Charles University in Prague, Faculty of Science
Molecular and Cell Biology, Genetics and Virology

Molecular dynamics of proteins interacting with substrate or ligand: mitochondrial processing peptidase and FixL oxygen sensor

RNDr. Klára Dvořáková Holá

Ph.D. Thesis
2007

Supervisor: Ing. Jiří Janata, CSc.
Institute of Microbiology AS CR

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... Klára Dvořáková Holá

Date 14^{the} Jun 2007

ACKNOWLEDGEMENTS

I would like to thank the Hlavka Foundation for a financial support covering the travel costs to the 9th International Symposium on the Genetics of Industrial Microorganism in 2002 (Gyeongju, Korea) as well as to my fellowship at the Ecole Polytechnique in 2004 (Palaiseau Cedex, France).

I am deeply indebted to my supervisor Ing. Jiří Janata, CSc. for offering me the opportunity to participate in a project focused on protein dynamics of mitochondrial processing peptidase and for all kinds of support.

I am obliged to Dr. Marten H. Vos and Dr. Ursula Liebel for giving me the chance to meet a research on dynamics of FixL oxygen sensor as a Marie Curie Training Site PhD student.

Finaly, I wish to express my appreciation to my husband Robert and to my parents for understanding.

LIST OF ABBREVIATIONS

*Bj*FixLH	heme domain of FixL from *Bradyrhizobium japonicum*
FixLH	heme domain of FixL
FG loop	position Thr^{209} to Arg^{220} in FixL from *Bradyrhizobium japonicum*
GRL	glycin rich loop
α-MPP*	modified α-MPP with a single remaining tryptophan residue (Trp^{223})
Mb	myoglobin
MD	molecular dynamics
MPP	mitochondrial processing peptidase
pMDH	malate dehydrogenase precursor
TR^3	time resolved resonance Raman

CONTENTS

1. List of publications 5

2. Overview 6

3. Ligand binding and protein dynamics 7

4. Mitochondrial processing peptidase (MPP) 9

    4.1. MPP: Introduction 9

    4.2. MPP: Aims of experiments 15

    4.3. MPP: Resuls and discussion 15

5. FixL oxygen sensor 22

    5.1. FixL: Introduction 22

    5.2. FixL: Aims of experiments 28

    5.3. FixL: Results and discussion 28

6. Methods 32

    6.1. Molecular biology approach: site-directed mutagenesis, gene expression, protein purification 32

    6.2. Basic principles of used spectroscopic methods 33

7. Proteins involved in biosynthesis of secondary metabolites 38

8. References 40

1. LIST OF PUBLICATIONS

(1) **Dvorakova-Hola K.**, Matuskova A., Parkhomenko N., Kubala M., Vecer J., Herman P., Gakh O., Kutejova E., Amler E. and Janata J. **(2007)** Glycine-rich loop of mitochondrial processing peptidase α-subunit recognizes substrates by analogical mechanism with mitochondrial receptor Tom20. *Prepared for submission to J. Mol. Biol.*

(2) Kruglik S.G., Jasaitis A., **Hola K.**, Liebl U., Martin J.L. and Vos M. H. **(2007)** Subpicosecond oxygen trapping in the heme pocket of the oxygen sensor FixL observed by time-resolved resonance Raman spectroscopy. *Proc. Natl. Acad. Sci. U S A* 104: 7408-7413

(3) Jasaitis A., **Hola K.**, Bouzhir-Sima L., Lambry J.C., Balland V., Vos M.H. and Liebl U. **(2006)** Role of distal arginine in early sensing intermediates in the heme domain of the oxygen sensor FixL. *Biochemistry* 45: 6018-6026

(4) Janata J., **Hola K.**, Kubala M., Gakh O., Parkhomenko N., Matuskova A., Kutejova E. and Amler E. **(2004)** Substrate evokes translocation of both domains in the mitochondrial processing peptidase α-subunit during which the C-terminus acts as a stabilizing element. *Biochem. Biophys. Res. Commun.* 316: 211-217

(5) **Hola K**., Janata J, Kopecky J. and Spizek J. **(2003)** LmbJ and LmbIH protein levels correlate with lincomycin production in *Streptomyces lincolnensis. Lett. Appl. Microbiol*. 37: 470-474

(6) Janata J., Najmanova L., Novotna J., **Hola K.**, Felsberg J. and Spizek J. **(2001)** Putative *lmbI* and *lmbH* genes form a single *lmbIH* ORF in *Streptomyces lincolnensis* type strain ATCC25466. *Antonie Van Leeuwenhoek* 79: 277–284

2. OVERVIEW

The presented doctoral thesis includes five published scientific articles and one manuscript prepared for submission. All describe studies on three protein models. The four papers numbered (1) – (4) in the list of publications, share a common objective e.i. observing and describing functional protein dynamics and conformation change induced by ligand or substrate binding, and represent the main result of my PhD work.

The papers (1) and (4) offer results of a project from my home laboratory at the Institute of Microbiology AS CR, Laboratory for Biology of Secondary Metabolism, under the supervision of Jiří Janata, PhD. The project is focused on the protein-protein dynamics interaction of mitochondrial processing peptidase (MPP) from *Saccharomyces cerevisiae* with its preproteins substrates.

The papers (2) and (3) describe results of a project, in which I have participated during my Marie Curie fellowship at the Ecole Polytechnique (Palaiseau Cedex, France), Laboratory for Optics and Biosciences, in 2004. The project concerns research on protein structural dynamics of the heme-based oxygen sensor FixL from *Bradyrhizobium japonicum*, in which oxygen binding to the heme sensor domain induces conformation change, which regulates the activity of neighboring kinase domain.

In both projects, analogy in methodical approach, i.e. series of molecular biology and biochemistry techniques, was used; the highly conserved amino acid residue and/or region, which are suggested to play a key role in ligand binding and discrimination, were modified by site directed mutagenesis and the protein mutant forms were purified (Chapter 6.1.). The ligand binding dynamics of the mutant forms was examined by various biophysical methods: i) the precursor protein binding to MPP was observed by time-resolved and steady-state fluorescence spectroscopy ii) the dynamics of the oxygen binding to the FixL heme domain was demonstrated by femtosecond transient absorption spectroscopy and time-resolved resonance Raman spectroscopy. The basic principles of all used spectroscopic techniques are described in Chapter 6.2. Finally, the molecular dynamics simulations were performed in both projects to gain insight into the mechanism of the ligand binding, recognition and ligand specificity at a molecular level.

The publications (5) and (6) give the results of another project from the Laboratory for Biology of Secondary Metabolism – a study on proteins involved in the biosynthesis of

secondary metabolites. Although the used methodical approach (techniques of molecular biology and biochemistry) was similar to those of the above mentioned projects, the aims of the experiments were different from the observing the protein dynamics. Thus, these papers are only very briefly summarized in Chapter 7.

3. LIGAND BINDING AND PROTEIN DYNAMICS

The functions of the most proteins, whether signaling or transport or catalysis, depend on the ability to bind other molecules − ligands or substrates. As the protein ligand is a molecule (small molecule or a macromolecule) that recognizes and binds to specific site on the protein surface, the enzyme substrate is a molecule that binds to the enzyme active site and is subsequently transformed in an enzymatic reaction. In the case of MPP protein model, we monitor just the initial phase of the substrate-enzyme interaction, i.e. recognition, (not following steps ending in a cleavage reaction), thus, the term protein "ligand" can be used for the MPP preprotein substrate in this phase of interaction.

Binding specificity arises from complementarity of shape, charge distribution and from the distribution of donors and acceptors of hydrogen bonds. Small-molecule ligands usually bind at depressions on the protein surface (pockets, cavities or clefts). Deep binding pockets allow protein to envelop the ligand and thus use complementarity of shape to provide specifity. Clefts or cavities can easily provide unusual microenvironments and enable even a small molecule to have enough contact points to bind strongly if that is needed. On the other hand, the specific recognition of a macromolecule by the protein usually involves interactions over a large surface area (hundreds of sqare Ångstroms) or over several discrete binding regions. The most frequently observed binding sites are protruding loops or large cavities because these provide specific shape complementarity, but relatively flat binding sites are also found.

The classic model for the ligand specific binding is that of a "key fitting into lock". However, this analogy implies rigidity of the protein (the lock) and of the ligand (the key). In reality, the proteins (and in case of larger molecules also ligands) are naturally flexible, so the classic view has been augmented by a model of induced fit: During binding, each can adjust its structure to the presence of the other and the protein can be said to "snuggle" around the substrate. Change in the protein conformation may be necessary for binding to occur and serves to enclose the substrate, thereby preventing its release from the protein and ideally

positioning it for the protein to perform its function. Alternatively, even a small change in the structure of a ligand or protein can abolish binding. A special class of conformational transitions is found in so-called allosteric proteins. Substrate binding to one subunit of these multimeric proteins triggers a conformational change that alters the substrate afinity of the other subunits, thereby sharpening the switching response of these proteins.

The protein conformational changes range from very subtle, local changes, to global movements, for example:

- Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)
    - Atomic fluctuations
    - Sidechain Motions
    - Loop Motions
- Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)
    - Helix Motions
    - Domain Motions
    - Subunit motions
- Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)
    - Helix coil transitions
    - Dissociation/Association
    - Folding and Unfolding

The present thesis reports two examples of a protein conformation change of different extents and time scales:

- **In the case of mitochondrial processing peptidase we studied the domain motion as well as local loop motion on the nanosecond time scale.**
- **On the other hand, the change of the conformation of the heme as the prosthetic group of FixL protein was observed on the picosecond and/or subpicosecond time scale.**

4. MITOCHONDRIAL PROCESSING PEPTIDASE

4.1. MPP: Introduction

In 1999, Günter Blobel received the Nobel Prize in Physiology or Medicine for discovering that proteins have intrinsic signals that govern their transport and localization in the cell; phenomenon called protein targeting. Targeting signals are contained in the polypeptide chain or in the folded protein. The continuous stretch of amino acid residues is called targeting or signal peptide. There are two types of signal peptides: The internal targeting peptides and the presequences that are usually found at the N-terminal extension of the proteins.

Mitochondria make up ~20% of the mass of eukaryotic cells and contain ~1000 different proteins. Only a few mitochondrial proteins are encoded by the organellar genome and synthesised in the innermost compartment, the matrix. Over 98% of mitochondrial proteins are encoded by nuclear genes, synthesised as precursor proteins on cytosolic ribosomes and posttranslationally transported to the mitochondria (Schatz and Dobberstein, 1996; Bauer *et al*., 2000; Pfanner and Wiedemann, 2002). Specific transport of the proteins into mitochondria is facilitated by mitochondrial targeting signals, most often in the form of N-terminal presequences, which are recognized by a series of protein translocases. The presequences are proteolytically removed during or subsequent import into the mitochondria by three types of processing peptidases. **Mitochondrial processing peptidase (MPP) located in the matrix is a metallopeptidase that cleaves off most N-terminal presequences from precursor proteins** (Pfanner and Neupert, 1990; Ito, 1999; Gakh *et al*., 2002).

The mitochondrial presequences vary in lenght and share a little sequence similarity (Pfanner and Neupert, 1990; Glick *et al*., 1992; Hartl and Neupert, 1990), but each is cleaved at a single specific site by MPP. The lengths of the known presequences of MPP substrates vary from 6 to 136 amino acid residues, the majority falling between 20-50 residues. Some general sequence features was identified as follows: A proximal basic amino acid residue, usually arginine, at the -2 position in respect to the cleavage site, distal N-terminal basic residues generally 3-10 residues from the proximal arginine and an aromatic or less often another type of bulky hydrophobic residue at +1 position (Hendrick *et al*., 1989; Gavel and von Heijne, 1990; Niidome *et al*., 1994; Song *et al*., 1996). These attributes are proposed to be useful for a

cleavage reaction in MPP active site, localised inside enzyme cavity. However, a general "entering password" of a large variety of diverse presequences essential for a primary contact and recognition by MPP remained unresolved.

In spite of the lack of a distinct sequence motif, mitochondrial presequences display common physicochemical properties – a positive charge and a potential to form an amphiphilic α-helix. (Roise *et al*, 1986; 1988; von Heine, 1986). Some presequences were found to adopt a helical conformation in the presence of detergent micelles and organic solvent, *i.e.* in a membrane model system (Hammen *et al*., 1994; Chupin *et al*. 1996; De Jongh, 2000). It has been proposed that both surfaces of mitochondrial presequences, the positively charged one and the hydrophobic one, are recognised by different import components during the preprotein import from cytosol to mitochondria (Brix *et al*., 1997).

Mitochondrial protein import from cytosol to matrix

Mitochondrial proteins possessing N-terminal presequences cross the mitochondrial membranes by virtue of several molecular machines, (Neupert, 1997; Pfanner *et al*., 1997, Phanner and Wiedemann, 2002); a TOM complex in the outer membrane and one of two TIM complexes in the inner membrane. The TOM complex is the **t**ranslocase in the **o**uter **m**itochondrial membrane, responsible for recognition and transport of the proteins into the intermembrane space (**Fig. 1a**). The core of this complex contains Tom40, Tom22, Tom7 and Tom6. The protein-translocation pore is formed from Tom40. The function of the TOM complex is completed by the Tom20 and Tom70 receptors, which recognize and bind the wide-range of proteins targeted to mitochondria. Previous studies suggested that Tom70 functions as a receptor for a subset of proteins as well as a docking site for a cytosolic targeting factor (Hines and Schatz, 1993; Hachiya *et al*., 1995; Komiya *et al*., 1997).

Tom20 functions as a general import receptor for mitochondrial proteins and passes them onto the Tom40 channel (Ramage *et al*., 1993). After passing through the Tom40 channel, substrate proteins can interact with **t**ranslocases of **i**nner mitochondrial **m**embrane, including Tim23 forming a channel that enables positively charged presequences to continue into mitochondrial matrix. Although the positive charges of the presequence are essential during import process, **presequence binding to the "entrance" of the import machinery – receptor Tom20 – is mediated mainly by hydrophobic rather than ionic interactions (Abe *et al*., 2000).** The NMR structure of rat Tom20 in complex with a presequences demostrated that the bound presequence forms an amphiphilic helical structure with hydrophobic leucines

aligned on one side to interact with a hydrophobic patch in the Tom20 groove (**Fig. 1b**). Additionally, the experiments investigating mutant forms of the presequence clearly proved that the rat **Tom20 recognizes the potential of the presequence to form an amphiphilic α-helical structure, but not its positive charge (Abe *et al*., 2000).**

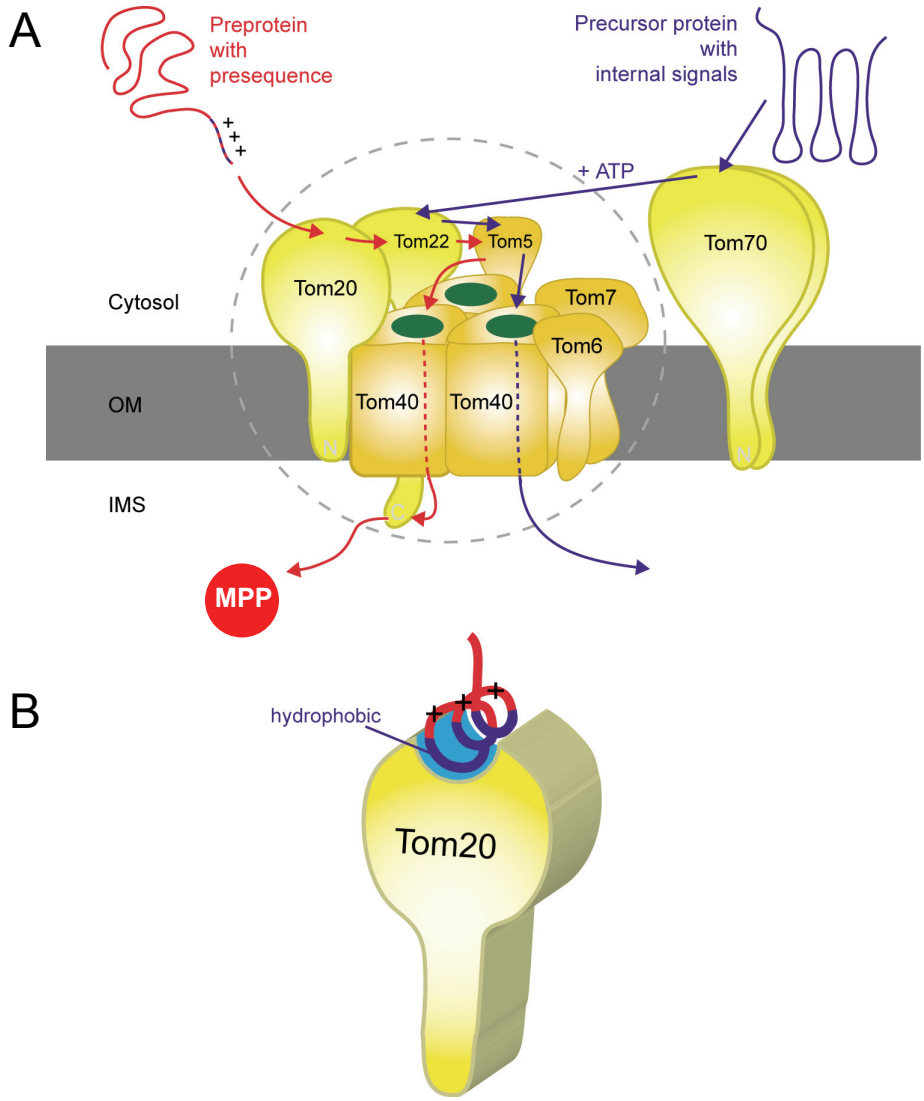

Fig. 1. The preprotein translocase of the outer mitochondrial membrane. **(a)** The TOM complex contains three main receptors:Tom20, Tom22 and Tom70. Preproteins with a presequence are initially recognised by Tom20, transferred to Tom22, Tom5 and Tom40, and translocated across the outer membrane (OM). When the presequence emerges on the inter-membrane space (IMS) side of the TOM complex, it can bind to the carboxy-terminal domain of Tom22 before the transfer to the inner membrane. **(b)** Model of the cytosolic domain of the receptor Tom20 with a bound presequence. Mitochondrial presequences can form amphiphilic α helices. The presequence binds with the hydrophobic side to a hydrophobic binding groove of Tom20. (derived from Phanner and Wiedemann, 2002; Figure2)

Perry et al. (2006) reported that comparison of the sequence and structure of Tom20 from plants and animals suggested that these two presequnce binding receptors evolved from

two distinct ancestral genes. However, the need to bind equivalent mitochondrial presequences has driven the convergent evolution of two distinct proteins to a common structure and function. The stuctural features of plant Tom20s are equivalent to the structure of Tom20 in animals and fungi, but in reverse order.

Mitochondrial processing peptidase localized in matrix

MPP consists of two homologous subunits α and β **(Fig. 2)**, both essential for its enzymatic activity. The β-subunit contains an HxxEH zinc binding motif, which is an inversion of the thermolysin zinc binding motif, HExxH (Becker and Roth, 1992). Although the α and β subunits of MPP have very similar sequences (i.e., 48% identical residues or conservative replacement between *Saccharomyces cerevisiae* MPP subunits), the catalytic center of containing a zinc binding motif is conserved only in the β-subunit and is localized in a polar internal cavity inside the enzyme dimer. Crystal structures of yeast MPP demonstrated the nearly identical protein architecture of the two subunits; each consists of two domains of ~210 residues with very similar folding topology, which are connected by a flexible linker of ~20 residues (Taylor *et al.*, 2001). Thus, the MPP dimer consists of four structurally related domains.

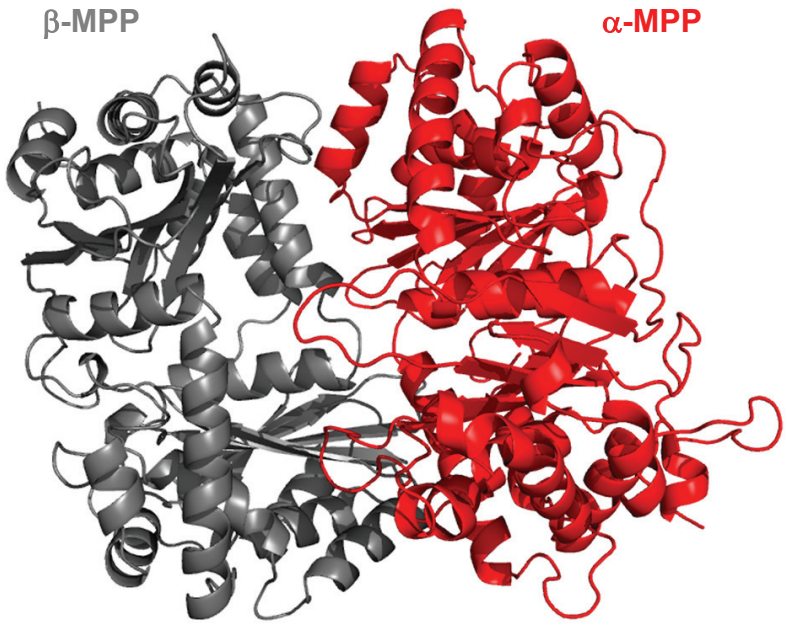

Fig. 2. Crystallographic structure of the yeast MPP dimer; α-subunit (red) and β-subunit (grey). Protein Data Bank with ID code 1HR6 (Taylor *et al.*, 2001) The figure was prepared using PYMOL.

As Tom20 import receptor binds the presequences in the α-helical structure, a rather structureless presequence form was predicted to bind MPP because the precursor polypeptide, including the presequence, undergoes unfolding and conformational changes during protein import and the maturation process (Kojima *et al.*, 2001). This prediction is consistent with the crystal structures of two different synthetic substrate peptides co-crystallized with mutant MPP, deficient in cleavage function, showing the presequences bound in an extended conformation at the active site (Taylor *et al.*, 2001). Nevertheless, this finding reflects the final step of enzyme-presequence interaction (after fixation of the extended presequence directly towards the active site) in which conformation of the presequence may differ from the primary interaction.

Although the roles of each subunit in substrate binding are not indisputable yet, most of studies pointed out the relevancy of the α-subunit in substrate binding and recognition. First, cross-linking and surface plasmon resonance analyses showed that the α-MPP subunit alone, but not β-MPP subunit, binds substrates as efficiently as the MPP complex (Yang, *et al.*, 1991; Luciano *et al.*, 1997). Second, truncation of the α-MPP C-terminal 41 amino acid residues led to a loss of binding and processing activity (Shimokata *et al.*, 2000). Finally, lifetime analysis of tryptophan fluorescence of the yeast MPP showed that the presence of the substrate evokes a conformational change of α-MPP, while in the case of the β-MPP neither the lifetime distribution nor the average lifetime of the tryptophan excited state is influenced (Gakh *et al.*, 2001). However, it remained unclear which particular regions of the α-MPP are responsible for the primary substrate-enzyme interaction.

Highly conserved glycine-rich loop of the α-MPP

It appears that the essential function of the α-subunit is localized just in the most conserved part of all known α-MPPs – the glycine-rich site. The point mutations in this region notably affected the MPP activity; particularly deletions completely destroyed the enzyme activity (Nagao *et al.*, 2000).

This sequence motif (HfbHfbGGGGSFSAGGPGKGMF/YSRLYT/LxVLN; capitals mark the highly conserved residues, Hfb indicates hydrophobic residues) is present in all known α-MPPs, in the yeast α-MPP occupies positions 284-308. Besides the characteristic glycine residues, there are also several conserved hydrophobic and/or aromatic amino acid residues. **The glycine-rich region forming the flexible loop is situated at the entrance into the MPP**

dimer cavity and is exposed to both the catalytic centre of the β-MPP and to the substrate. The intrinsic fully unstructured loop (glycine-rich loop; GRL) is bordered by a short α-helix **(Fig. 3)**.

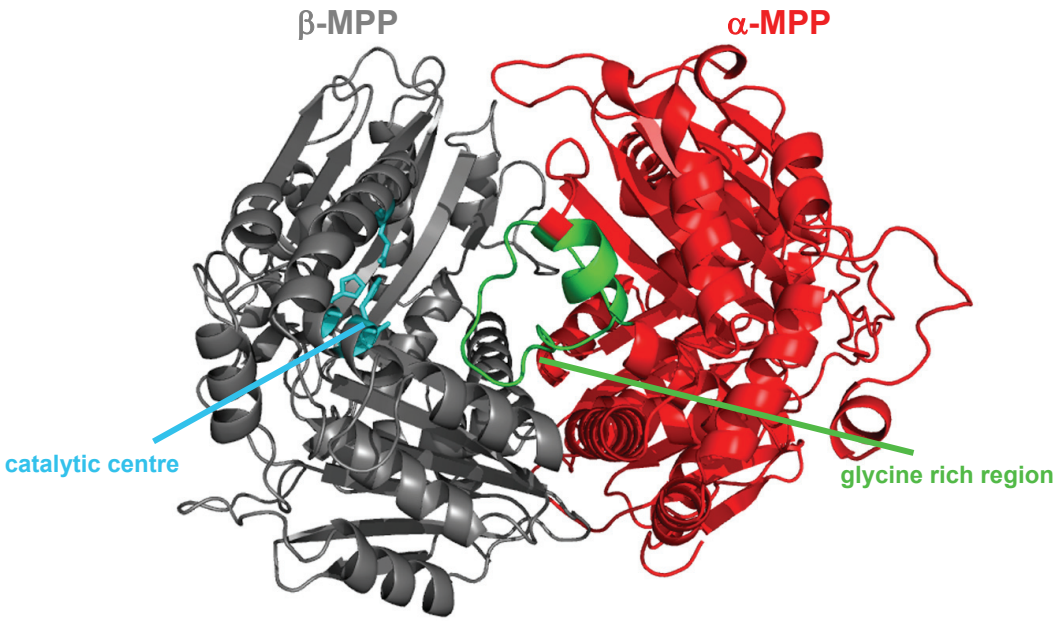

Fig. 3. Structure of the yeast MPP dimer; the highly conserved glycine-rich site of the α-MPP (green) is surface-exposed as well as exposed to the catalytic center (cyan) of the β-MPP. The figure was prepared using PYMOL.

4.2. MPP: Aims of experiments

1. To characterize overall conformation change of the yeast α-MPP subunit in the presence of substrate and to investigate the role of C-terminus of the α-MPP in the substrate binding.

→ *Janata et al., 2004*

2. To monitor the local dynamics of the glycine-rich loop of the yeast α-MPP in the presence of substrate and to confirm this region as the "primary contact site" of MPP with the substrate presequence.

→ *Dvorakova-Hola et. al., 2007*

4.3. MPP: Results and discussion

*Janata J., **Hola K.**, Kubala M., Gakh O., Parkhomenko N., Matuskova A., Kutejova E. and Amler E. (2004) Substrate evokes translocation of both domains in the mitochondrial processing peptidase α-subunit during which the C-terminus acts as a stabilizing element. Biochem Biophys Res Commun 316: 211-217*

The previous results of our laboratory demonstrated a conformational change of α-MPP in the presence of substrate (Gakh *et al.*, 2001): Lifetime analysis of tryptophan fluorescence of the yeast MPP subunits proved that substrate binding evoked a conformational change of α-MPP, but not β-MPP. In order to determine more precisely which region of α-MPP is responsible for the conformational change detected previously, all three tryptophan residues presented in the yeast α-MPP native sequence (Trp^{147}, Trp^{223} and Trp^{481}) were subsequently substituted (**Fig. 4**). Alignment performed on all known α-MPPs revealed that positions corresponding to these Trp residues are not conserved. The substitutions of Trp^{147} and/or Trp^{481}

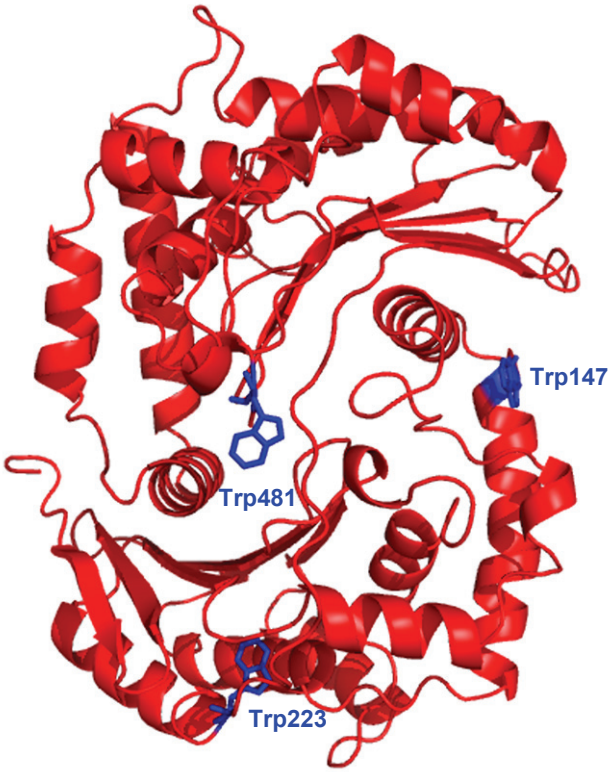

Fig. 4. Structure of the yeast α-MPP. The positions of the all three native tryptophan residues in blue. The figure was prepared using PYMOL.

did not affect neither the solubility of the proteins nor processing activity of the MPP enzyme. On the contrary, amino acid changes at position 223 surprisingly yielded insoluble protein. Consequently, we prepared a modified α-MPP with a single remaining Trp^{223} residue (α-MPP-W147N/W481Y, in the following text denoted as **α-MPP***). In that case, solubility of the protein as well as processing activity of the MPP dimer was comparable to that of the wild type.

Steady-state tryptophan fluorescence measurement of yeast α-MPP with reduced one or two tryptophan residues (W147N and W147N/W481Y), provided a detailed view of the α-subunit conformational change evoked by substrate: **Translocation of both, the N- and C-terminal domains of α-MPP as well as the conformational change of the linker connecting both domains were demonstrated. In particular, the C-terminal penultimate Trp^{481} residue approaches to the Trp^{223} localized at the border of the N-terminal domain and interdomain linker in the middle portion of the protein.** In case of the α-MPP containing both tryptophan residues (Trp^{223} and Trp^{481}), there was a negative peak in the short-

wavelength part of the spectrum in the presence of the substrate (yeast malate dehydrogenase precursor; pMDH). Because this part of the spectrum is dominated by the Trp^{223} emission, it was deduced that Trp^{223} is quenched as a donor in an energy transfer from Trp^{223} to Trp^{481} when precursor substrate is present. **Moreover, Trp^{223}, buried into the hydrofobic core of the subunit, becomes more exposed to the solvent in the presence of substrate,** as one may deduce from the red shift of the spectral maximum. It could reflect relaxation of the flexible interdomain linker accompanying the radical overall subunit conformational change during the substrate recognition process.

Based on these results, we propose a putative mechanism of MPP action: 1. A specific region of α-MPP is responsible for the weak primary interaction with the presequences of protein substrates. The highly conserved glycine-rich region localized at the entrance into the MPP dimer cavity is a potential candidate of the "primary contact site". 2. The weak primary contact of the substrate with the specific region of α-MPP acts as the release pulse of the short but distinct overall conformational change of α-MPP. The documented α-MPP conformational change could be related to the initiation of the presequence entrance to the polar cavity of the dimer. 3. In cooperation of both MPP subunits, the presequence is fully pulled into the cavity, extended, strongly fixed and directed towards the active site of the β-subunit (Kojima *et al*., 1998; Taylor *et al.*, 2001).

The additional experiments of this study investigated the role of C-terminus of the α-MPP, as the data obtained from tryptophan fluorescence measurements indicated considerable complicity of the region in the overall subunit conformational change. Shimokata *et al*. (1998) reported that truncation of the α-MPP C-terminal 41 amino acid residues led to a loss of binding and processing activity. Here we demonstrated that excision of C-terminal 30 amino acid residues (ΔC30) led a to a improper protein folding resulting in a complete loss of protein function. The other two shorter deletions (ΔC17 and ΔC2), affecting only the region without any definable secondary structure, conserve processing activity of the MPP dimer. Nevertheless, any shortening of the α-MPP C-terminus destabilized the protein slightly (Δ2 AA) or dramatically (Δ17 or Δ30 AA). Thus, we suggest that the extreme C-terminus of α-MPP provides mechanical support to the C-terminal domain of the subunit during its conformational change upon substrate binding.

***Dvorakova-Hola K.**, Matuskova A., Parkhomenko N., Kubala M., Vecer J., Herman P., Gakh O., Kutejova E., Amler E. and Janata J. (2007) Glycine-rich loop of mitochondrial processing peptidase α-subunit recognizes substrates by analogical mechanism with mitochondrial receptor Tom20. Prepared for submission to J. Mol. Biol.*

To obtain a system for monitoring the local dynamics of the glycine-rich loop **(GRL)** of the α-MPP using tryptophan fluorescence assays, reporter tryptophan residues were individually introduced at three positions into the GRL or in its vicinity. Individual substitutions of aromatic amino acid residues within the glycine-rich region of the α-MPP* (F289W, Y299W and Y303W; **Fig. 5**) were performed resulting in the following "fluorescent labels" - Trp^{289}, Trp^{299} or Trp^{303}. The introduction of reporters directly to the GRL (Trp^{289}, Trp^{299}) led to a decrease in MPP activity, albeit to a very different extents. A significant portion of the wild-type activity (~ 60%) remained preserved in the case of MPP dimer with substitution Y299W at the border of the GRL. In contrast, the substitution of the absolutely conserved Phe^{289} (F289W) resulted in a dramatic reduction of the processing activity; nevertheless, remaining detectable activity was observed (~ 3 % as compared to wild type). On the other hand, the introduction of the reporter residue within a short conserved α-helix immediately adjacent to the GRL (Y303W) did not influence enzymatic activity at all.

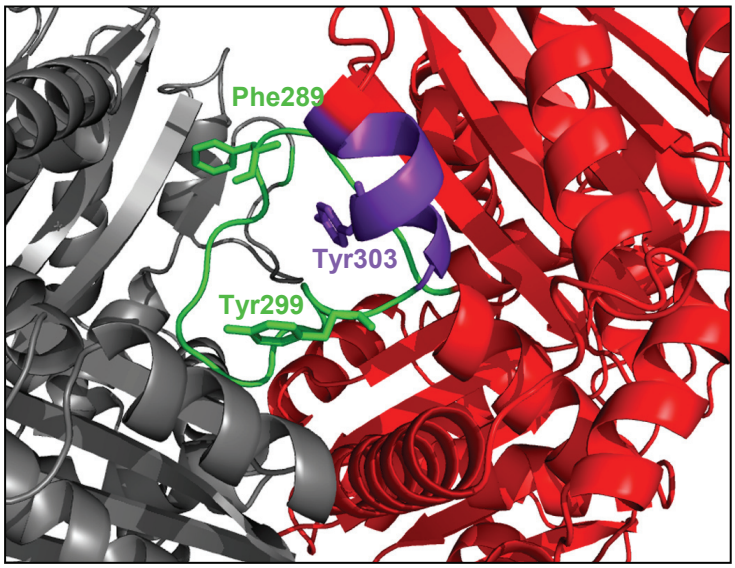

Fig. 5. Detail of the glycine-rich site of the α-MPP. The unstructured loop (green, positions 284-299) is bordered by the short α-helix (purpleblue), formed by residues between positions 301-308. Phe289, Tyr299, Tyr303 were individually replaced by reporter Trp residues. The figure was prepared using PYMOL.

Consequently, we combined this mutation that does not affect protein function with the deletion of a single glycine residue ΔG292 that causes the reduction of the enzyme activity (Gakh O., Ph.D. thesis, 2001). The MPP dimer comprising the α-MPP*-Y303W/ΔG292 did not show any detectable activity. Thus, the two forms of the α-MPP carrying the same reporter residue (Trp^{303}) were prepared to observe the GRL dynamics using tryptophan fluorescence measurement: the fully functional and the functionless one with the modified conformation of the loop by its shortening.

Subsequently, steady-state as well as time-resolved tryptophan fluorescence spectroscopy was used to characterize the local dynamics of the GRL; without substrate and in the presence of yeast pMDH. As expected, the segmental motion of both reporter residues Trp^{289} and Trp^{299}, situated directly in the flexible loop, was demonstrated in the absence of substrate. **The change of the loop dynamics after substrate addition was clearly proved in the case of α-MPP*-Y299W. The decrease of segmental motion evoked by the substrate is significant and reflects a primary contact of the GRL with the presequence.** The evaluation of the interaction dynamics failed in time-resolved fluorescence measurement of the α-MPP*-F289W because no significant changes of an anisotropy decay were detected. We explain this result by the fact that substitution of the absolutely conserved Phe^{289} decreased the GRL function to such a degree that any response to the presequence presence is under a detectable limit of this methodological approach. Nevertheless, **steady-state measurement demonstrated the local conformational change within the loop for both reporter tryptophans.** The Trp^{289} spectrum suggested that the residue becomes more accessible to the solvent while the significant blue-shift of the Trp^{299} spectrum could be a result of a direct shielding of the residue by the substrate amino acid residues.

Both steady-state and time-resolved fluorescence measurement of Trp^{303} indicated that a short α-helix immediately adjacent to the GRL, where this reporter is situated, is rigid and does not participate in the local conformation change. In other words, the conformation change of the GRL evoked by substrate is localized just to its intrinsic region, not to its vicinity. However, the Trp^{303} has served as an excellent observation point for substrate binding: Whereas the substrate binding was clearly demonstrated in the case of the fully active α-MPP*-Y303W, the α-MPP*-Y303W/ΔG292 is defective in substrate binding. **Thus, comparison of a**

pair of the α-MPP mutant forms with the same reporter tryptophan differing in the glycine-rich loop length provided evidence of the essential role of the GRL for initial contact with the substrate.

As our experimental results clearly proved that the GRL is the "primary contact site" of the MPP enzyme with the substrate presequences, we formulated new hypothesis concerning the principle of the substrate recognition by MPP: Regarding the surface exposed localization and a presence of the "hydrophobic spots" within the GRL, **we proposed an analogy between the presequence recognition by hydrophobic binding groove of the Tom20 import receptor and the GRL of the α-MPP.** As both systems were designed to recognize the same set of the substrate presequences, they might share a common "ID code for entrance to a system". Tom20 recognizes the potential of the presequence to form an amphiphilic α-helix and the presequence binding is mediated primarily by hydrophobic interactions (Abe *et al*., 2000). We propose that these features could be the recognition "ID code" required also by MPP enzyme and the GRL (similarly to Tom20 binding groove) induces the formation of a transient helical structural motif of approaching presequence, whereas the structure-less form is assumed to be present in cytosol or matrix.

To verify the proposed interaction of the GRL with the presequence, a molecular dynamics (MD) simulation model was employed. We made use of the fact that the structure of the rat ALDH presequence peptide (residues 12-22) in a complex with the Tom20 receptor has previously been elucidated by NMR (Abe *et al.*, 2000). For MD simulations, we used the initial conformation, posititioning and orientation of the presequence structure based on the analogy with Tom20 binding groove: a hydrophobic side of amphiphilic helix interacting with GRL surface. Besides remarkable surface complementarity between the GRL and peptide was shown (**Fig. 6**) the simulations supported our hypothesis particulary in the following aspect – the stability of proposed interaction between the hydrophobic side of the amphiphilic helix and the GRL was proved. Moreover, the model confirmed a decrease of the loop flexibility in the presence of substrate, in agreement with fluorescent measurement.

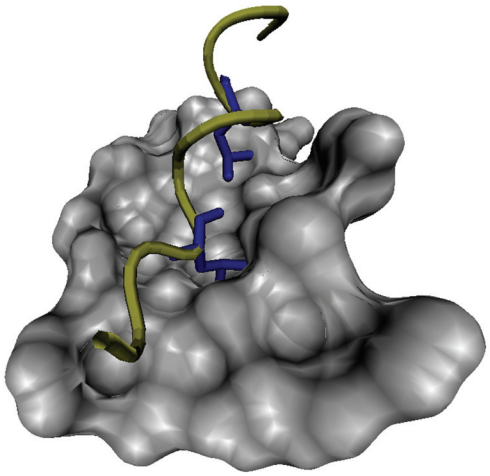

Fig. 6. Representation of the glycine-rich site of the yeast α-MPP interacting with the rat ALDH presequence peptide (residues 12-22) – the initial fitting position for MD simulations. Molecular surface of the GRL (residues 284-307). The interacting peptide is drawn as a tube with side chains in blue for two highly conserved Leu15 and Leu19, respectively. The figure was prepared using VMD.

As the NMR structure of Tom20 in complex with a peptide demonstrated, the side chains of three leucines oriented toward the binding groove (Leu^{15}, Leu^{18} and Leu^{19}) make close contacts with the side chains of residues located in the hydrophobic patch of Tom20 (Abe *et al*., 2000). Our dynamic model clearly showed that whereas the side chain of the Leu^{18} (nonconserved hydrophobic residue) changes its orientation, the two highly conserved Leu^{15} and Leu^{19} (**Fig. 6**) remain in the close contact with the side chains of the GRL during the entire MD simulation. Thus, these two conserved residues appear to be the stabilizing points of the hydrophobic interaction.

We hypothesize that the GRL of the MPP α-subunit is the crucial evolutional outcome of the presequence recognition by MPP and it represents a functional parallel to Tom20 import receptor. The proposed analogy relates just to the moment of the presequence recognition, i.e. the initial step of the protein-protein interaction. Obviously, the immediately following steps are different in both systems. The detailed mechanism of presequence translocation from the glycine-rich loop of the α-MPP to the distant active site at the β-MPP subunit accompanied by extension of the unstable α-helix of the presequence remains to be elucidated.

5. FIXL OXYGEN SENSOR

5.1. FixL: Introduction

The biological reduction of the atmospheric nitogen gas to ammonium – nitrogen fixation – and its subsequent incorporation into biomolecules are essential for life on earth. Only bacteria and blu-green algae are capable of nitogen fixation, an energetically highly unfavourable reaction. Some of these microorganisms, bacteria of the genera Rhizobia and Bradyrhizobium, live in symbiotic ralationships with plants, in specialised structures known as nodules that forms in the root systems of infected plants. The nitrogen fixation proceeds under anaerobic conditions because the nitrogenase complex, which carries out the reduction reaction, is extremely sensitive to oxygen.

Heme proteins have been known to play important roles in biology as oxygen transport (hemoglobin), storage (myoglobin) and oxidation catalysts (cytochrome P450). However, a another class of heme proteins has been discovered that serve as biological sensors, in which the heme cofactor plays a central role not only in binding the respective effector molecules but also in regulating the associated enzymatic function via heme and protein conformational changes induced by ligand binding. The best described heme-based sensors include the **FixL protein of nitrogen fixing bacteria that senses O_2 from *Bradyrhizobium japonicum*** (Gilles-Gonzalez *et al.*, 1991), the soluble guanylate cyclase of vertebrates that senses NO (Garbers and Lowe, 1994) and the CooA protein of *Rhodospirillum rubrum* that senses CO (Shelver *et al.*, 1997).

FixL is a part of signal transduction system that enables nitrogen fixing bacterium to adapt its respiratory energy metabolism to the aerobic or microaerobic state of its environment. The protein is comprised of two distinct domains: An N-terminal heme-based sensory domain and a C-terminal histidine kinase domain. The site of regulation in FixL is located in its heme domain. A conformational change in the heme domain upon oxygen binding to the heme induces a change in the histidine kinase active site – the kinase activity of FixL strongly diminishes.Thus, a local perturbation at the heme domain is transduced over a relatively long distance within the protein. Diminished kinase activity has also been observed with CO or NO adducts, but to a far lesser degree (Tuckerman *et al.*, 2002; Dunham *et al.*,

2003). In the absence of oxygen, the kinase domain of FixL is active and undergoes autophosphorylation and then transfers the phosphoryl group to FixJ, its associated response regulator. Phosphorylated FixJ binds the nifK promoter initiating nitrogen fixation gene expression (Reyrat *et al.*, 1993; Agron *et al.*, 1993).

"Model of heme protein" – myoglobin (Mb)

Myoglobin has enjoyed a central position as a model system to investigate the ligand binding dynamics to the heme. This well described O_2-storage protein is capable of releasing O_2 during periods of hypoxia or anoxia in muscle. The protein consists of 8 α-helices which fold to make a compact globular stucture. The heme group sits in its hydrophobic pocket where only polar residues are two histidines (His^{93} and His^{64}), one on either side of the heme. The heme consits of four heterocyclic pyrole rings, that share a common plane and bind a iron ion (Fe^{II}) in the centre. The iron ion is capable to interact with six ligands, four of which are provided by the nitrogen atoms of the four pyroles. The imidazole side chain of proximal histidine (His^{93}) provides the fifth ligand, stabilizing the heme group and slightly displacing the iron ion away from the plane of the heme (0.4 Å from the heme plane). The out-of plane iron displacement is known as heme "doming" (**Fig. 7**). The sixth ligand position, unoccupied in deoxymyoglobin, serves as the binding site for oxygen, as well as for other potential ligands such as CO or NO. Binding of this small molecule changes the spin state of the Fe-porphyrin group (from high spin to low spin state) and pulls the iron atom into the porphyrin ring plane, triggering conformational change of the protein. This displacement provides the basis for the conformational changes that underlie the allosteric properties of tetrameric hemoglobin (Perutz, 1989).

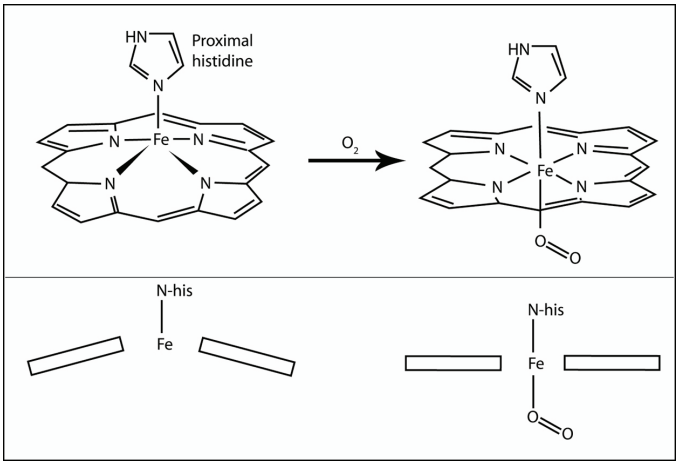

Fig. 7. The "deoxy" and "oxy"configuration of the heme conformation in myoglobin

FixL heme domain (FixLH)

In contrast to globular structure the classical heme proteins, the overall structural fold of the **FixL heme domain** is dominated by an antiparallel β-sheet and can be described as a left hand that encloses a heme. The β-strands A, B, H and I form the fingers, the β-strand G the thumb, and helices E and F the palm (**Fig. 8**). The hand clasps the heme group, which is covalently linked to His^{200} in helix F. The heme pocket is highly hydrophobic due to presence of a triad Ile^{215}/Leu^{236}/Ile^{238}, whose side chains point toward the heme.

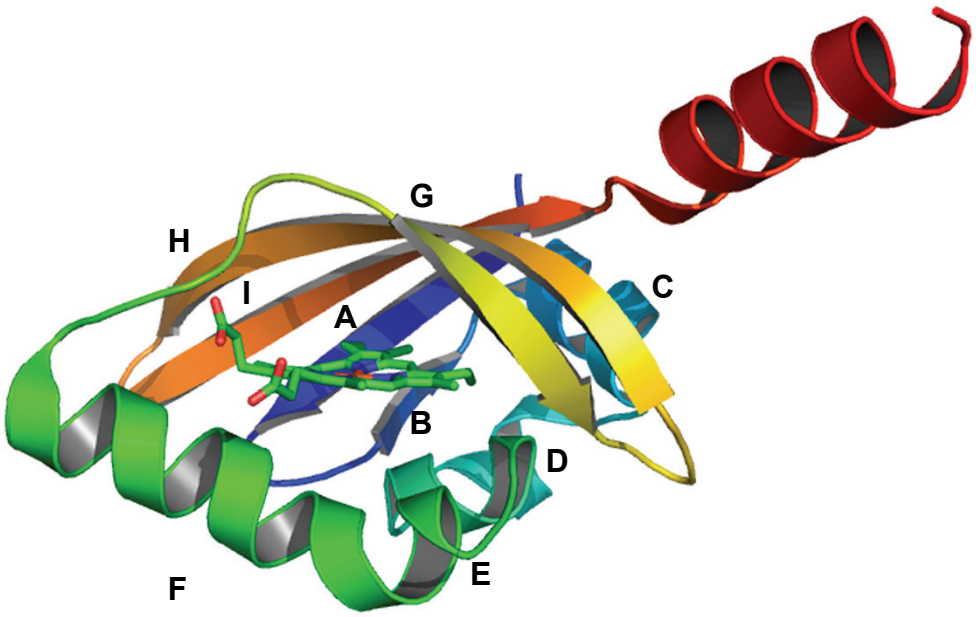

Fig. 8. Crystallographic structure of the heme domain in FixL of *Bradyrhizobium japonicum* (PDF entry 1LSW). The colouring goes from blue at the amino to red at the carboxyl end. The figure was prepared using PYMOL.

Similar to myoglobin, the FixLH has a *b*-type heme as prosthetic group with a proximal histidine (His^{200}) as axial ligand. The deoxy-Fe^{II} is high spin and 5-coordinated. Upon oxygen fixation, the heme becomes low spin 6-coordinated (Gilles-Gonzalez *et al.*, 1995). In myoglobin, this spin transition is accompanied by movement of the proximal histidine ligand of the heme; in case of tetrameric hemoglobin consequently transduced into allosteric changes. Unlike myoglobin, structural crystallographic studies of FixLH demostrated that, the axial histidine ligand does not move between the 'on' (high-spin) and 'off' (low-spin) conformations (Gong *et al.*, 1998; Gong *et al.*, 2000; Hao *et al.*, 2002). Instead, the conformation change of the heme pocket upon ligand binding was observed especially in the region between the Fα-helix and the

Gβ-strand, termed the **FG loop** (position Thr^{209} to Arg^{220} in *Bradyrhizobium japonicum*). The spatial position of one arginine (Arg^{220}) moves from the propionate 7 group of the heme toward the oxygen ligand. It was proposed that flattening of the heme plane upon ligand binding leads to a shift in the position of the heme propionates reducing the strength of the salt bridge between Arg^{220} and propionate 7 (Gong *et al.*, 2000). As a result, the arginine at position 220 shifts into the distal pocket of FixL and H-bond is formed with the terminal oxygen atom of heme-bound O_2 (**Fig. 9**).

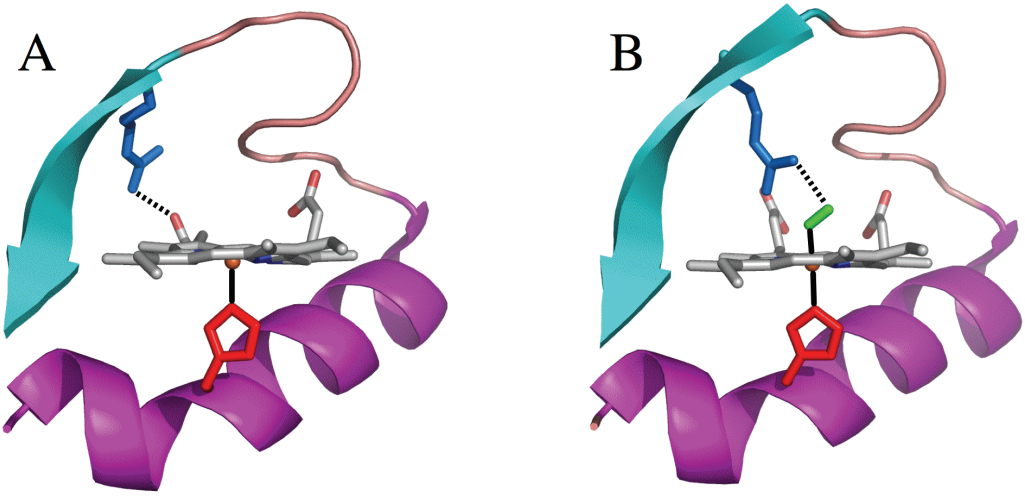

Fig. 9. Crystallographic structures: **(a)** ferrous deoxy complex (PDF entry 1LSW; Hao *et al.*, 2002) and **(b)** oxycomplex (PDF entry 1DP6; Gong *et al.*, 2000) of FixLH. The residues shown are the proximal histidine (H200, red), and the distal argine (R220, blue). The latter forms a salt bridge with heme propionate 7 in the deoxy form and a hydrogen bond with O_2 (green) in the oxycomplex. The figure was prepared using PYMOL.

Highly conserved Arg220 in FixLH

Movement of the Arg^{220} hase been proposed to induce the conformation change of the the FG loop, consequently resulting in the inactivation of the kinase domain of FixL protein (Dunham *et al.*, 2003; Gong *et al.*, 2000; Hao *et al.*, 2002). Arg^{220} is highly conserved residue among all known FixL proteins. This residue has been proposed to play a crucial role in ligand discrimination and in the signal transmission pathway, because of its striking rearrangement between the oxy- and deoxy complex. Moreover, crystalographic structure of R220A mutant of *Bj*FixL demostrated the structural modification involved the heme planarity and the propianate 7 geometry (Dunham *et al.*, 2003). Indeed, the histidine kinase activity was strongly affected by this mutation.

The steady-state absorption spectra of the oxy- and deoxy complexes of FixLH closely resemble those of myoglobin (Gilles-Gonzalez *et al.*, 1991), and also heme geometry in the both oxycomplexes are very similar (Gong *et al.*, 2000). However, the ligand-binding properties are very different (Perutz *et al.*, 1999). In particular the oxygen affinity and oxygen-binding rates are much lower in FixL (Gilles-Gonzalez *et al.*, 1994), possibly adapting its sensitivity to changes in the environmental oxygen concentration.

Whereas crystallographic structures of FixLH in unliganded and liganded states (Gong *et al.*, 1998; Gong *et al.*, 2000, Hao *et al.*, 2002; Key and Moffat, 2005) provided models for the starting and end points of signal transmission within the heme domain, intermediates of the transmission patway can in principle be experimentally elucidated using time-resolved studies, initiated by ligand binding or dissociation. In FixL protein, dissociation of the heme-bound ligand is the first step in signaling toward the enzymatic domain. The fact that in FixL the signaling molecule bound to a heme cofactor can be exploited (as diatomic ligands can be photodissociated using a short light pulse) allows subsequent observation of very short-lived intermediates in the liganded-unliganded pathway. The kinetics of ligand binding and the spectral characteristics of intermediates have been studied using various spectroscopic techniques for CO, NO and O_2 as ligands (Gilles-Gonzalez *et al.*, 1994; Rodgers *et al.*, 1999; Liebl *et al.*, 2002; Miksovska *et al.*, 2005). The study on ultafast ligand dynamics using femtosecond spectroscopy resulted in two main findings (Liebl *et al.*, 2002):

- In FixLH, oxygen geminate recombination after O_2 dissociation occurs with unprecedented efficiency for heme proteins: extremely fast (in ~5 ps) and high yield (~90%). Thus, despite the much lower oxygen affinity comparing to myoglobin, the heme pocket of FixLH appears to act as efficient oxygen trap.

- In contrast to myoglobin, the transient spectrum of the ligand-photodissociated heme in FixLH is modified with respect to the steady-state unliganded heme, especially for physiological ligand O_2. Spectrum of the transient photoproduct of oxy FixLH is close to that of the oxy-complex, whereas in myoglobin it is very similar to that of the deoxy-complex. The spectral difference, implies that the heme and its environment have not relaxed yet to the equilibrium unliganded conformation. This observation suggests a specific role of the protein environment in the transiently formed initial signaling intermediate.

Four site-directed mutants of FixLH from *Bradyrhizobium japonicum* at position 220 were constructed to study the influence of Arg^{220} on the heme conformation and on the nature of O_2 binding. The point mutations were chosen in order to modify the electrostatic properties and H-bonding capabilities of residue 220, with minor steric modifications of the side chain. Subsequently, mutant forms were characterized by their steady-state resonance Raman spectra and bimolecular ligand interaction properties (Balland *et al.*, 2005; Balland *et al.*, 2006). In addition, we constructed FixLH/R220A mutant in which steric modifications are quite substantial. For all mutant proteins, the overall oxygen affinity was found to be modified and the ligand binding and dissociation rates were found to be substantially increased (Balland *et al.*, 2005).

5.2. FixL: Aims of experiments

1. To characterize the influence of the mutations at position 220 in the FixLH from *Bradyrhizobium japonicum* (*Bj*FixLH)on the initial signaling dynamics, in terms of transient spectra of the heme after ligand dissociation, using femtosecond spectroscopy.

→ *Jasaitis et al., 2006*

2. To characterize the heme configuration of the transient primary photoproduct in the *Bj*FixLH wild type and R220X mutant forms by time-resolved resonance Raman spectroscopy.

→ *Kruglik et. al., 2007*

5.3. FixL: Results and discussion

*Jasaitis A., **Hola K.**, Bouzhir-Sima L., Lambry J.C., Balland V., Vos M.H., Liebl U. (2006) Role of distal arginine in early sensing intermediates in the heme domain of the oxygen sensor FixL. Biochemistry 45: 6018-6026*

To assess the role of Arg^{220} in the dynamics and properties of the initial intermediates in ligand release, we have investigated the effects of R220X (X = I, Q, E, H, or A) mutations in the FixLH domain on spectral properties of the heme upon photolysis of O_2, NO, and CO using femtosecond transient absorption spectroscopy.

We show for substitutions of arginine at position 220, that the spectral perturbations with respect to myoglobin transient spectrum reported previously (described on page 26; Liebl *et al.*, 2002) **are strongly diminished compared to the wild type spectrum and that the heme adopts a "deoxy-like" configuration within a few**

picoseconds of ligand dissociation. This implies that Arg^{220} is strongly involved in the postdissociation transmission steps involving the heme.

The decrese of the spectral pertrubation due to substitution of Arg^{220} was slightly observed also in case of of the NO- and CO-dissociated state, in spite of the fact that Arg^{220} does not appear to interact directly with heme-bound NO and CO (Gong *et al.*, 2000). This finding suggests that this residue is also involved in the transmission of the NO and CO signals that also modulate FixL activity, even to a far lesser extent.

Moreover, we observed decreased O_2 rebinding efficiency (as well as NO, CO) in the mutants comparing to the wild type, especially in case of FixLH-R220A. We proposed that the efficient O_2 rebinding kinetics in the wild type heme domain is due to volume reduction of the distal heme pocket by Arg^{220}. In view of the marked difference in the size of alanine and arginine, this indicates that the volume of the heme pocket plays an important role in the kinetics of ligand rebinding. Indeed, the order of efficiency decreasing of rebinding correlates as well with a decrease in residue size (histidine > glutamine > alanine).

As reported previously (Liebl *et al.*, 2002), only ~10% O_2 escape form the heme pocket after dissociation in the case of FixLH wild type. **We demostrated that the yield of dissociated oxygen after recombination is increased in all mutants, up to almost unity in FixLH-R220A, indicating a key role of Arg^{220} in caging the oxygen near the heme.**

Molecular dynamics simulations corroborated these findings. In wild type FixLH, after deletion of the heme iron-oxygen bond, in most simulations the oxygen molecule does not move away but fluctuates in the vicinity of its initial position close to the heme iron. In all trajectories of the mutants, oxygen exhibits a higher mobility and moves farther from the heme.

Moreover, in a few of the simulations in which O_2 does move out of the heme pocket (50 ps time scale) R220 remains hydrogen bonded with O_2 and also sweeps away from the heme. This suggests a correlated movement of R220 and O_2 that might be essential for initiation of the signal transduction.

As FixL heme domain acts as a "bistable switch", which allows a fraction of dissociated oxygen to bring about the ensemble of further intermediates in the sensing process, Arg^{220} was proposed to be a pivoting element of this picosecond switch.

*Kruglik S.G., Jasaitis A., **Dvorakova-Hola K.**, Liebl U., Martin J.L., Vos M. H.: Subpicosecond Oxygen Trapping in the Heme Pocket of the Oxygen Sensor FixL Observed by Time-Resolved Resonance Raman Spectroscopy. Proc. Natl. Acad. Sci. U S A 104: 7408-7413*

Time-resolved resonance Raman spectroscopy (TR^3) spectroscopy is a powerful technique to study transient heme structures after ligand photodissociation and has been successfully used to characterize photodissociated heme-CO, heme-NO, and heme-methionine complexes in various six-coordinated heme proteins on the picosecond time scale (Petrich *et al.*, 1987; Franzen *et al.*, 1994; Mizutani *et al.*, 1997; Mizutani *et al.*, 2001; Cianetti *et al.*, 2004; Négrerie *et al.*, 2006). These studies clearly indicate that deoxy-like heme conformation, with heme iron out-of-plane motion (doming), appears within 1 ps after photoexcitation. However, such studies have not been reported yet for the physiologically relevant ligand oxygen, even for the well studied oxygen carriers myoglobin and hemoglobin.

Here, we have succeeded in monitoring TR^3 spectra of oxy-FixLH and oxymyoglobin, within the very first moments after photoexcitation, with an instrument response of ~0.7 ps. Importantly, for MbO_2, our results clearly demonstrate that upon dissociation of the physiological ligand O_2, a five-coordinate deoxy complex is formed faster than 0.5 ps, in a very similar way as previously shown for dissociation of the inhibitor molecule CO from hemoglobin and myoglobin.

In case of FixLH, we surprisingly discovered that, the deoxy-like structure with heme doming does not appear upon dissociation of the oxygen ligand, even at time delays as short as 0.5 ps. Based on this striking finding we propose that the rebinding of dissociated O_2 is mostly completed at 0.5 ps, much faster than previously proposed (in ~5 ps; Liebl *et al.*, 2002). A ~5 ps decay component associated with a spectrum, that is red-shifted with respect to the ground state oxy-spectrum, but far less than expected for a steady-state deoxy spectrum, most likely reflects cooling of a hot 6-coordinate heme (Ye *et al.*, 2002; Négrerie *et al.*, 2006,).

Consequently, we suggest that ultrafast oxygen rebinding to the heme on the femtosecond time scale hinders heme doming. In this view, the dissociated oxygen molecule cannot move away substantially from the heme, due to steric constraints, and remains in a favorable position for rebinding. This rebinding may then occur prior to, or in competition with, heme doming, which is expected to take ~200 fs (Li *et al.*, 1992; Klug *et al.,* 2002).

To study the origin of this finding in more detail, and in particular the role of arginine 220, we compared the behavior of the wild type with that of the R220H and R220Q mutant proteins.

Comparison of TR^3 results on wild type and mutant FixLH indicates that a hydrogen bond between the residue at position 220 (arginine in wild type) and the terminal oxygen atom of the oxygen molecule (present in wild type and FixLH-R220H) plays an essential role in the steric constraints maintaining the oxygen molecule in place.

In FixLH-R220H-O_2, substituted histidine is hydrogen bonded to the terminal oxygen atom (Balland *et al.*, 2005) as well as arginine 220 in wild type FixLH-O_2. In FixLH-R220Q-O_2, however, glutamine makes a hydrogen bond with the iron-bound oxygen atom, but not with the terminal oxygen atom (Balland *et al*., 2005).

While the heme conformation in the FixLH-R220H mutant protein is similar to that of wild type FixLH, we observed the domed heme in the case of FixLH-R220Q-O_2. This correlates well with the relatively high yield of long-lived deoxy heme in FixLH-R220Q previously observed (Jasaitis *et al*., 2006), and indicates that in wild type FixLH, it is via constraints on the terminal oxygen atom that O_2 is 'pushed' towards the heme.

Thus, we propose that the rigid O_2 configuration imposed by arginine 220, in combination with the hydrophobic and constrained properties of the distal cavity, keep dissociated oxygen in place. These results uncover the origin of the 'oxygen cage' properties of oxygen sensor FixL protein.

6. METHODS

6.1. Molecular biology approach:

site-directed mutagenesis, gene expression, protein purification

Strategy for substitutions of tryptophan residues in the yeast α-MPP (Trp^{147}, Trp^{223}, Trp^{481})

For all tree tryptophan residues substitutions, a PCR-based strategy was employed, taking advantage of unique restriction site in the DNA regions where mutation where planed. Synthetic mutation primers were designed to carry the intended mutation together with a unique restriction site at its 5` end. DNA amplification was then performed between the mutant primer and a second primer at another restriction site several tens or hundred nucleotides away. PCR product was used as a primer for the second PCR step with the third primer, which is closed to second unique site. The yielded fragment was digested by two enzymes cleaving the unique restriction sites and ligated into the construct carrying gene for the yeast α-MPP by replacing corresponding region.

Strategy for site-directed mutagenesis within the GRL of the yeast α-MPP

The point mutations were introduced into the GRL using a cassette allowing the replacement of 74 bp in the sequence coding for the whole GRL and its close neighborhood. This cassette was derived from sequence coding for the yeast α-MPP and was prepared by an analogous PCR-based strategy described above. The “key-stones” of this cassette were two restriction sites (*Pst* I, *Aat* II) bordering the sequence coding for the glycine-rich region. The fragment carrying the requested mutation was formed by hybridization of two synthetic complementary oligonucleotides. Subsequently, the fragment was ligated into the cassette using their cohesive ends (*Ps t*I/*Aat* II). Finally, the cassette replaced corresponding region in the construct carrying gene for the yeast α-MPP.

Strategy for substitutions of Arg220 in *Bj*FixLH

The substitutions of Arg220 were carried out using QuikChange® site-directed mutagenesis kit (Stratagene), based on PCR and subsequent treatment by *Dpn* I endonuclease. DNA polymerase replicates both strands (in the whole length) of dsDNA vector carrying sequence coding for using *Bj*FixLH. Thus, the synthetic primers containing the desired

mutation are incorporated and the mutated plasmid is yielded. The *Dpn* I is specific for methylated and hemimethylated DNA and is used to digest parental DNA template to select for mutation-containing synthesized DNA (DNA isolated from a *dam*$^+$ *E.coli* strain is dam methylated and therefore susceptible to *Dpn* I digestion).

Gene expression and protein purification

All used proteins (yeast α-MPP and its mutant forms, yeast β-MPP, yeast pMDH, *Bj*FixLH and its mutant forms) were produced as heterologous proteins in *E. coli* BL21(DE3) using the pET expression system (Novagen). In case of α-MPP forms, *E. coli* cells were co-transformed by pGroESL (DuPont; vector bearing sequences coding for GroES and GroEL proteins) and co-expressing α-MPP with chaperonins.

The α-MPP forms as well as FixLH forms were produced with His-tag on the the N-terminus; a Hi-Trap Chelating column (Amersham) was used to purify the α-MPP, Ni-NTA Spin kit (Qiagen) for purification of FixLH forms.

Insoluble β-MPP as well as yeast pMDH precursors (used as MPP substrate) were isolated from inclusion bodies. The renaturation procedure was used as described by Géli *et al*.(1990).

6.2. Basic principles of used spectroscopic methods

Tryptophan fluorescence spectroscopy

Fluorescence spectroscopy is widely used spectroscopic techniques in the study of changes in the structural and dynamic properties of biomolecules and biomolecular complexes. As fluorescence typically occurs from aromatic molecules, a tryptophan residue is an intrinsic fluorophore and is widely employed as a “native fluorescence probe” to study protein conformation changes.

Fluorescence spectroscopy involves using a beam of light, usually ultraviolet light, that excites the electrons in molecules of certain compounds and causes them to emit light of a lower energy (indole group of tryptophan absorbs near 280 nm, and emits near 340 nm). By analysing the different frequencies of light emitted in fluorescent spectroscopy, the structure of different vibrational levels can be determined. Like most biophysical techniques, fluorescence

spectroscopic studies can be carried out at many levels ranging from simple measurement of steady-state emission intensity to quite sophisticated time-resolved studies.

Steady-state fluorescence measurement

Steady-state measurements are performed with constant ilumination and observation. The sample is iluminated with continuous beam of light, and the intensity or the emission spectrum is recorded. The steady-state spectra can provide information concerning protein folding and/or protein-protein interaction, expressed as the change of the tryptophan environment, where two main paramenters are observed: the intensity of fluorescence and a shift of spectral maximum.

In the unfolded state most tryptophan residues in proteins have spectra similar to that of N-acetyl tryptophanamide in water, with a maximum of ~355 nm. The average energy of the emission of the tryptophan residues usually shifts to the red on unfolding because the solvent exposure. However, the shift can also reflect the contact with a polar amino acid residue during protein-protein interaction. On the other hand, the emission of tryptophan may be blue shift if the group is buried within a protein.

The second observable parametr is decrease of the fluorescence intensity, termed fluorescence quenching that can occurs by different mechanisms. Collisional quenching occurs when the extited-state fluorophore is deactivated upon contact with some other molecule. The quenching of fluorescence emission of tryptophan in proteins arises from excited state encounters of the tryptophan with the functional groups of the amino acids in the surrounding protein matrix or from encounters with the solvent.

Time-resolved fluorescence measurement

Time-resolved fluorescence spectroscopy is an extension of fluorescence spectroscopy, where the fluorescence of a sample is monitored as a function of time. The main characteristics is the fluorescence lifetime that is defined by the time the molecule spends in the excited state prior to return to the ground state (generally ns timescale). The lifetime is influenced by an interaction of the fluorophore with its environment. For the measurements the sample is exposed to a pulse of light, where the pulse width is typically shorter than the decay time of the sample. To simply describe the relationship between steady-state and time-resolved fluorescence measurement, the steady-state observation is an average of the time-resolved phenomena over intensity decay of the sample. Indeed, much of the molecular information

available from the fluorescence is lost during the time averaging process. For example, the time-resolved measurements are used for measuring anisotropy decays that are frequently more complex than a single exponetial.

Anisotropy measurements provide information on the size and shape of protein and its flexibility and are used to measure protein-protein associations, fluidity of membranes, and for immunoassays of numerous substances. Anisotropy measurements are based on the principle of photoselective excitation of fluorophores by polarized light. These experiments reveal the angular displacement of the fluorophore that occurs between absorption and subsequent emission of a photon. This angular displacement is dependent upon the rate and extent of rotational diffussion during the lifetime of the excited state. Thus, the dependence of fluorescence anisotropy upon fluorophore motion has resulted in application of this technique in the study of protein dynamics.

Definition of fluorescence anisotropy

Upon excitation with polarized light the emission form sample is also polarized. The extent of polarization of the emission is described in terms of the anisotropy *(the anisotropy is defined as the ratio of the difference between the emission intensity parallel to the polarization of the electric vector of the exciting light and that perpendicular to that vector divided by the total intensity)*. The origin of the anisotropy is the existence of transition moments for absorption and emission that lie along specific direction within the fluorophore structure. In homogenous solution the ground-state fluorophores are all randomly oriented. When exposed to polarized light, those fluorophores that have their absorption moments oriented along the electric vector of the incident light are preferentially extited. Hence, the excited population is partially oriented.

Fluorescence resonance energy transfer

Resonance energy transfer is a very powerful method for obtaining both structural and dynamic information about protein and protein complexes. It is based on the throughspace dipolar transfer of excited-state energy to acceptor molecules up to 75 Å removed from the donor fluorophore. The efficiency of transfer is determined by the distance between the donor and acceptor molecules. Other factors such as the overlap of the emission of the donor with the absorption of the acceptor and the relative orientations of the two chromophores also affect the energy transfer. Thus, distances, distance distributions, and even dynamic distance distributions can be obtained using this approach.

Femtosecond transient absorption spectroscopy

Absorption spectroscopy refers to a wide range of techniques where one measures how much light of a particular wavelength (color) is absorbed by a sample. Since wavelength can often be correlated with the presence and or structure of a particular chemical, absorption spectroscopy is widely used for both qualitative and quantitative evaluation of a sample. In case of heme proteins the measurement is targetted to the the blue wavelength region of the visible spectrum (400-450 nm), corresponding to a wavelength of maximum absorption. This maximum, termed "Soret band" is used to describe the absorption of heme-containing moieties. Ultrafast transient absorption spectroscopy is an extension of absorption spectroscopy, where the absorbance at a particular wavelength or range of wavelengths of a sample is measured as a function of time after excitation by a flash of light. The measurement is conducted on extremely short time scales, in case of femtosecond spectroscopy on 10^{-15} seconds.

The most common form of femtosecond spectroscopy is a “pump-probe” method that needs two light pulses on femtosecond time scale. The first one, the so-called “pump” pulse, perturbs the sample at time t=0. The second one, the so-called “probe” pulse, delayed with respect to the pump pulse, crosses the perturbed sample and act as a probe, investigating the changes produced by the pump.

Because femtosecond pulses are shorter than the time scale for most intramolecular motions, the technique provides tools for directly studying the dynamics and relaxation of molecular vibrations. More importantly, this technology allows the "real-time"observation of photochemical reactions as they proceed on the initially populated excited-state potential surfaces, which are the "transition states" of photoinitiated reactions. Thus, intermediate products, that differ from the starting and end products, can be observed.

Resonance Raman spectroscopy

Raman spectroscopy involves analyzing the scattered photons from a laser beam focused into the sample solution and provides information on molecular vibrations that, in turn, yield data on molecular structure and confirmation.

When light is scattered from a molecule most photons are elastically scattered. The scattered photons have the same energy (frequency) and, therefore, wavelength, as the incident photons. However, a small fraction of light (approximately 1 in 10^7 photons) is scattered at optical frequencies different from, and usually lower than the frequency of the

incident photons — Raman effect. The difference in energy between the incident photon and the Raman scattered photon is equal to the energy of a vibration of the scattering molecule. Thus, it is the shift in wavelength of the inelastically scattered radiation that provides the chemical and structural information.

Typically, a sample is illuminated with a laser beam. Light from the illuminated spot is collected with a lens and sent through a monochromator. Wavelengths close to the laser line (due to elastic scattering) are filtered out and those in a certain spectral window away from the laser line are dispersed onto a detector.

Raman scattering is a relatively weak process. However, there are several techniques which can be used to enhance the sensitivity of a Raman measurement; one of these is resonance Raman spectroscopy. If the wavelength of the exciting laser coincides with an electronic absorption of a molecule, the intensity of Raman active vibrations are enhanced by a factor of 10^2 to 10^4. This resonance enhancement (resonance Raman effect) can be extremely useful, not just in significantly lowering the detection limits, but also in introducing electronic selectivety. These features make the resonance Raman technique useful for providing both structural and electronic insight into species of interest.

Metalloporphyrins (heme), carotenoids and several other classes of biologically important molecules have strongly allowed electronic transitions in the visible region, making them ideal candidates for resonance Raman spectroscopy. The spectrum of the chromophoric moiety is resonance enhanced and that of the surrounding environment is not. This means that absorbing chromophores in active centres of proteins can be specifically probed by visible excitation wavelengths, and not the surrounding protein matrix, which would require UV lasers to bring into resonance. Thus, resonance Raman spectroscopy and its time-resolved variant (TR3) are powerful techniques for documenting structures of heme protein stable forms and also allow characterization of intermediate structures.

7. PROTEINS INVOLVED IN BIOSYNTHESIS OF SECONDARY METABOLITES

Regarding the dissimilarity of the subject fields, two additional papers are not included in the compilation of the studies concerning the protein structural dynamics. These papers describe the results of the other project of our laboratory – an analysis of the lincomycin biosyntetic pathway.

Lincomycin, produced by *Streptomyces lincolnensis*, and its derivatives are important clinically used antibiotics. In lincomycin biosynthesis, two building units, the aglycone propyl-L-proline from L-tyrosine and the sugar moiety – methyl-thiolincosamide, are first synthesized separately. Their condensation to N-demethyllincomycin (NDL) is catalyzed by NDL synthetase, the central enzyme of the biosynthetic pathway. The final biosynthetic step is N-methylation of the propylproline moiety of NDL.

The lincomycin-production gene cluster of the industrial overproduction strain *Streptomyces lincolnensis* 78-11 has been sequenced (Peschke *et al*., 1995) and twenty-seven putative open reading frames with biosynthetic or regulatory functions (*lmb* genes) identified. Two distinct hypothetical genes, *lmbI* and *lmbH*, were postulated downstream of the *lmbJ* gene, coding for LmbJ protein, which catalyses the last lincomycin biosynthetic step, i.e. conversion of NDL to lincomycin.

Aims of experiments:

1. To investigate the existence of a single open reading frame *lmbIH* in *S. lincolnensis* ATCC25466, instead of two smaller *lmbI* and *lmbH* previously shown in *S. lincolnensis* 78-11 (Peschke *et al*., 1995).

→ *Janata et al., 2001*

2. To demonstrate the expression of two overlapping genes *lmbJ* and *lmbIH* in high-production type *S. lincolnensis* JK-17-45 and to document LmbJ and LmbIH protein levels during the lincomycin production phase.

→ *Hola et. al., 2007*

*Janata J., Najmanova L., Novotna J., **Hola K.**, Felsberg J. and Spizek J. (2001) Putative lmbI and lmbH genes form a single lmbIH ORF in Streptomyces lincolnensis type strain ATCC25466. Antonie Van Leeuwenhoek 79: 277–284.*

The presence of a single open reading frame *lmbIH* in the lincomycin *Streptomyces lincolnensis* ATCC 25466 was demonstrated. The protein product LmbIH is homologous with the TldD protein family. Moreover, our experiments indicate co-regulation of *lmbJ* and *lmbIH* expression. This translation coupling probably reflects an eight nucleotide overlap between the *lmbJ* and *lmbIH* genes, as well as the lack of a Shine-Dalgarno sequence upstream of the *lmbIH* gene.

***Hola K**., Janata J, Kopecky J. and Spizek J. (2003) LmbJ and LmbIH protein levels correlate with lincomycin production in Streptomyces lincolnensis. Lett. Appl. Microbiol. 37: 470-474.*

In *Streptomyces lincolnensis* JK-17-45, protein LmbJ was detected at stable level for a long time period covering the whole lincomycin production phase. This fairly corresponds to the catalytic function of the protein in the antibiotic biosynthesis (NDL methyltransferase). On the contrary, the absence of LmbIH protein at a detectable level during the major part of the antibiotic production phase casts doubt on its possible catalytic function. Rather a different connection with the final biosynthetic steps, e.g. regulatory, can be envisaged. However, the important goal of this study is the confirmed expression of a newly found *lmbIH* putative regulatory gene *in vivo*, during lincomycin production.

8. REFERENCES

- Abe Y., Shodai T., Muto T., Mihara K., Torii H., Nishikawa S., Endo T., Kohda D. (2000) *Cell* 100: 551-560
- Agron P.G., Ditta G.S., Helinski D.R. (1993) *Proc. Natl. Acad. Sci.USA* 90: 3506-3509
- Balland V., Bouzhir-Sima L., Anxolabéhère-Mallart E., Boussac A., Vos M. H., Liebl U., and Mattioli T. A. (2006) *Biochemistry* 45: 2072-2084
- Balland V., Bouzhir-Sima L., Kiger L., Marden M. C., Vos M. H., Liebl U. and Mattioli T. A. (2005) *J. Biol. Chem.* 280: 15279-15288
- Bauer M.F., Hofmann S., Neupert W. and Brunner M. (2000) *Trends Cell Biol* 10:25-31
- Becker A.B. and Roth R.A. (1992) *Proc. Natl. Acad. Sci. USA*. 89: 3835-3839
- Brix J., Dietmeier K. and Pfanner N. (1997) *J. Biol. Chem.* 272: 20730–20735
- Chupin V., Leenhouts J. M., de Kroon A. I. and de Kruijff B (1996) *Biochemistry* 35: 3141-3146
- Cianetti S., Négrerie M., Vos M. H., Martin J.-L. and Kruglik S. G. (2004) *J. Am. Chem. Soc.* 126**,** 13932 - 13933
- De Jongh, H. H. (2000) *Eur J Biochem* 267: 5796-5804
- Dunham C. M., Dioum E. M., Tuckerman J.R., Gonzalez G., Scott W. G. and Gilles-Gonzalez M.A. (2003) *Biochemistry* 42: 7701-7708
- Franzen S., Lambry J.-C., Bohn B., Poyart C. and Martin J.-L. (1994) *Nature Struct. Biol.* 1: 230 – 233
- Gakh O, Cavadini P. and Isaya G (2002) *Biochim Biophys Acta* 1592: 63-77
- Gakh O., Obsil T., Adamec J., Spizek J., Amler E., Janata J. and Kalousek F. (2001) *Arch. Biochem. Biophys.* 385: 392-396
- Garbers D.L and Lowe D.G. (1994) *J. Biol. Chem.* 269: 30741-30744
- Gavel Y. And von Heijne G. (1990) *Protein Eng.* 4: 33-37
- Geli V., Yang M.J., Suda K., Lustig A., Schatz G. (1990) *J. Biol. Chem.* 265: 19216-19222
- Gilles-Gonzalez M.A., Ditta G. and Helinski D.R. (1991) *Nature*; 350: 170-172
- Gilles-Gonzalez M.A., Gonzalez G., Perutz, M., Kiger L., Marden M. C. and Poyart C. (1994) *Biochemistry* 33: 8067-8073

- Glick B.S., Beasley E.M. and Schatz G. (1992) *Trends. Biochem. Sci. Nov.* 17: 453-459
- Gong W., Hao B. and Chan M.K. (2000) *Biochemistry* 39: 3955–3962
- Gong W., Hao B., Mansy S.S., Gonzales G., Gilles-Gonzales M.A., Chan M.K. (1998) *Proc. Natl. Acad. Sci. USA.* 95: 15177–15182
- Hachiya N., Mihara K., Suda K., Horst M., Schatz G. and Lithgow, T. (1995) *Nature* 376: 705–709
- Hammen P. K., Gorenstein D. G. and Weiner H. (1994) *Biochemistry* 33: 8610-8617
- Hao B., Isaza C., Arndt J., Soltis M., and Chan M. K. (2002) *Biochemistry* 41**:** 12952-12958
- Hartl F.U. and Neupert W. (1990) *Science* 247: 930-938
- Hendrick J.P., Hodges P.E. and Rosenberg L.E. (1989) *Proc. Natl. Acad. Sci. U S A.* 86: 4056-4060.
- Hines V. and Schatz G. (1993) *J. Biol. Chem.* 268: 449–454
- Ito A. (1999) *Biochem Biophys Res Commu,* 265: 611-616
- Key J. and Moffat K. (2005) *Biochemistry* 44: 4627-4635
- Klug D. D., Zgierski M. Z., Tse J. S., Liu Z., Kincaid J. R., Czarnecki K. and Hemley R. J. (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99: 12526-12530
- Kojima K., Kitada S., Ogishima T. and Ito A. (2001) *J. Biol. Chem.* 276: 2115-2121
- Kojima K., Kitada S., Shimokata K., Ogishima T. and Ito A. (1998) *J. Biol. Chem.* 273: 32542-32546
- Komiya T., Rospert S., Schatz G. and Mihara, K. (1997) *EMBO J.*, 16: 4267–4275
- Li X.-Y. and Zgierski M. Z. (1992) *Chem. Phys. Lett.* 188: 16-20
- Lakowicz J. R. (2006): Principles of Fluorescence Spectroscopy, Third Edition, Springer Science+Business Media, LLC
- Liebl U., Bouzhir-Sima L., Negrerie M., Martin J.-L. and Vos M. H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99: 12771-12776
- Luciano P., Geoffroy S., Brandt A., Hernandez J. F. and Geli V. (1997) *J. Mol. Biol.* 272: 213-225
- Miksovska J., Suquet C., Satterlee J. D. and Larsen R. W. (2005) *Biochemistry* 44: 10028-10036
- Mizutani Y. and Kitagawa, T. (1997) *Science* 278: 443-446
- Mizutani Y. and Kitagawa T. (2001) *J. Phys. Chem. B.* 105: 10992-10999

- Nagao Y., Kitada S., Kojima K., Toh H., Kuhara S., Ogishima T. and Ito A. (2000) *J. Biol. Chem.* 275: 34552-34556
- Négrerie M., Cianetti S., Vos M. H., Martin J.-L. and Kruglik S. G. (2006) *J. Phys. Chem. B.* 110: 12766 -12781
- Neupert W. (1997) *Annu. Rev.Biochem.* 66: 863–917
- Niidome T., Kitada S., Shimokata K., Ogishima T. and Ito A. (1994) *J. Bio.l Chem.* 269: 24719-24722
- Perutz M.F. (1989) *Mechanisms of Cooperativity and Allosteric Regulation in Proteins.* Cambridge University Press: Cambridge
- Perutz M.F., Paoli M. and Lesk A.M. (1999) *Chem. Biol.* 6: 291-297
- Perry A.J., Hulett J.M., Likic V.A., Lithgow T., Gooley P.R. (2006) *Curr Biol*. 16:221-229
- Peschke U., Schmidt H., Zhang H.Z. and Piepersberg W. (1995) *Mol. Microbiol.* 16: 1137–1156.
- Petrich J. W., Martin J.-L., Houde D., Poyart C. and Orszag A. (1987) *Biochemistry* 26**,** 7914-7923
- Pfanner N. and Neupert W. (1990) *Annu Rev.Biochem* 59: 331-353
- Pfanner N. and Geissler A. (2001) *Nat Rev Mol Cell Biol* 2: 339-349
- Pfanner N. and Wiedemann N. (2002) *Curr Opin Cell Biol,* 14: 400-11
- Pfanner N., Craig E.A. and Hönlinger A. (1997) *Annu. Rev. Cell Dev. Biol.* 13: 25–51
- Ramage L., Junne T., Hahne K., Lithgow T. and Schatz G. (1993) *EMBO J.*, 12: 4115–4123
- Reyrat J.M., David M., Blonski C., Boistard P., Batut J. (1993) *J.Bacteriol.*175: 6867-6870
- Rodgers K. R., Lukat-Rodgers G. S., and Tang L. (1999) *J. Am. Chem. Soc.* 121: 11241-11242
- Roise D., Horvath S.J., Tomich J.M., Richards J.H. and Schatz (1986) *EMBO J,* 5: 1327–1334
- Roise D., Theiler F., Horvath S.J., Tomich J.M., Richards J.H., Allison D.S. and Schatz G. (1988) *EMBO J.* 7: 649–653
- Shelver D., Kerby R.L., He Y. and Roberts G.P. (1997) *Proc. Natl. Acad. Sci. USA*; 94: 11216-11219
- Shimokata K., Kitada S., Ogishima T., and Ito A. (1998) *J. Biol. Chem.* 273: 25158-

2516

- Schatz G. and Dobberstein B. (1996) *Science* 271: 1519-1526
- Song M.C., Shimokata K., Kitada S., Ogishima T. and Ito A. (1996) *J. Biochem*.(Tokyo) 120:1163-1166
- Taylor A. B., Smith B. S., Kitada S., Kojima K., Miyaura H., Otwinowski Z., Ito A. And Deisenhofer J. (2001) *Structure* 9: 615-625
- Tuckerman J.R., Gonzalez G., Dioum E.M. and Gilles-Gonzalez M.A. (2002) *Biochemistry* 41: 6170-6177
- von Heijne G. (1986) *EMBO J.* 5: 1335–1342
- Vreede J., van der Horst M.A., Hellingwerf K. J., Crielaard W., and van Aalten D. M. F. (2003) *J. Biol. Chem.* 278**:** 18434-18439
- Yang M.J., Geli V., Oppliger W., Suda K., James P. and Schatz G. (1991). *J. Biol. Chem.* 266: 6416-6423
- Ye X., Demidov A. and Champion P. M. (2002) *J. Am. Chem. Soc.* 124: 5914 -592