (Figure 49A).

However, in analytical gel filtration analysis with various buffer systems, *tmNusB* clearly migrated as a monomer (Figure 49B). Therefore, it was concluded that in solution *tmNusB* does not assemble into stable dimers, but molecules may transiently associate in the fashion seen in crystal forms 4 and 5.

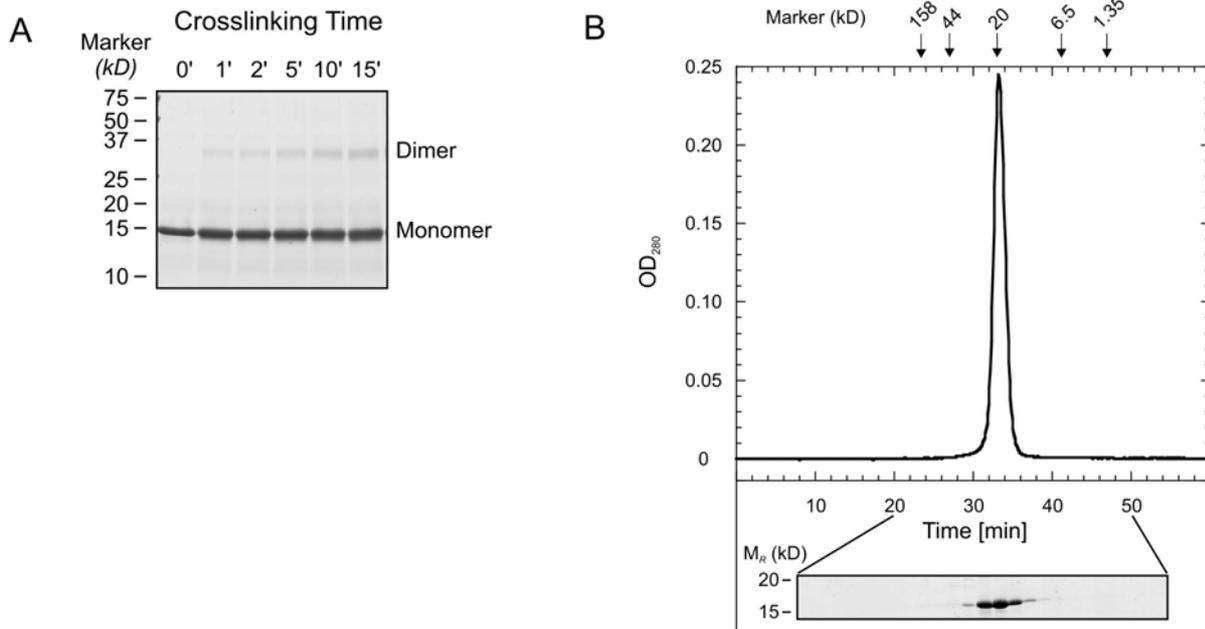


Figure 49. Oligomeric state of *tmNusB* in solution. **A.** SDS-PAGE showing the time course of a glutaraldehyde crosslinking experiment with *tmNusB*. Samples were taken between 1 and 15 min after the onset of the reaction, as indicated. **B.** Representative gel filtration analysis of *tmNusB*. The elution time of the single peak corresponds to a monomer molecular mass (~18 KDa). Arrows indicate elution peaks of reference molecules. The protein profile of the run as seen on an SDS polyacrylamide gel is shown below the graph.

5.3.2.5 Binding of *tmNusB* to *boxA*-like RNA

Whereas *ecNusB* has been shown to bind *boxA* RNA (Mogridge et al., 1995; Court et al., 1995; Lüttgen et al., 2002) Gopal and coworkers have failed to detect binding of *mtNusB* to *boxA* RNA *in vitro* (Gopal et al., 2000). In order to test RNA binding by *tmNusB*, two variants of the *E. coli rrn-boxA* sequence (5'-UGCUCUUUA-3' and 5'-UGCUCUUUAACA-3', termed RNA1 and RNA2, respectively) and two sequences from the *T. maritima rrn* leader and spacer regions (5'-GGGUCUUCA-3' and 5'-GGCUCAUUG-3'; RNA3 and RNA4, respectively), which had been suggested previously to constitute *boxA* elements in *Thermotoga* (Berg et al., 1989), were chemically synthesized and probed in electrophoretic

gel mobility shift assays. RNA3 was seen to bind to *tmNusB* better than the other RNAs (Figure 50A), indicating some specificity of the interaction. However, binding was weak or transient, because the complex seemed to dissociate during electrophoresis. Binding was enhanced when unlabeled RNA of the same type was added to the reactions (Figure 50A). Such cooperative effects could be due to a conformational change in the protein upon binding or due to the uncovering of another binding site after association with the first RNA molecule. In any case, the RNA binding capacity of NusB molecules seems to parallel their oligomeric state in solution: *ecNusB* or *tmNusB* proteins, which exist preferentially as monomers, exhibit binding capacity, while dimeric *mtNusB* does not.

In contrast with the results of the electrophoretic gel mobility shift assays, all RNA sequences tested could be covalently attached to *tmNusB* by UV-induced crosslinking (Figure 50B). These results confirm that *tmNusB* associates transiently with various RNA sequences. The crosslink to labeled RNA was enhanced by the addition of small amounts (increasing with the protein concentration) of unlabeled oligomer; larger amounts of unlabeled RNA led to the decrease in the crosslinking to labeled RNA (Figure 50B). Increasing the amount of protein in the reactions did not enhance the crosslink (Figure 50C). These results confirm that *boxA*-like sequences bind cooperatively to NusB, e.g. by eliciting a conformational change in the protein.

RESULTS AND DISCUSSION

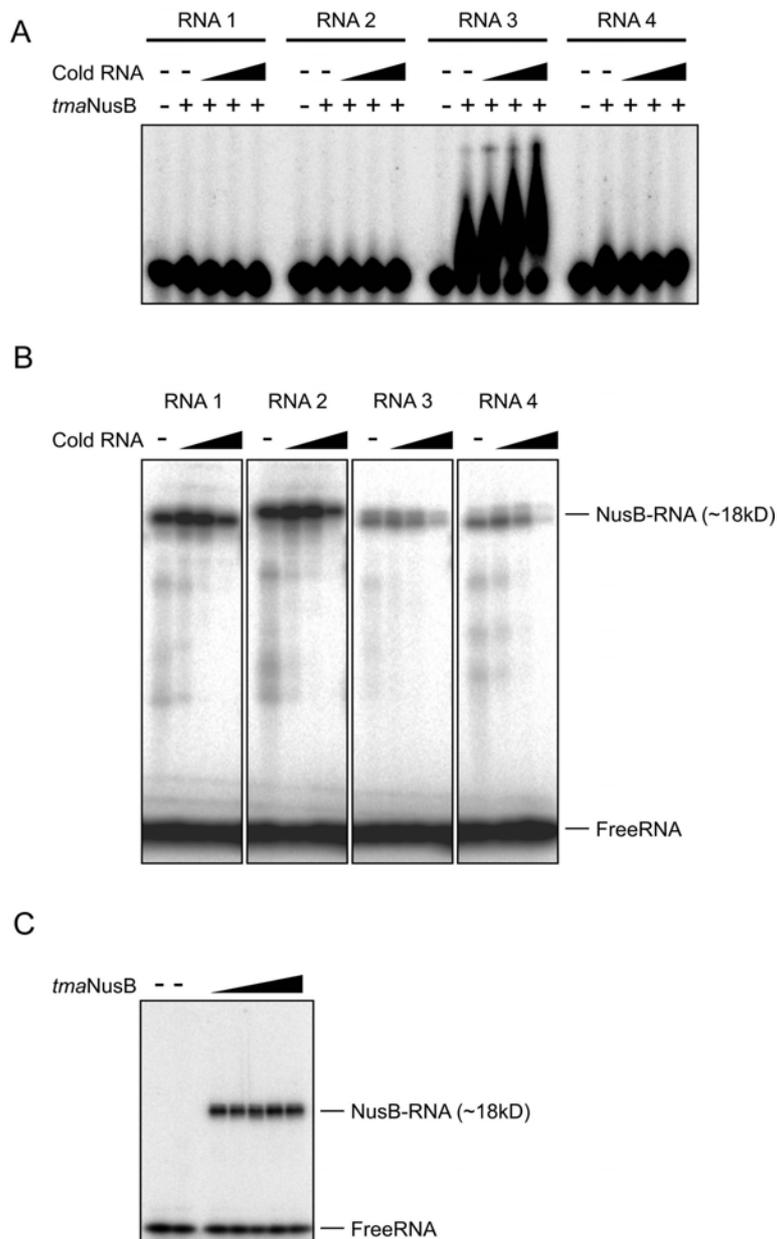


Figure 50. RNA binding of *tmNusB*. **A.** Electrophoretic gel mobility shift analysis of *tmNusB-boxA* complexes. Only RNA3 was shifted by the protein under the experimental conditions. The shapes of the retarded bands suggest that the *tmNusB-RNA3* complex is not stable, but partially dissociates during the run. Complex stability is increased by the addition of unlabeled RNA3. **B.** SDS-PAGE analysis of *tmNusB* UV-crosslinked to various *boxA*-like RNA oligomers. Each reaction contained 300 pmol of NusB and 4 pmol of radiolabeled RNA. The first to fourth lane of each block contained increasing amounts (0, 20, 250 and 4000 pmol) of unlabeled ('cold') competitor RNA (same sequence) as indicated above the gel. Because the short RNAs do not markedly modulate the migration behavior of the proteins on a denaturing SDS polyacrylamide gel, the crosslink at ~18 kDa presumably corresponds to monomeric NusB-*boxA* complexes. **C.** SDS-PAGE analysis of increasing amounts of *tmNusB* (0-300 pmol) UV-crosslinked to various *boxA*-like RNA oligomers. The first two lanes do not contain protein. The first lanes were not exposed to UV radiation.

5.3.2.6 A Putative RNA Binding Region at the NusB N-terminus

The above RNA interaction studies validate an interpretation of the *tmNusB* structures in terms of an RNA binding fold. About one third of the first 30 residues of *tmNusB* are Arg and Lys residues. This arginine-rich motif (ARM) is conserved in other NusB molecules (Figure 46). It gives rise to a broad stretch of positive surface potential, which extends along the center of the molecules (Figure 50) and appears well suited for interaction with the negatively charged sugar phosphate backbone of a nucleic acid.

Several anions bound to this region in various *tmNusB* crystal structures were located, which mark possible attachment sites for the RNA sugar phosphate backbone. In the form 5 crystals, two citrate ions are bound to the surfaces of the two $\alpha 1$ helices of the dimer. Each citrate ion lies along the aliphatic portion of the Lys14 side chain. One of the terminal carboxy groups forms salt bridges with Arg113 and a hydrogen bond to the Tyr114 hydroxy group (Figure 51). The 2-carboxy group is hydrogen-bonded to the amide group of Gln18 of the other monomer. In the $P2_1$ and $P2_12_12_1$ crystal forms, several sulfate ions were bound to Lys44, Arg48 and Arg76, which are flanking the binding site for the citrate ions.

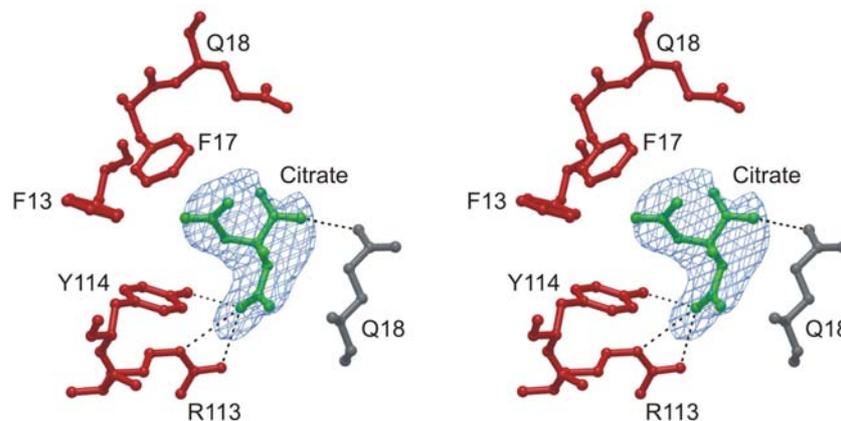


Figure 51. Binding region of one citrate ion in the form 5 structure. A F_o-F_c 'omit' electron density map, for which the ligand had been omitted, is displayed at the 2.5σ level (blue). Residues contributing to the citrate binding site from one monomer are shown in red, those from the other monomer are in gray; the citrate is green.

In other RNA binding proteins, aromatic residues are often essential for stacking to nucleic acid bases (Oubridge et al., 1994). As pointed out, residues Phe13, Phe17, Phe21, and the side chain of Tyr114 lie along the surface of helix $\alpha 1$ and, thus, could play a similar role

in *tmNusB*. Among these aromatic residues, Phe17 and Tyr114 show the highest degree of conservation and are particularly interesting candidates for RNA contact sites. Consistent with this interpretation, the *ecNusB5* mutant (Y18D in *ecNusB*; Phe17 in *tmNusB*, Figure 42) abrogates λ N-mediated antitermination activity (Court et al., 1995). NusB has been found to act as termination enhancer at a suboptimal Rho site. To perform this function, it has to interact with a variant *boxA* sequence upstream of the terminator (Carlomagno and Nappo, 2001). The data presented by Carlomagno and Nappo (Carlomagno and Nappo, 2001) strongly support the suggestion that Phe17 (Tyr18 in *ecNusB*) is directly involved in *boxA* RNA binding.

The above mentioned localization of a NusB-RNA interaction site is also consistent with previous structural studies. Chemical shift changes upon titration of *ecNusB* with a *boxA* RNA dodecamer (Huenges et al., 1998) suggested contacts of the RNA with the N-terminal region of the protein. Furthermore, Altieri and coworkers pointed out that the region around Arg82 and Arg86 exhibits features typical for an RNA binding site (Altieri et al., 2000). These two highly conserved arginines form a positively charged surface patch, which in *tmNusB* is enlarged by Arg23. This patch lies above the C-terminal end of helix α 1. Finally, the *mtNusB* crystal structure includes a phosphate ion bound in the amino terminal region to Arg10 (Arg5 in *tmNusB*; Gopal et al., 2000), in the vicinity of helix α 1. Taken together, the sequence composition and anion binding by *tmNusB*, as well as previous structural analysis of NusB proteins, imply as a possibility that NusB harbors a continuous RNA interaction region along the center of the protein, largely following the trace of helix α 1 (Figure 51).

5.3.2.7 Hypothesis - Dimerization as a Silencing Mechanism for Some NusB Proteins

In stable NusB dimers, such as *mtNusB*, the region implicated above in RNA binding overlaps with the dimerization interface (Figure 52).

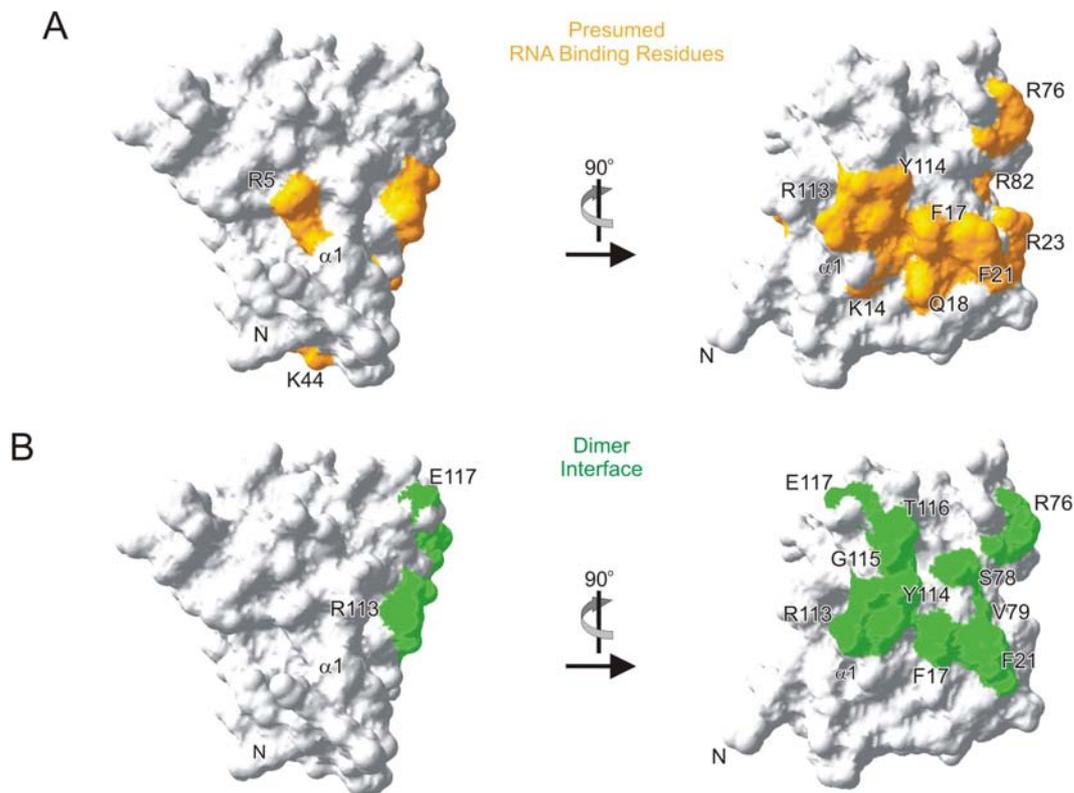


Figure 52. Surface topology of *tmNusB*. **A.** Mapping of residues on the surface of *tmNusB* which in the present or former structures have been implicated in RNA binding. Positions of individual residues are indicated. **B.** Two orthogonal views of the dimer interface residues (green), as seen in the *tmNusB* crystalline dimers, mapped on the surface of *tmNusB*. A similar but more extended dimer interface is seen in *mtNusB*. Residues are labeled and the position of the helix $\alpha 1$ is indicated and the N-termini are indicated.

Thus, concomitant dimerization and interaction with the RNA are seemingly incompatible. The dilemma could be resolved, if NusB proteins that form dimers underwent a large conformational change to expose helix $\alpha 1$ or were converted to monomers for RNA interaction. Such structural changes of NusB may depend on other factors within antitermination complexes. In this scenario, dimerization could be exploited in some bacteria to inactivate NusB until it is needed for transcriptional regulation. The crystalline dimers of *tmNusB* feature a large gap below the putative RNA binding site and, thus, still allow access of small anions, but may also obstruct binding of larger RNA molecules. Conspicuously, mechanisms of tunable RNA recognition have previously been identified in NusA (Mah et al., 2000) and have been suggested for NusG (Steiner et al., 2002). Therefore, they may emerge

RESULTS AND DISCUSSION

as a general theme in nucleic acid binding by the Nus-factors, exploited for regulating termination and antitermination processes.

6 References

- Abdel-Rassol, T. (2003) Construction of the Genomic Library of *Proteus vulgaris* and its Screening for the Gene Encoding (2R)-Hydroxycarboxylate-Viologen-Oxidoreductase. Universität Karlsruhe, Germany.
- Aiyar, A. (2000) The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol. Biol.*, **132**, 221-241.
- Altieri, A.S., Mazzulla, M.J., Horita, D.A., Coats, R.H., Wingfield, P.T., Das, A., Court, D.L. and Byrd, R.A. (2000) The structure of the transcriptional antiterminator NusB from *Escherichia coli*. *Nature Structural Biology*, **7**, 470-474.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389-3402.
- Arnoux, P., Sabaty, M., Alric, J., Frangioni, B., Guigliarelli, B., Adriano, J.M. and Pignol, D. (2003) Structural and redox plasticity in the heterodimeric periplasmic nitrate reductase. *Nature Structural Biology*, **10**, 928-934.
- Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography*, **50**, 760-763.
- Barton, G.J. (1993) Alscript - a Tool to Format Multiple Sequence Alignments. *Protein Engineering*, **6**, 37-40.
- Beinert, H., Holm, R.H. and Munck, E. (1997) Iron-sulfur clusters: Nature's modular, multipurpose structures. *Science*, **277**, 653-659.
- Berg, K.L., Squires, C. and Squires, C.L. (1989) Ribosomal-Rna Operon Anti-Termination - Function of Leader and Spacer Region Box-B-Box-a Sequences and Their Conservation in Diverse Microorganisms. *Journal of Molecular Biology*, **209**, 345-358.
- Bläse, M., Bruntner, C., Tshisuaka, B., Fetzner, S. and Lingens, F. (1996) Cloning, expression, and sequence analysis of the three genes encoding quinoline 2-oxidoreductase, a molybdenum-containing hydroxylase from *Pseudomonas putida* 86. *Journal of Biological Chemistry*, **271**, 23068-23079.
- Bohm, G., Muhr, R. and Jaenicke, R. (1992) Quantitative-Analysis of Protein Far Uv Circular-Dichroism Spectra by Neural Networks. *Protein Engineering*, **5**, 191-195.
- Boll, M., Fuchs, G. and Lowes, D.J. (2001) Single turnover EPR studies of benzoyl-CoA reductase. *Biochemistry*, **40**, 7612-7620.
- Bonin, I., Martins, B.M., Purvanov, V., Fetzner, S., Huber, R. and Dobbek, H. (2004) Active site geometry and substrate recognition of the molybdenum hydroxylase quinoline 2-oxidoreductase. *Structure*, **12**, 1425-1435.
- Boyington, J.C., Gladyshev, V.N., Khangulov, S.V., Stadtman, T.C. and Sun, P.D. (1997) Crystal structure of formate dehydrogenase H: Catalysis involving Mo, molybdopterin, selenocysteine, and an Fe₄S₄ cluster. *Science*, **275**, 1305-1308.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography*, **54**, 905-921.
- Bycroft, M., Hubbard, T.J.P., Proctor, M., Freund, S.M.V. and Murzin, A.G. (1997) The solution structure of the S1 RNA binding domain: A member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235-242.

REFERENCES

- Canne, C., Stephan, I., Finsterbusch, J., Lingens, F., Kappl, R., Fetzner, S. and Huttermann, J. (1997) Comparative EPR and redox studies of three prokaryotic enzymes of the xanthine oxidase family: Quinoline 2-oxidoreductase, quinaldine 4-oxidase, and isoquinoline 1-oxidoreductase. *Biochemistry*, **36**, 9780-9790.
- Carlomagno, M.S. and Nappo, A. (2001) The antiterminator NusB enhances termination at a sub-optimal Rho site. *Journal of Molecular Biology*, **309**, 19-28.
- Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W.W. and Rees, D.C. (1995) Structure of a Hyperthermophilic Tungstopterin Enzyme, Aldehyde Ferredoxin Oxidoreductase. *Science*, **267**, 1463-1469.
- Coughlan, M.P., Johnson, J.L. and Rajagopalan, K.V. (1980) Mechanisms of Inactivation of Molybdoenzymes by Cyanide. *Journal of Biological Chemistry*, **255**, 2694-2699.
- Court, D.L., Patterson, T.A., Baker, T., Costantino, N., Mao, X.H. and Friedman, D.I. (1995) Structural and Functional Analyses of the Transcription-Translation Proteins NusB and Nuse. *Journal of Bacteriology*, **177**, 2589-2591.
- Cramer, P., Bushnell, D.A., Fu, J.H., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. and Kornberg, R.D. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, **288**, 640-649.
- Czjzek, M., Dos Santos, J.P., Pommier, J., Giordano, G., Mejean, V. and Haser, R. (1998) Crystal structure of oxidized trimethylamine N-oxide reductase from *Shewanella massilia* at 2.5 angstrom resolution. *Journal of Molecular Biology*, **284**, 435-447.
- DeLano, W.L. (2002) The PyMOL User's Manual., DeLano Scientific, San Carlos, CA, USA.
- Dobbek, H. (2000) Röntgenstrukturanalyse von molybdän-haltigen kohlenmonoxid-dehydrogenasen. Technische Universität München, Germany.
- Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R. and Meyer, O. (2002) Catalysis at a dinuclear CuSMo(=O)OH cluster in a CO dehydrogenase resolved at 1.1-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15971-15976.
- Dobbek, H., Gremer, L., Meyer, O. and Huber, R. (1999) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanyl cysteine. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 8884-8889.
- Dym, O. and Eisenberg, D. (2001) Sequence-structure analysis of FAD-containing proteins. *Protein Science*, **10**, 1712-1728.
- Eisenmann, A., Schwarz, S., Rosch, P. and Schweimer, K. (2004) Letter to the Editor: Sequence-specific H-1, C-13, N-15 resonance assignments and secondary structure of the carboxyterminal domain of the E-coli transcription factor NusA. *Journal of Biomolecular Nmr*, **28**, 193-194.
- Enroth, C., Eger, B.T., Okamoto, K., Nishino, T. and Pai, E.F. (2000) Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: Structure-based mechanism of conversion. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10723-10728.
- Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *Journal of Molecular Graphics & Modelling*, **15**, 132-&.
- Fetzner, S., Tshisuaka, B., Lingens, F., Kappl, R. and Huttermann, J. (1998) Bacterial degradation of quinoline and derivatives - Pathways and their biocatalysts. *Angewandte Chemie-International Edition*, **37**, 577-597.

REFERENCES

- Finkelstein, A.V., Ptitsyn, O.B. and Kozitsyn, S.A. (1977) Theory of Protein Molecule Self-Organization .2. Comparison of Calculated Thermodynamic Parameters of Local Secondary Structures with Experiments. *Biopolymers*, **16**, 497-524.
- Fraaije, M.W., Van Berkel, W.J.H., Benen, J.A.E., Visser, J. and Mattevi, A. (1998) A novel oxidoreductase family sharing a conserved FAD-binding domain. *Trends in Biochemical Sciences*, **23**, 206-207.
- Frerichs-Deeken, U., Goldenstedt, B., Gahl-Janssen, R., Kappl, R., Huttermann, J. and Fetzner, S. (2003) Functional expression of the quinoline 2-oxidoreductase genes (qorMSL) in *Pseudomonas putida* KT2440 pUF1 and in *P-putida* 86-1 Delta qor pUF1 and analysis of the Qor proteins. *European Journal of Biochemistry*, **270**, 1567-1577.
- Freund, C., Kuhne, R., Yang, H.L., Park, S., Reinherz, E.L. and Wagner, G. (2002) Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *Embo Journal*, **21**, 5985-5995.
- Friedman, D.I., Schauer, A.T., Baumann, M.R., Baron, L.S. and Adhya, S.L. (1981) Evidence That Ribosomal Protein-S10 Participates in Control of Transcription Termination. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78**, 1115-1118.
- Ghisla, S. and Massey, V. (1989) Mechanisms of Flavoprotein-Catalyzed Reactions. *European Journal of Biochemistry*, **181**, 1-17.
- Gibson, T.J., Thompson, J.D. and Heringa, J. (1993) The Kh Domain Occurs in a Diverse Set of Rna-Binding Proteins That Include the Antiterminator Nusa and Is Probably Involved in Binding to Nucleic-Acid. *Febs Letters*, **324**, 361-366.
- Gill, S.C., Weitzel, S.E. and Vonhippel, P.H. (1991) Escherichia-Coli-Sigma-70 and Nusa Proteins .1. Binding Interactions with Core Rna-Polymerase in Solution and within the Transcription Complex. *Journal of Molecular Biology*, **220**, 307-324.
- Gopal, B., Haire, L.F., Cox, R.A., Colston, M.J., Major, S., Brannigan, J.A., Smerdon, S.J. and Dodson, G. (2000) The crystal structure of NusB from *Mycobacterium tuberculosis*. *Nature Structural Biology*, **7**, 475-478.
- Greenblatt, J. and Li, J. (1981) Interaction of the Sigma Factor and the Nusa Gene Protein of Escherichia-Coli with Rna-Polymerase in the Initiation-Termination Cycle of Transcription. *Cell*, **24**, 421-428.
- Greenblatt, J., Nodwell, J.R. and Mason, S.W. (1993) Transcriptional Antitermination. *Nature*, **364**, 401-406.
- Gremer, L. and Meyer, O. (1996) Characterization of xanthine dehydrogenase from the anaerobic bacterium *Veillonella atypica* and identification of a molybdopterincytosine-dinucleotide-containing molybdenum cofactor. *European Journal of Biochemistry*, **238**, 862-866.
- Grether-Beck, S., Igloi, G.L., Pust, S., Schilz, E., Decker, K. and Brandsch, R. (1994) Structural analysis and molybdenum-dependent expression of the pAO1-encoded nicotine dehydrogenase genes of *Arthrobacter nicotinovorans*. *Molecular Microbiology*, **13**, 929-936.
- Hänzelmann, P., Dobbek, H., Gremer, L., Huber, R. and Meyer, O. (2000) The effect of intracellular molybdenum in *Hydrogenophaga pseudoflava* on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase. *Journal of Molecular Biology*, **301**, 1221-1235.
- Hille, R. (1996) The mononuclear molybdenum enzymes. *Chemical Reviews*, **96**, 2757-2816.
- Hille, R., Retey, J., Bartlewski-Hof, U., Reichenbecher, W. and Schink, B. (1999) Mechanistic aspects of molybdenum-containing enzymes. *Fems Microbiology Reviews*, **22**, 489-501.

REFERENCES

- Holm, L. and Sander, C. (1993) Protein-Structure Comparison by Alignment of Distance Matrices. *Journal of Molecular Biology*, **233**, 123-138.
- Horwitz, R.J., Li, J. and Greenblatt, J. (1987) An Elongation Control Particle Containing the N-Gene Transcriptional Antitermination Protein of Bacteriophage-Lambda. *Cell*, **51**, 631-641.
- Huber, R., Hof, P., Duarte, R.O., Moura, J.J.G., Moura, I., Liu, M.Y., LeGall, J., Hille, R., Archer, M. and Romao, M.J. (1996) A structure-based catalytic mechanism for the xanthine oxidase family of molybdenum enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 8846-8851.
- Huenges, M., Rolz, C., Gschwind, R., Peteranderl, R., Berglechner, F., Richter, G., Bacher, A., Kessler, H. and Gemmecker, G. (1998) Solution structure of the antitermination protein NusB of Escherichia coli: a novel all-helical fold for an RNA-binding protein. *Embo Journal*, **17**, 4092-4100.
- Johnson, M.K., Rees, D.C. and Adams, M.W.W. (1996) Tungstoenzymes. *Chemical Reviews*, **96**, 2817-2839.
- Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13-20.
- Jones, T.A. (1992) A, yaap, asap, @#*? A set of averaging programs. *Molecular Replacement*, E.J. Dodson, S. Gover and W. Wolf, eds. SERC Daresbury Laboratory, Warrington, 91-105.
- Kaim, W. and Schwederski, B. (1994) *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*, John Wiley & Sons eds., New York.
- Kisker, C., Schindelin, H., Pacheco, A., Wehbi, W.A., Garrett, R.M., Rajagopalan, K.V., Enemark, J.H. and Rees, D.C. (1997a) Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell*, **91**, 973-983.
- Kisker, C., Schindelin, H. and Rees, D.C. (1997b) Molybdenum-cofactor-containing enzymes: Structure and mechanism. *Annual Review of Biochemistry*, **66**, 233-267.
- Kornberg, R.D. (1999) Eukaryotic transcriptional control. *Trends in Biochemical Sciences*, **24**, M46-M49.
- Krissinel, E. and Henrick, K. (2003) Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Ca alignment, scored by a new structural similarity function. *Proceedings of the 5th International Conference on Molecular Structural Biology*, A.J. Kungl and P. J. Kungl, eds., Vienna, September 3-7, 88.
- Laemmli, U.K. (1970) Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature*, **227**, 680-&.
- Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, **26**, 283-291.
- Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J.K., Kim, H.K., Kwon, S.T. and Suh, S.W. (2000) Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *Embo Journal*, **19**, 1119-1129.
- Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. *Cell*, **93**, 289-299.
- Lewin, B. (2000) *Genes VII*, Oxford University Press Inc. ed., New York.
- Lüttgen, H., Robelek, R., Mühlberger, R., Diercks, T., Schuster, S.C., Köhler, P., Kessler, H., Bacher, A. and Richter, G. (2002) Transcriptional regulation by antitermination. Interaction of RNA with NusB protein and NusB/NusE protein complex of Escherichia coli. *Journal of Molecular Biology*, **316**, 875-885.

REFERENCES

- Luzzati, V. (1952) Traitement Statistique Des Erreurs Dans La Determination Des Structures Cristallines. *Acta Crystallographica*, **5**, 802-810.
- Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K. and Greenblatt, J. (2000) The alpha subunit of E-coli RNA polymerase activates RNA binding by NusA. *Genes & Development*, **14**, 2664-2675.
- Mah, T.F., Li, J., Davidson, A.R. and Greenblatt, J. (1999) Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Molecular Microbiology*, **34**, 523-537.
- Massey, V. (2000) The chemical and biological versatility of riboflavin. *Biochemical Society Transactions*, **28**, 283-296.
- Matthews, B.W. (1968) Solvent Content of Protein Crystals. *Journal of Molecular Biology*, **33**, 491-&.
- Merritt, E.A. and Bacon, D.J. (1997) Raster3D: Photorealistic molecular graphics. In *Macromolecular Crystallography, Pt B*, Vol. 277, pp. 505-524.
- Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E. and Greenblatt, J. (1998a) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Molecular Cell*, **1**, 265-275.
- Mogridge, J., Mah, T.F. and Greenblatt, J. (1995) A Protein Rna Interaction Network Facilitates the Template-Independent Cooperative Assembly on Rna-Polymerase of a Stable Antitermination Complex Containing the Lambda-N-Protein. *Genes & Development*, **9**, 2831-2845.
- Mogridge, J., Mah, T.F. and Greenblatt, J. (1998b) Involvement of boxA nucleotides in the formation of a stable ribonucleoprotein complex containing the bacteriophage lambda N protein. *Journal of Biological Chemistry*, **273**, 4143-4148.
- Morris, R.J., Perrakis, A. and Lamzin, V.S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. In *Macromolecular Crystallography, Pt D*, Vol. 374, pp. 229-244.
- Navaza, J. (1994) AMoRe, an automated package for molecular replacement. *Acta Crystallographica Section A-Foundations of Crystallography*, **50**, 157-163.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein Folding and Association - Insights from the Interfacial and Thermodynamic Properties of Hydrocarbons. *Proteins-Structure Function and Genetics*, **11**, 281-296.
- Nishino, T., Ariyoshi, M., Iwasaki, H., Shinagawa, H. and Morikawa, K. (1998) Functional analyses of the domain structure in the Holliday junction binding protein RuvA. *Structure*, **6**, 11-21.
- Nodwell, J.R. and Greenblatt, J. (1991) The Nut Site of Bacteriophage-Lambda Is Made of Rna and Is Bound by Transcription Antitermination Factors on the Surface of Rna-Polymerase. *Genes & Development*, **5**, 2141-2151.
- Nudler, E. and Gottesman, M.E. (2002) Transcription termination and anti-termination in E-coli. *Genes to Cells*, **7**, 755-768.
- Okamoto, K., Eger, B.T., Nishino, T., Kondo, S. and Pai, E.F. (2003) An extremely potent inhibitor of xanthine oxidoreductase - Crystal structure of the enzyme-inhibitor complex and mechanism of inhibition. *Journal of Biological Chemistry*, **278**, 1848-1855.
- Okamoto, K., Matsumoto, K., Hille, R., Eger, B.T., Pai, E.F. and Nishino, T. (2004) The crystal structure of xanthine oxidoreductase during catalysis: Implications for reaction mechanism and enzyme inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 7931-7936.

REFERENCES

- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, Pt A*, Vol. 276, pp. 307-326.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal-Structure at 1.92 Angstrom Resolution of the Rna-Binding Domain of the U1a Spliceosomal Protein Complexed with an Rna Hairpin. *Nature*, **372**, 432-438.
- Page, C.C., Moser, C.C., Chen, X.X. and Dutton, P.L. (1999) Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*, **402**, 47-52.
- Patterson, T.A., Zhang, Z.S., Baker, T., Johnson, L.L., Friedman, D.I. and Court, D.L. (1994) Bacteriophage-Lambda N-Dependent Transcription Antitermination - Competition for an Rna Site May Regulate Antitermination. *Journal of Molecular Biology*, **236**, 217-228.
- Peariso, K., Chohan, B.S., Carrano, C.J. and Kirk, M.L. (2003) Synthesis and EPR characterization of new models for the one-electron reduced molybdenum site of sulfite oxidase. *Inorganic Chemistry*, **42**, 6194-6203.
- Ptashne, M. and Gann, A. (2002) *Genes & Signals*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Rappe, A.K. and Goddard, W.A. (1980) Bivalent Spectator Oxo Bonds in Metathesis and Epoxidation Alkenes. *Nature*, **285**, 311-312.
- Rebelo, J., Macieira, S., Dias, J.M., Huber, R., Ascenso, C.S., Rusnak, F., Moura, J.J.G., Moura, I. and Romão, M.J. (2000) Gene sequence and crystal structure of the aldehyde oxidoreductase from *Desulfovibrio desulfuricans* ATCC 27774. *Journal of Molecular Biology*, **297**, 135-146.
- Rebelo, J.M., Dias, J.M., Huber, R., Moura, J.J.G. and Romão, M.J. (2001) Structure refinement of the aldehyde oxidoreductase from *Desulfovibrio gigas* (MOP) at 1.28 angstrom. *Journal of Biological Inorganic Chemistry*, **6**, 791-800.
- Rees, D.C. and Howard, J.B. (2003) The interface between the biological and inorganic worlds: Iron-sulfur metallocusters. *Science*, **300**, 929-931.
- Richardson, J.P. (2003) Loading rho to terminate transcription. *Cell*, **114**, 157-159.
- Richardson, J.P. and Greenblatt, J. (1996) Control of RNA chain elongation and termination. *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*, Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umberger, H. E., eds., American Society for Microbiology, Washington, DC, 822-848.
- Romão, M.J., Archer, M., Moura, I., Moura, J.J.G., Legall, J., Engh, R., Schneider, M., Hof, P. and Huber, R. (1995) Crystal-Structure of the Xanthine Oxidase-Related Aldehyde Oxidoreductase from *D-Gigas*. *Science*, **270**, 1170-1176.
- Romão, M.J., Rosch, N. and Huber, R. (1997) The molybdenum site in the xanthine oxidase-related aldehyde oxidoreductase from *Desulfovibrio gigas* and a catalytic mechanism for this class of enzymes. *Journal of Biological Inorganic Chemistry*, **2**, 782-785.
- Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J. and Pelletier, H. (1997) Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: Evidence for an induced fit mechanism. *Biochemistry*, **36**, 11205-11215.
- Schindelin, H., Kisker, C. and Rees, D.C. (1997) The molybdenum-cofactor: a crystallographic perspective. *Journal of Biological Inorganic Chemistry*, **2**, 773-781.
- Schneider, F., Lowe, J., Huber, R., Schindelin, H., Kisker, C. and Knablein, J. (1996) Crystal structure of dimethyl sulfoxide reductase from *Rhodobacter capsulatus* at 1.88 angstrom resolution. *Journal of Molecular Biology*, **263**, 53-69.
- Schneider, T.R. and Sheldrick, G.M. (2002) Substructure solution with SHELXD. *Acta Crystallographica Section D-Biological Crystallography*, **58**, 1772-1779.

REFERENCES

- Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal-Structure of a Cap-DNA Complex - the DNA Is Bent by 90-Degrees. *Science*, **253**, 1001-1007.
- Schulz, G.E. (1992) Binding of nucleotides by proteins. *Current Opinion in Structural Biology*, **2**, 61-67.
- Shao, X.G. and Grishin, N.V. (2000) Common fold in helix-hairpin-helix proteins. *Nucleic Acids Research*, **28**, 2643-2650.
- Skordalakes, E. and Berger, J.M. (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell*, **114**, 135-146.
- Sparkowski, J. and Das, A. (1992) Simultaneous Gain and Loss of Functions Caused by a Single Amino-Acid Substitution in the Beta Subunit of Escherichia-Coli Rna-Polymerase - Suppression of Nua and Rho Mutations and Conditional Lethality. *Genetics*, **130**, 411-428.
- Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R. and Wahl, M.C. (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *Embo Journal*, **21**, 4641-4653.
- Sticht, H. and Rosch, P. (1998) The structure of iron-sulfur proteins. *Progress in Biophysics & Molecular Biology*, **70**, 95-136.
- Stiefel, E.I. (1997) Chemical keys to molybdenum enzymes. *Journal of the Chemical Society-Dalton Transactions*, 3915-3923.
- Suzuki, T., Miyata, Y., Saeki, K., Kawazoe, Y., Hayashi, M. and Sofuni, T. (1998) In vivo mutagenesis by the hepatocarcinogen quinoline in the lacZ transgenic mouse: evidence for its in vivo genotoxicity. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis*, **412**, 161-166.
- Thanos, C.D., Goodwill, K.E. and Bowie, J.U. (1999) Oligomeric structure of the human EphB2 receptor SAM domain. *Science*, **283**, 833-836.
- Thapper, A., Donahue, J.P., Musgrave, K.B., Willer, M.W., Nordlander, E., Hedman, B., Hodgson, K.O. and Holm, R.H. (1999) The unperturbed oxo-sulfido functional group cis-Mo(VI)OS related to that in the xanthine oxidase family of molybdoenzyme: Synthesis, structural characterization, and reactivity aspects. *Inorganic Chemistry*, **38**, 4104-4114.
- Truglio, J.J., Theis, K., Leimkuhler, S., Rappa, R., Rajagopalan, K.V. and Kisker, C. (2002) Crystal structures of the active and alloxanthine-inhibited forms of xanthine dehydrogenase from Rhodobacter capsulatus. *Structure*, **10**, 115-125.
- Tshisuaka, B., Kappl, R., Huttermann, J. and Lingens, F. (1993) Quinoline Oxidoreductase from Pseudomonas-Putida-86 - an Improved Purification Procedure and Electron-Paramagnetic-Resonance Spectroscopy. *Biochemistry*, **32**, 12928-12934.
- Turk, D. (1996) MAIN 96: An interactive software for density modifications, model building, structure refinement and analysis. *Meeting of the International Union of Crystallography Macromolecular Computing School*, International Union of Crystallography, P.E. Bourne, and K. Watpaugh, eds. Western Washington University, WA.
- Van Gilst, M.R., Rees, W.A., Das, A. and von Hippel, P.H. (1997) Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry*, **36**, 1514-1524.
- Van Gilst, M.R. and von Hippel, P.H. (1997) Assembly of the N-dependent antitermination complex of phage lambda: NusA and RNA bind independently to different unfolded domains of the N protein. *Journal of Molecular Biology*, **274**, 160-173.
- Voet, D., Voet, J.G. and Pratt, C.W. (1999) *Fundamentals of Biochemistry*, John Wiley & Sons eds., New York.

REFERENCES

- Weisberg, R.A., Gottesman, M.E., Hendrix, R.M. and Little, J.W. (1999) Family values in the age of genomics: Comparative analyses of temperate bacteriophage HK022. *Annual Review of Genetics*, **33**, 565-602.
- Wilson, K.S. and von Hippel, P.H. (1995) Transcription Termination at Intrinsic Terminators - the Role of the Rna Hairpin. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 8793-8797.
- Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Analytical Biochemistry*, **179**, 131-137.
- Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber, R. and Wahl, M.C. (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Molecular Cell*, **7**, 1177-1189.
- Xia, T.B., Frankel, A., Takahashi, T.T., Ren, J.S. and Roberts, R.W. (2003) Context and conformation dictate function of a transcription antitermination switch. *Nature Structural Biology*, **10**, 812-819.
- Young, C.G. (1997) Models for the molybdenum hydroxylases. *Journal of Biological Inorganic Chemistry*, **2**, 810-816.
- Zor, T. and Selinger, Z. (1996) Linearization of Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Analytical Biochemistry*, **236**, 302-308.

7 Appendix

7.1 Abbreviations

Å	Ångström, 1 Å = 10 ⁻¹⁰ m
AR1-AR2	Acidic repeat 1 and 2
ARM	Arginine-rich motif
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
CD	Circular dichroism
°C	Celsius degree
ddH ₂ O	<i>aqua bidestilata</i>
DEAE	Diethyl aminoethyl
DESY	Deutsches Elektronen Synchrotron
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
<i>ecNusB</i>	<i>Escherichia coli</i> NusB
<i>g</i>	9.81 m/s ²
HCl	Hydrochloric acid
HhH	helix-hairpin-helix
HTH	Helix-turn-helix motif
INT	Iodonitrotetrazolium chloride
ITC	Isothermal titration calorimetry
K	Kelvin degree
K _{av}	Partition coefficient
Kb	Kilo base
K _d	Dissociation constant
KDa	Kilo Dalton, 1 Da = 1 g mol ⁻¹
M	Molarity (mol/l)
MAD	Multi-wavelength anomalous dispersion
M _r	Relative molecular mass
MDa	Mega Dalton
MES	Morpholinoethanesulfonic acid
min	minutes
MR	Molecular replacement
mRNA	messenger RNA
<i>mtNusB</i>	<i>Mycobacterium tuberculosis</i> NusB
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCS	Non-crystallographic symmetry
nm	nanometer
NMR	Nuclear magnetic resonance
Nus	N-utilization substance
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
PMSF	Phenylmethanesulfonyl fluoride
ppm	parts per million
Qor	Quinoline 2-oxidoreductase
r.m.s.d.	root-mean-square deviation
RNA	Ribonucleic acid

APPENDIX

RNAP	RNA polymerase
rRNA	ribosomal RNA
SAM	Sterile alpha motif
SDS	Sodium dodecylsulfate
TCA	Trichloroacetic acid
Tris-HCl	N-Tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
<i>tmNusB</i>	<i>Thermotoga maritima</i> NusB
v/v	volume per volume
w/v	weight per volume

Nucleic Acid Bases

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

Amino Acids

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartate	P	Pro	Proline
E	Glu	Glutamate	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

7.2 Index of Figures

Figure 1. The mononuclear molybdenum enzyme family.....	13
Scheme 1. Hydroxylation reaction catalyzed by the molybdenum hydroxylases.	14
Figure 2. Pathway of quinoline degradation.....	16
Scheme 2. Catalytic reaction of Qor.	17
Figure 3. The molybdenum cofactor Moco.....	18
Figure 4. Structures of the most common types of iron-sulfur centers.	20
Figure 5. The canonical [2Fe-2S] cluster.....	21
Figure 6. Chemical structure of the FAD moiety.....	22
Figure 7. Yeast RNAP II.....	24
Figure 8. Crystal structure of a CAP-DNA complex.....	25
Figure 9. Variant of the λ tR2 terminator.....	26
Figure 10. Rho-dependent termination.....	27
Figure 11. The <i>trp</i> operon.....	28
Figure 12. Functional segments of λ N.....	30
Figure 13. Model of the λ N-mediated antitermination complex.....	31
Figure 14. Domain arrangement in NusA from <i>T. maritima</i>	32
Figure 15. Qor purification.....	47
Figure 16. Crystals of Qor.....	48
Figure 17. Diffraction image of Qor crystals.....	49
Figure 18. Ribbon plot representation of the Qor dimer and arrangement of its cofactors.....	53
Figure 19. Ribbon plot representation of the iron-sulfur protein (QorS).....	54
Figure 20. Ribbon plot representation of the flavoprotein (QorM).....	55
Figure 21. Shielding of the isoalloxazine ring of FAD from the solvent by Tyr190.....	56
Figure 22. Ribbon representation of the molybdoprotein (QorL).....	57
Figure 23. Alignment of ten sequences of bacterial molybdenum hydroxylases.....	58
Figure 24. Stereo view of the electrostatic surface of the active site channel.....	59
Figure 25. Stereo view of the Moco.....	60
Figure 26. Stereo view of the active site residues.....	62
Figure 27. Stereo view of the active site superposition of Qor, <i>oc</i> CODH and <i>rc</i> XDH.....	62
Figure 28. Comparison of the active sites of Qor and <i>dg</i> ALO.....	64
Figure 29. Stereo view of the observed positive density in the active site of Qor.....	65
Figure 30. Binding of the λ N-derived peptide λ N(34-40) by NusA AR1-AR2.....	69
Figure 31. Circular dichroism spectra of NusA AR1-AR2.....	70
Figure 32. Crystal of (NusA AR1) ₂ - λ N.....	71
Figure 33. Sequence alignment and stereo plot of the (NusA AR1) ₂ - λ N complex.....	75
Figure 34. Mass spectrometric analysis on NusA AR1-AR2.....	76
Figure 35. Electrostatic surface potentials of the binding partners.....	78
Figure 36. Electrophoretic gel mobility shift analysis on NusA AR1-AR2- λ N- <i>nut</i> RNA.....	80
Figure 37. Probing of the NusA- λ N interactions by ITC.....	81
Figure 38. Three orthogonal stereo views detailing the AR1- λ N(34-40) interaction.....	85
Figure 39. Crystals of <i>tm</i> NusB.....	91
Figure 40. Close-up stereo ribbon plot of the <i>tm</i> NusB monomer.....	94
Figure 41. Superposition of known NusB structures.....	95
Figure 42. Alignment of nine bacterial NusB sequences.....	96
Figure 43. Stereo ribbon plot of a <i>tm</i> NusB monomer as seen in crystal form 3.....	97
Figure 44. Stereo ribbon plot of a <i>tm</i> NusB dimeric arrangement as seen in crystal form 4.....	98
Figure 45. Stereo plot of the aromatic interactions in the dimer interface of crystal form 4.....	99
Figure 46. Electrostatic surface potential.....	100
Figure 47. Close-up view of the dimer interface in default orientation.....	100
Figure 48. Superposition of the dimeric <i>mt</i> NusB structure on a <i>tm</i> NusB.....	101
Figure 49. Oligomeric state of <i>tm</i> NusB in solution.....	102
Figure 50. RNA binding of <i>tm</i> NusB.....	104

Figure 51. Binding region of one citrate ion in the form 5 structure.....	105
Figure 52. Surface topology of <i>tmNusB</i>	107

7.3 Index of Tables

Table 1. Selected representatives of the molybdenum hydroxylase enzyme family.....	14
Table 2. Quinoline minimal medium according to Tshisuaka et al., 1993.....	36
Table 3. Data collection and refinement statistics for Qor.....	51
Table 4. Kinetic parameters of recombinant Qor and Qor variants.....	67
Table 5. Secondary structure elements calculated for the NusA AR1-AR2 CD-spectrum.....	70
Table 6. Crystallographic data for the (NusA AR1) ₂ -λN complex.....	73
Table 7. Interactions of NusA AR1-AR2 with λN peptides.....	83
Table 8. Crystallization conditions for <i>tmNusB</i>	90
Table 9. Data collection and refinement statistics for <i>tmNusB</i>	92

7.4 Curriculum vitae

Personal Data

Name: Irena Bonin
Place of birth and Date: Koper, 14.08.1973

Education and Research Activities

2000-present	Ph.D. Thesis at the Department of Structural Biology headed by Prof. Robert Huber, FRS, Max-Planck-Institute for Biochemistry, Martinsried, Germany
1999-2000	Graduate Research Assistant International Centre for Science and High Technology (ICS-UNIDO), Trieste, Italy <i>"Computer Design of New PR HIV-1 Inhibitors"</i> Supervisor Prof. Dr. Stanislav Miertus
1999	Graduate Research Assistant University of Trieste, Chemical Sciences Department, Italy. <i>"Design and Synthesis of Pseudopeptide Inhibitors of HIV-1 Protease"</i> Supervisor Prof. Dr. Fabio Benedetti
1998	Master of Science in Chemistry and Pharmaceutical Technology, University of Trieste, Italy. Master thesis: <i>"Usage of Computer Chemistry in Planning New Inhibitors against HIV Aspartic Protease"</i> Supervisor Prof. Dr. Domenico Romeo

1991 – 1998

Student of the Faculty of Pharmacy, University of Trieste, Italy

Publications

Bonin, I., Martins, B.M., Purvanov, V., Fetzner, S., Huber, R. and Dobbek, H. (2004). Active site geometry and substrate recognition of the molybdenum hydroxylase quinoline 2-oxidoreductase. *Structure*, **12**, 1425-1435.

Bonin, I., Robelek, R., Benecke, H., Urlaub, H. Bacher, A., Richter, G., and Wahl, M.C. (2004). Crystal structures of the antitermination factor NusB from *Thermotoga maritima* and implications for RNA binding. *Biochem. J.* (in press)

Bonin, I., Mühlberger, R., Bourenkov, G.P., Huber, R., Bacher, A., Richter, G. and Wahl, M.C. (2004). Structural basis for the interaction of *Escherichia coli* NusA with protein N of phage λ . PNAS (in press)

Nair, A.C., **Bonin, I.**, Tossi, A., Welsh, W.J. and Miertus, S. (2002). Computational studies of the resistance patterns of mutant HIV-1 aspartic proteases towards ABT-538 (ritonavir) and design of new derivatives *J. Mol. Graph. & Model.* **21**, 171-179.

Marcorin, G.L., Da Ros, T., Castellano, S., Stefancich, G., **Bonin, I.**, Miertus, S. and Prato, M. (2000). Design and synthesis of novel [60]fullerene derivatives as potential HIV aspartic protease inhibitors. *Org. Lett.* **2**, 3955-3958.

Nair, A.C., **Bonin, I.**, Tossi, A., Welsh, W.J. and Miertus, S. (2000). Computational studies of the resistance patterns of mutant HIV-1 aspartic proteases toward ABT-538 (ritonavir) and design of new derivatives. Abstracts of Papers of the American Chemical Society 219: 140-COMP Part 1.

Tossi, A., **Bonin, I.**, Antcheva, N., Norbedo, S., Benedetti, F., Miertus, S., Nair, A.C., Maliar, T., Dal Bello, F., Palu, G. and Romeo, D. (2000). Aspartic protease inhibitors - An integrated approach for the design and synthesis of diaminiol-based peptidomimetics. *Eur. J. Biochem.* **267**, 1715-1722.

Kotamarthi, B., **Bonin, I.**, Benedetti, F. and Miertus, S. (2000). Computational design of new cyclic urea inhibitors for improved binding of HIV-1 aspartic protease. *BBRC* **268**, 384-389.