From translation initiation factors to enzymes –
a crystallographic study

Inauguraldissertation

der Philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Mario Bumann

von Saas-Fee/VS

Leiter der Arbeit

Prof. Dr. Ulrich Baumann

Departement für Chemie und Biochemie

From translation initiation factors to enzymes –
a crystallographic study

Inauguraldissertation

der Philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Mario Bumann

von Saas-Fee/VS

Leiter der Arbeit

Prof. Dr. Ulrich Baumann

Departement für Chemie und Biochemie

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, den 25. Juni, 2004   Der Dekan

Prof. Dr. G. Jäger

Summary

The work presented in this thesis is divided into three chapters: (I) analysis of two translation initiation factors from eukaryotes (eIF), namely eIF4G and eIF4A; (II) determination of the structure of Ypr118w, a protein homologous to eIF2B regulatory subunits, which is involved in the methionine salvage pathway; and (III) determination of the structure of Enzyme I, the gateway protein to the bacterial phosphotransferase system.

Chapter I:

Initiation of translation is a multi-step pathway which positions the ribosome at the correct AUG codon on the mRNA. Many different eukaryotic initiation factors (eIFs) are involved in the initiation step. Translation initiation factor eIF4G is an essential protein for cap-dependent translation. Regions of eIF4G that interact with other proteins have been identified indicating that eIF4G serves as a scaffolding protein that brings together the mRNA cap-structure (via eIF4E interaction) and the poly(A) tail of the mRNA (via Pabp interaction). eIF4G possesses an eIF4A binding region in the middle domain. eIF4A binds ATP and is a bi-directional helicase which unwinds RNA duplexes in mRNA in order to facilitate ribosome binding. The eIF4G-eIF4A interaction is essential for translation and cell growth.

Four different complexes between eIF4A and eIF4G could be crystallized and were studied by X-ray crystallography, but no structure could be determined. Only the amino-terminal part of eIF4A was localized on one of these crystals by molecular replacement.

Chapter II:

Ypr118w, a non-essential, low-copy number gene product from *Saccharomyces cerevisiae* belongs to the PFAM family PF01008, which contains the α−, β-, and δ-subunits of eIF2B.

Apart from the shared 26-28% sequence identity with eIF2B α-subunits from *Schizzosaccharomyces pombe* and *Saccharomyces cerevisiae,* Ypr118w exhibits 37% sequence identity with the established 5-methylthioribose-1-phosphate isomerase (MTNA) from *Bacillus subtilis*. MTNA is also a member of the PF01008 and participates in the methionine salvage pathway, catalyzing the isomerisation of 5-methylthioribose-1-phosphate to 5-methylthioribulose-1-phosphate. 5-Methylthioribose-1-phosphate isomerase has attracted interest due to its tumor suppression function and as target of antiprotozoal drugs.

The structure of Ypr118w was determined by the method of multiple wavelength anomalous dispersion. It reveals a dimeric protein with two domains and a putative active site cleft. Finally, genetic and biochemical experiments prove that Ypr118w is a 5-methylthioribose-1-phosphate isomerase involved in the methionine salvage pathway.

Chapter III:

Enzyme I (EI) is the first protein in the phosphoryl transfer sequence from phospho*enol*pyruvate (PEP) to sugar in carbohydrate uptake via the bacterial PTS.

In this part we present the three-dimensional structure of the C-terminal domain of enzyme I from *Thermoanaerobacter tengcongensis*. The structure was determined by the method of multiple wavelength anomalous dispersion. The dimeric protein folds into a $(\beta\alpha)_8$ barrel with some large extrusions in the loops connecting $\beta2$ and $\alpha2$, $\beta3$ and $\alpha3$ and $\beta6$ and $\alpha6$. The structure is correlated with the wealth of existing biochemical data.

Contents

Abbreviations

Å	Ångstrøm
AMP	Adenosine monophosphate
β-ME	β-mercaptoethanol
CCD	Charge coupled device
CCP4	Collaborative computing project 4
Da	Dalton
DESY	Deutsches Elektronen-Synchrotron
DLS	Dynamic light scattering
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF	Eukaryotic translation initiation factor
ESRF	European synchrotron radiation facility
GST	Glutathion-S-transferase
HPr	Histidine containing protein
IPTG	Isopropyl-thiogalactopyranoside
MPD	2-Methyl-2,4-pentane-diol
MTA	Methylthioadenoside
NaAc	Sodium azide
NMR	Nuclear magnetic resonance
OD_x	Optical density at x nanometer
PAGE	Polyacrylamide gelelectrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PTS	Phosphotransferase system
RCSB	Research collaboratory for structural bioinformatics
Rmsd	Root mean square deviation
RT	Room temperature
SDS	Sodiumdodecylsulfate
SLS	Swiss light source
Tris	Tris-hydroxymethylaminoethane
XDS	X-ray detector software

Introduction

A. The eIF4G and eIF4A complex

Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. On account of this, synthesis of a protein is one of the most central and basic events in a living system.

There are two major steps in protein synthesis: transcription and translation. The first step in the production of proteins is the transcription of DNA to a mRNA molecule. In the second major step of protein synthesis, ribosomes and Met-tRNA attach to the start codon and begin to translate the mRNA into the amino acid sequence of a protein.

The translation step is divided into initiation, elongation and termination, where initation is the most crucial and rate-limiting step. Translation initiation in eukaryotic cells is catalysed by a large number of eukaryotic initiation factors (eIFs).

Cap-dependent initiation of translation starts with the recognition of the 5`$m^7$GpppG cap structure of the mRNA by the eukaryotic initiation factor eIF4F, a complex of different translation initiation factor polypeptide chains [1]. eIF4F consists of three initiation factors: the scaffold protein eIF4G, the 5`m^7GpppG cap binding protein eIF4E which associates with the N terminus of eIF4G, and the ATP-dependent RNA helicase eIF4A (and its activator eIF4B). eIF4A associates with the middle region of eIF4G. Poly(A)-binding protein (PABP) binds to the 3' poly(A) tail of the mRNA and to the N terminus of eIF4G [2] leading to circularisation of the mRNA.

The 43S complex, consisting of a 40S ribosomal subunit, eIF3 and an eIF2-GTP-initiator tRNA complex, binds to the 5´end of mRNA. The initiator methionyl-tRNA is brought to the ribosome by eIF2. eIF2 is active when bound to GTP but is released as inactive eIF2-GDP after each round of initiation. The exchange of GDP for GTP in order to regenerate active eIF2-GTP is mediated by another factor, eIF2B, which is crucial for translation initiation and

its control. The ribosome-initiation factor complex scans the mRNA in 5` to 3` direction for the first AUG triplet, which acts as the start codon of protein synthesis. eIF5 stimulates GTP hydrolysis by eIF2 which leads to the release of bound factors from the small ribosome subunit, leaving the initiator tRNA in the P-site of the 40S subunit. The 60S ribosomal subunit joins the 40 S ribosomal subunit and protein synthesis begins (reviewed in [3]). Figure 1 gives an overview of the cap-dependent translation initiation.
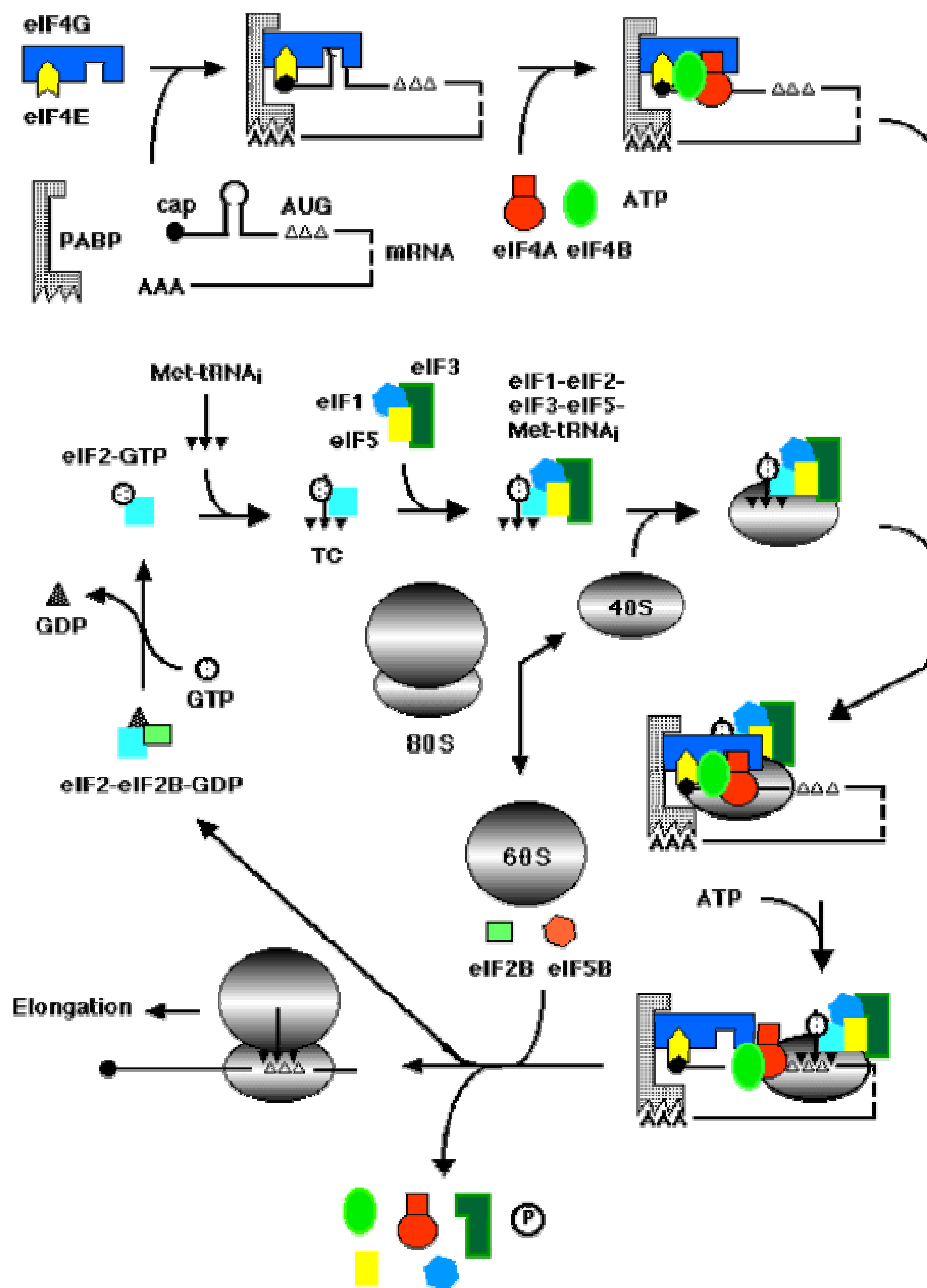

Figure 1: model of cap–dependent translation initiation

(http://ntbiouser.unibe.ch/Trachsel/teaching/translation/initiation/initiationOnline.html)

Among many different eukaryotic initiation factors involved in initiation of translation, eIF4G plays a central role in the assembly of the pre-initiation complex, acting as scaffold protein that interacts with several other initiation factors. The N-terminal part of eIF4G contains binding sites for poly(A)-binding protein (PABP) and eIF4E [4-6]. The central region of eIF4G carries a binding site for the RNA helicase eIF4A [7-9]. The C-terminal region of mammalian eIF4G carries a second binding site for eIF4A [7] and a binding site for the protein kinase Mnk1 [10], which phosphorylates eIF4E. Yeast eIF4G, which was used in this work, lacks the domain corresponding to the C-terminal region of mammalian eIF4G.

The initiation factor eIF4A is a member of the DEAD-box protein family (where DEAD stands for the amino acids Asp-Glu-Ala-Asp). Proteins of the DEAD-box family are highly conserved from bacteria to humans and are involved as helicases in a wide range of cellular processes, including translation, ribosome biogenesis, mRNA splicing, and RNA maturation and degradation [11-13]. eIF4A is an ATP-dependent RNA helicase that is thought to melt the 5` proximal secondary structure of eukaryotic mRNAs to facilitate attachment of the 40S ribosomal subunit [14-16]

The structure determination of the complex eIF4G and eIF4A was one of the objectives of this PhD thesis. Cloning, purification crystallisation and preliminary X-ray analysis of different complexes of eIF4G and eIF4A of *Saccharomyces cerevisiae* are the subjects of chapter I.

B. **Ypr118w**

The yeast protein encoded by the gene *YPR118W* shares homology with the three regulatory subunits of eIF2B (in yeast Gcn3, Gcd7, and Gcd2, *i.e.* the regulatory αβδ-subcomplex).

eIF2B is a GDP-GTP exchange factor which regulates the activity of translation initiation factor eIF2. eIF2B is composed of five subunits – α, β, γ, δ and ε. eIF2B plays an important part in the regulation of translation initiation (reviewed in [3]). eIF2 gets phosphorylated by a number of kinases in response to viral infections, haem-deficiency and certain stress conditions. Phosphorylated eIF2 binds 150 times more strongly to eIF2B than the nonphosphorylated form and therefore acts as a competitive inhibitor in the nucleotide exchange reaction, which consequently leads to inhibition of translation initiation [17].

Apart from the shared 26-28% sequence identity of Ypr118w to eIF2B α-subunits from *Schizzosaccharomyces pombe* and *Saccharomyces cerevisiae,* Ypr118w has 37% sequence identity with the established 5-methylthioribose-1-phosphate isomerase from *Bacillus subtilis,* an enzyme of the methionine salvage pathway. This pathway functions to salvage methylthioadenosine (MTA), which is formed as a by-product of polyamine metabolism (see Figure 2). Polyamines are small, aliphatic amines involved in a variety of cellular processes including transcription and apoptosis [18].

Phosphoshorolysis of MTA by methylthioadenosine phosphorylase (MTAP) results in the conversion of MTA into adenine and 5-methylthioribose-1-phosphate. A series of reactions then salvages the methyl-thio group from methylthioribose-1-phosphate to form methionine. This pathway has been most extensively studied in *Klebsiella pneumoniae* [19-22] but has also been shown to exist in rat liver [23-25] and in *Saccharomyces cerevisiae* [26, 27].
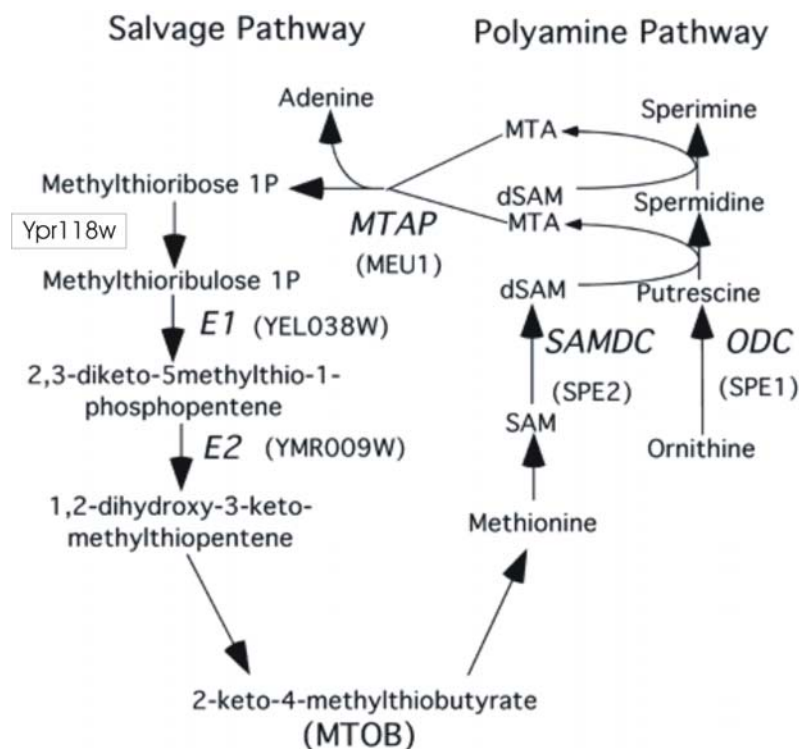
Figure 2: Methionine salvage and polyamine pathways. Enzymes are shown in *italic typing*. Yeast genes are shown in *parenthesis. MTAP,* methylthioadenosine phosphorylase; MTA, methylthioadenosine; *SAMDC, S*-adenosylmethionine decarboxylase; dSAM, decarboxylated *S*-adenosylmethionine; *ODC,* ornithine decarboxylase.

Loss of MTAP activity through gene deletion is common in many kinds of human cancers [28-34]. It was shown that expression of MTAP in an MTAP-deleted breast adenocarcinoma cell line (MCF-7) resulted in a dramatic inhibition of tumourigenicity [35]. The rate-limiting enzyme in the production of polyamines is ornithine decarboxylase (ODC). Experiments with different tumours show that overexpression of ODC is tumourigenic [36-39].

It was demonstrated that products of the methionine salvage pathway regulate polyamine biosynthesis and suggest that MTAP deletion may lead to ODC activation in human tumours [40].

Chapter II presents the X-ray structure of the Ypr118w from *Saccharomyces cerevisiae.* We could show that Ypr118w is indeed a 5-methylthioribose-1-phosphate isomerase.

C. Enzyme I

In collaboration with Prof. Dr. Bernard Erni´s group at the University of Berne we had the opportunity to solve the structure of C-terminal domain of the enzyme I, a protein of the phosphoenolpyruvate:sugar phosphotransferase system (PTS). The enzyme I was cloned by Christoph Bächler, and expressed, purified and crystallised by Philipp Schneider.

The PTS effects the coupled transport and phosphorylation of sugars in bacteria [41]. The reaction is accomplished in five subsequent steps of phosphoryl group (P) transfer from phosphoenolpyruvate (PEP) to carbohydrate, which are catalyzed by PTS "Enzyme I" (EI), phosphocarrier protein HPr, and various sugar-specific "Enzymes II" (Figure 3). For the carbohydrate transport, the EI becomes autophosphorylated by PEP and transfers P to the HPr protein. Next, the phosphoryl group is delivered to the carbohydrate via PTS "Enzyme II" (EII), composed of three protein subunits or domains IIA, IIB, and IIC. The IIA domain becomes phosphorylated by HPr and subsequently passes the phosphate to the IIB domain. Then, a coupled phosphotransfer and transport, involving the membrane-associated IIB-IIC complex, can convert extracellular glucose to intracellular glucose-6-phosphate.
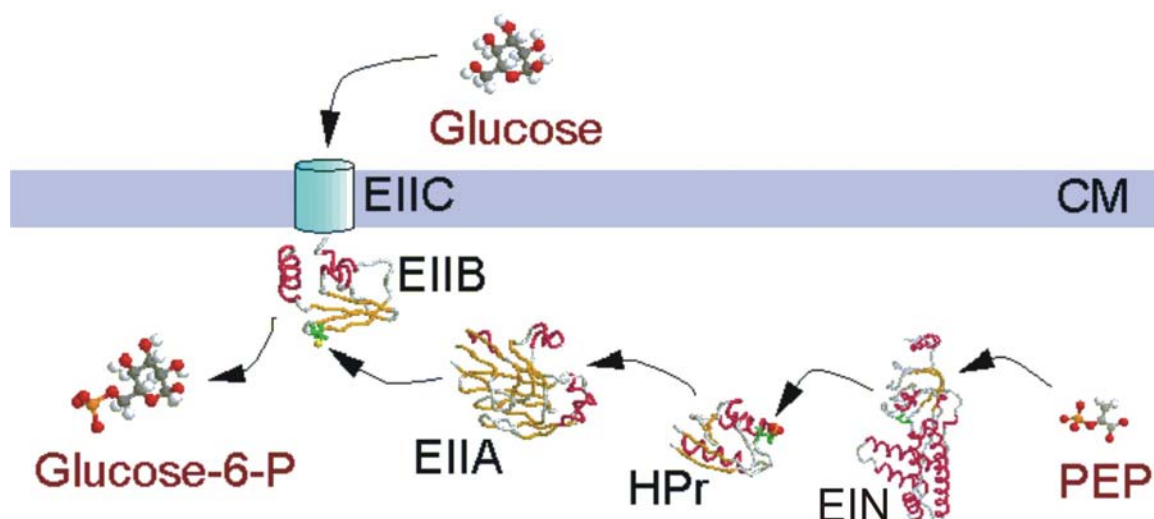

Figure 3: Phosphotransfer chain of the glucose PTS. Proteins are phosphorylated at His or Cys. The structure of EIN includes amino acids 2–249. The histidine-containing phosphocarrier protein (HPr) is a molecule of 85 amino acids. The atom coordinates for amino acids 19 - 168 from EIIA are known. The structure of domain IIB includes amino acids 386-477.

PTS does not only function as a general sugar transport system but is also involved in numerous regulatory pathways [42]. PTS occurs in eubacteria but not in eukaryotes and archeabacteria [43-47].

Enzyme I, a ~64kDa two domain protein, is the gateway to the PTS. Proteolysis and fluorescence studies [48-51] of the *E.coli* protein have shown that enzyme I has a compact amino terminal domain (EIN) containing the active-site His189 and a less structured, flexible carboxyl terminal domain (EIC) which contains the PEP binding site [52] and is more susceptible to proteolysis. The PEP-binding site of E. coli EI has been identified by affinity labelling with the mechanism-based inhibitor Z-cloro-PEP [53] which labelled Cys-502 covalently during catalytic turnover. X-ray crystallographic [54] and NMR solution [55] structures of the amino terminal domain of EI show that the nonphosphorylated active-site His189 is buried near the interface between two subdomains: an α/β-subdomain (residue 1-20 and 148-230) and an α-subdomain (residues 33-143) consisting of four helices arranged in two hairpins. Upon phosphorylation, His189 rotates towards the surface with only small structural changes detected [56]. A solution structure for dephospho-EIN complexed to HPr [57] shows that HPr binds to the α-subdomain of EIN, and this agrees with binding studies using cloned EI variants [58]. The α/β-subdomain is structurally similar to the phosphohistidine swivel domain of pyruvate phosphate dikinase (PPDK) [59].

The mode of action of *E. coli* EI and the control of its activity is not yet fully understood. EIC is responsible for dimer formation. The dimer dissociation constant is strongly affected by temperature [60, 61], divalent cations and PEP [62, 63]. The monomer association rate constant is 2-3 orders of magnitude slower than in other dimeric proteins, indicating that dimerization is accompanied by major conformational rearrangements of the interacting EIC domains [62, 64].

Chapter III presents the X-ray structure of the EIC domain from *Thermoanaerobacter tengcongensis*, a gram-negative, saccharolytic bacterium which grows optimally at 75°C [65,

66]. This structure allows an accurate description of the active site and the rationalization of

inhibitor studies.

References

1.   Hershey, J.W.B.M., W.C., Pathway and mechanism of initiation of protein synthesis. In: Translational Regulation of Gene Expression. 2000: Cold Spring Harbor Laboratory Press. 33-88.
2.   Gingras, A.C., B. Raught, and N. Sonenberg, eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem*, 1999. **68**: p. 913-63.
3.   Dever, T.E., Gene-specific regulation by general translation factors. *Cell*, 2002. **108**(4): p. 545-56.
4.   Imataka, H., A. Gradi, and N. Sonenberg, A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *Embo J*, 1998. **17**(24): p. 7480-9.
5.   Lamphear, B.J., et al., Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem*, 1995. **270**(37): p. 21975-83.
6.   Mader, S., et al., The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol Cell Biol*, 1995. **15**(9): p. 4990-7.
7.   Imataka, H. and N. Sonenberg, Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol Cell Biol*, 1997. **17**(12): p. 6940-7.
8.   Neff, C.L. and A.B. Sachs, Eukaryotic translation initiation factors 4G and 4A from Saccharomyces cerevisiae interact physically and functionally. *Mol Cell Biol*, 1999. **19**(8): p. 5557-64.
9.   Dominguez, D., Altmann, M., Benz, J., Baumann, U. & Trachsel, H., Interaction of translation initiation factor eIF4G with eIF4A in the yeast Saccharomyces cerevisiae. *J Biol Chem*, 1999. **274**: p. 26720-26726.
10.   Pyronnet, S., et al., Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *Embo J*, 1999. **18**(1): p. 270-9.
11.   de la Cruz, J., D. Kressler, and P. Linder, Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. *Trends Biochem Sci*, 1999. **24**(5): p. 192-8.
12.   Gorbalenya, A.E., et al., Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res*, 1989. **17**(12): p. 4713-30.
13.   Wassarman, D.A. and J.A. Steitz, RNA splicing. Alive with DEAD proteins. *Nature*, 1991. **349**(6309): p. 463-4.
14.   Merrick, W.C.a.H., J.W.B., *The pathway and mechanism of protein synthesis in translational control*, in *Translational Control*, J.W.B. (Hershey, Mathews, M.B. and Sonenberg, N., Editor. 1996, Cold Spring Harbor Press, Cold Spring Harbor, NY. p. pp. 31–69.

15.   Grifo, J.A., et al., RNA-stimulated ATPase activity of eukaryotic initiation factors. *J Biol Chem*, 1984. **259**(13): p. 8648-54.

16.   Pause, A. and N. Sonenberg, Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *Embo J*, 1992. **11**(7): p. 2643-54.

17.   Dever, T.E., et al., Modulation of tRNA(iMet), eIF-2, and eIF-2B expression shows that GCN4 translation is inversely coupled to the level of eIF-2.GTP.Met-tRNA(iMet) ternary complexes. *Mol Cell Biol*, 1995. **15**(11): p. 6351-63.

18.   Pegg, A.E., Polyamine metabolism and its importance in neoplastic growth and a target for chemotherapy. *Cancer Res*, 1988. **48**(4): p. 759-74.

19.   Trackman, P.C. and R.H. Abeles, The metabolism of 1-phospho-5-methylthioribose. *Biochem Biophys Res Commun*, 1981. **103**(4): p. 1238-44.

20.   Trackman, P.C. and R.H. Abeles, Methionine synthesis from 5'-S-Methylthioadenosine. Resolution of enzyme activities and identification of 1-phospho-5-S methylthioribulose. *J Biol Chem*, 1983. **258**(11): p. 6717-20.

21.   Furfine, E.S. and R.H. Abeles, Intermediates in the conversion of 5'-S-methylthioadenosine to methionine in Klebsiella pneumoniae. *J Biol Chem*, 1988. **263**(20): p. 9598-606.

22.   Myers, R.W. and R.H. Abeles, Conversion of 5-S-methyl-5-thio-D-ribose to methionine in Klebsiella pneumoniae. Stable isotope incorporation studies of the terminal enzymatic reactions in the pathway. *J Biol Chem*, 1990. **265**(28): p. 16913-21.

23.   Backlund, P.S., Jr. and R.A. Smith, Methionine synthesis from 5'-methylthioadenosine in rat liver. *J Biol Chem*, 1981. **256**(4): p. 1533-5.

24.   Backlund, P.S., Jr., C.P. Chang, and R.A. Smith, Identification of 2-keto-4-methylthiobutyrate as an intermediate compound in methionine synthesis from 5'-methylthioadenosine. *J Biol Chem*, 1982. **257**(8): p. 4196-202.

25.   Backlund, P.S., Jr. and R.A. Smith, 5'-Methylthioadenosine metabolism and methionine synthesis in mammalian cells grown in culture. *Biochem Biophys Res Commun*, 1982. **108**(2): p. 687-95.

26.   Cone, M.C., et al., Utilization by Saccharomyces cerevisiae of 5'-methylthioadenosine as a source of both purine and methionine. *J Bacteriol*, 1982. **151**(1): p. 510-5.

27.   Marchitto, K.S. and A.J. Ferro, The metabolism of 5'-methylthioadenosine and 5-methylthioribose 1-phosphate in Saccharomyces cerevisiae. *J Gen Microbiol*, 1985. **131 (Pt 9)**: p. 2153-64.

28.   Della Ragione, F., et al., 5'-Deoxy-5'-methylthioadenosine phosphorylase and p16INK4 deficiency in multiple tumor cell lines. *Oncogene*, 1995. **10**(5): p. 827-33.

29.   Batova, A., et al., Frequent deletion in the methylthioadenosine phosphorylase gene in T-cell acute lymphoblastic leukemia: strategies for enzyme-targeted therapy. *Blood*, 1996. **88**(8): p. 3083-90.

30.   Nobori, T., et al., Genomic cloning of methylthioadenosine phosphorylase: a purine metabolic enzyme deficient in multiple different cancers. *Proc Natl Acad Sci U S A*, 1996. **93**(12): p. 6203-8.

31.   Garcia-Castellano, J.M., et al., Methylthioadenosine phosphorylase gene deletions are common in osteosarcoma. *Clin Cancer Res*, 2002. **8**(3): p. 782-7.

32.   Schmid, M., et al., Homozygous deletions of methylthioadenosine phosphorylase (MTAP) are more frequent than p16INK4A (CDKN2) homozygous deletions in primary non-small cell lung cancers (NSCLC). *Oncogene*, 1998. **17**(20): p. 2669-75.

33.   Stadler, W.M. and O.I. Olopade, The 9p21 region in bladder cancer cell lines: large homozygous deletion inactivate the CDKN2, CDKN2B and MTAP genes. *Urol Res*, 1996. **24**(4): p. 239-44.

34. Nobori, T., et al., Absence of methylthioadenosine phosphorylase in human gliomas. *Cancer Res*, 1991. **51**(12): p. 3193-7.

35. Christopher, S.A., et al., Methylthioadenosine phosphorylase, a gene frequently codeleted with p16(cdkN2a/ARF), acts as a tumor suppressor in a breast cancer cell line. *Cancer Res*, 2002. **62**(22): p. 6639-44.

36. O'Brien, T.G., et al., Ornithine decarboxylase overexpression is a sufficient condition for tumor promotion in mouse skin. *Cancer Res*, 1997. **57**(13): p. 2630-7.

37. Megosh, L., et al., Increased frequency of spontaneous skin tumors in transgenic mice which overexpress ornithine decarboxylase. *Cancer Res*, 1995. **55**(19): p. 4205-9.

38. Auvinen, M., et al., Ornithine decarboxylase activity is critical for cell transformation. *Nature*, 1992. **360**(6402): p. 355-8.

39. Cohen, S.S., A Guide to the Polyamines. 1998: Oxford University Press, Oxford. pp. 296–319.

40. Subhi, A.L., et al., Methylthioadenosine phosphorylase regulates ornithine decarboxylase by production of downstream metabolites. *J Biol Chem*, 2003. **278**(50): p. 49868-73.

41. Kundig, W.G., S. Roseman, S., Phosphate Bound to Histidine in a Protein as an Intermediate in a Novel Phospho-Transferase System. *Proc Natl Acad Sci U S A*, 1964. **52**: p. 1067-74.

42. Postma, P.W.L., J. W. Jacobson, G. R., Escherichia coli and Salmonella: cellular and molecular biology. 1996: Washington, DC.

43. Reizer, J.R., A. Saier, M. H., Jr., Novel phosphotransferase system genes revealed by bacterial genome analysis--a gene cluster encoding a unique Enzyme I and the proteins of a fructose-like permease system. *Microbiology*, 1995. **141 (Pt 4)**: p. 961-71.

44. Reizer J, C.A., Reizer A, Saier MH, Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar-specific permease domains of the phosphotransferase system and pentose catabolic enzymes. *Genome Sci Technol*, 1996. **1**: p. pp. 53-57.

45. Reizer, J.R., A. Merrick, M. J. Plunkett, G., 3rd Rose, D. J. Saier, M. H., Jr., Novel phosphotransferase-encoding genes revealed by analysis of the Escherichia coli genome: a chimeric gene encoding an Enzyme I homologue that possesses a putative sensory transduction domain. *Gene*, 1996. **181**(1-2): p. 103-8.

46. Reizer, J.B., S. Reizer, A. Arnaud, M. Saier, M. H., Jr. Stulke, J., Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of Bacillus subtilis. *Microbiology*, 1999. **145 (Pt 12)**: p. 3419-29.

47. Tchieu, J.H.N., V.Edwards, J. S.Saier, M. H., Jr., The complete phosphotranferase system in Escherichia coli. *J Mol Microbiol Biotechnol*, 2001. **3**(3): p. 329-46.

48. Chauvin, F.B., L. Roseman, S., Enzyme I: the first protein and potential regulator of the bacterial phosphoenolpyruvate: glycose phosphotransferase system. *Res Microbiol*, 1996. **147**(6-7): p. 471-9.

49. LiCalsi, C.C., T. S. Freire, E. Roseman, S., Sugar transport by the bacterial phosphotransferase system. Structural and thermodynamic domains of enzyme I of Salmonella typhimurium. *J Biol Chem*, 1991. **266**(29): p. 19519-27.

50. Han, M.K.R., S. Brand, L., Sugar transport by the bacterial phosphotransferase system. Characterization of the sulfhydryl groups and site-specific labeling of enzyme I. *J Biol Chem*, 1990. **265**(4): p. 1985-95.

51. Lee, B.R.L., P.Pannell, L.Jaffe, H.Peterkofsky, A., Identification of the N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar

phosphotransferase system produced by proteolytic digestion. *Arch Biochem Biophys*, 1994. **312**(1): p. 121-4.

52. Seok, Y.J.L., B. R.Gazdar, C.Svenson, I.Yadla, N.Peterkofsky, A., Importance of the region around glycine-338 for the activity of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry*, 1996. **35**(1): p. 236-42.

53. Garcia-Alles, L.F., I. Alfonso, and B. Erni, Enzyme I of the phosphotransferase system: induced-fit protonation of the reaction transition state by Cys-502. *Biochemistry*, 2003. **42**(16): p. 4744-50.

54. Liao, D.I.S., E.Seok, Y. J.Lee, B. R.Peterkofsky, A.Davies, D. R., The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the E. coli PEP: sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure*, 1996. **4**(7): p. 861-72.

55. Garrett, D.S.S., Y. J.Liao, D. I.Peterkofsky, A.Gronenborn, A. M.Clore, G. M., Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry*, 1997. **36**(9): p. 2517-30.

56. Garrett, D.S.S., Y. J.Peterkofsky, A.Clore, G. M.Gronenborn, A. M., Tautomeric state and pKa of the phosphorylated active site histidine in the N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. *Protein Sci*, 1998. **7**(3): p. 789-93.

57. Garrett, D.S., et al., Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat Struct Biol*, 1999. **6**(2): p. 166-73.

58. Zhu, P.P., et al., Reconstitution studies using the helical and carboxy-terminal domains of enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry*, 1999. **38**(47): p. 15470-9.

59. Herzberg, O., et al., Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc Natl Acad Sci U S A*, 1996. **93**(7): p. 2652-7.

60. Chauvin, F., L. Brand, and S. Roseman, Sugar transport by the bacterial phosphotransferase system. Characterization of the Escherichia coli enzyme I monomer/dimer equilibrium by fluorescence anisotropy. *J Biol Chem*, 1994. **269**(32): p. 20263-9.

61. Kukuruzinska, M.A., et al., Subunit association of enzyme I of the Salmonella typhimurium phosphoenolpyruvate: glycose phosphotransferase system. Temperature dependence and thermodynamic properties. *J Biol Chem*, 1984. **259**(19): p. 11679-81.

62. Chauvin, F., L. Brand, and S. Roseman, Sugar transport by the bacterial phosphotransferase system. Characterization of the Escherichia coli enzyme I monomer/dimer transition kinetics by fluorescence anisotropy. *J Biol Chem*, 1994. **269**(32): p. 20270-4.

63. Dimitrova, M.N., et al., Interdomain interaction and substrate coupling effects on dimerization and conformational stability of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry*, 2002. **41**(3): p. 906-13.

64. Han, M.K., et al., Sugar transport by the bacterial phosphotransferase system. Fluorescence studies of subunit interactions of enzyme I. *J Biol Chem*, 1990. **265**(4): p. 1996-2003.

65. Xue, Y., et al., Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol*, 2001. **51**(Pt 4): p. 1335-41.

66.   Bao, Q., et al., A complete sequence of the T. tengcongensis genome. *Genome Res*, 2002. **12**(5): p. 689-700.

Chapter I

Cloning, purification, crystallisation and preliminary X-ray analysis of different complexes of eIF4G and eIF4A of *Saccharomyces cerevisiae*

Introduction

Many different eukaryotic inititation factors (eIF) are involved in the initiation of protein synthesis in eukaryotic cells. The process of cap-dependent initiation can be divided in four major events (see for review [1]):

1. The heteromultimeric complex eukaryotic initiation factor 4F recognises the m^7G cap structure at the 5'end of mRNA. It consists of the cap binding protein eIF4E, the RNA-dependent helicase eIF4A and the scaffold factor eIF4G. The poly(A)tail at the 3'-end of mRNA binds PABP (the poly(A)-binding protein) and interacts with eIF4G. This leads to circularisation of mRNA via end-to-end complex formation.

2. The 43S pre-initiation complex containing the 40S ribosomal subunit, eIF3 and a ternary complex of eIF2, GTP and Met-tRNAi, binds to eIF4F and forms the 48S complex.

3. The 48S complex scans along the 5'-untranslated region (5'UTR) in 5' to 3' direction.

4. The first AUG triplet is recognised by base-pairing between the initiation codon and the anticodon of initiator tRNA in the scanning ribosomal complex which triggers hydrolysis of eIF2-bound GTP, dissociation of the initiation factors, and subsequent binding of the large 60S ribosomal subunit to the 40S subunit to form the translation competent 80S ribosome

Here, particular attention shall be paid to the first event of initiation, especially to initiation factors eIF4G and eIF4A.

eIF4G, which exists in two isoform in plants [2], in mammals [3], and the yeast *Saccharomyces cerevisiae* [4], plays an essential role in translation initiation. The N-terminal half of eIF4G contains a PABP interacting motif [5] [6] and a Tyr-X_4-Leu-Φ (Φ denotes a hydrophobic amino acid) eIF4E-recognition motif [7].

The central region of eIF4G carries binding sites for the RNA helicase eIF4A (phylogenetically conserved) [8-10] and for the multi-subunit factor eIF3 [8, 11, 12], which itself binds to the small ribosomal subunit and recruits it as the 43S initiation complex for binding to the mRNA. eIF3 binding to yeast eIF4G (in contrast to mammalian eIF4G) could not be shown so far (M. Altmann and H. Trachsel, unpublished work).

The C-terminal part of mammalian eIF4G contains a second binding site for eIF4A [8] and a binding site for the protein kinase Mnk1, which phosphorylates eIF4E [13]. This C-terminal part is missing in yeast eIF4G but it was shown that this domain is not essential for translation initiation, whereas the central domain is essential for cap-dependent and internal initiation [8, 11].

The three-dimensional structure of the phylogenetically conserved middle domain of human eIF4GII (745-1003) has been determined by X-ray crystallography (Figure 1) [14]. This part is responsible for interactions with eIF4A. The crescent-shaped molecule consists of ten α-

helices. A right-handed solenoid is formed by the polypeptide chain. Comparisons of this structure disclosed similarities to various HEAT repeat proteins. HEAT repeat proteins participate in a wide variety of cellular processes that are dependent on assembling large multiprotein complexes [15].

Figure 1: Ribbon drawing of the conserved central region of eIF4GII viewed along the cylindrical axes of the α -helices [14].

eIF4A is an ATP-dependent RNA helicase belonging to the DEAD-box protein family. Sequence analysis and biochemical and genetic experiments with mutant eIF4A have revealed nine conserved sequence motifs. The ATPase A motif, also called Walker A motif (AxxGxGKT) [16], has been shown to be required for ATP binding, the ATPase B motif (DEAD) [17] is needed for ATP hydrolysis, the SAT and HRIGR motifs are important for helicase activity and protein-RNA interactions, respectively [18-21].

As described above, initiation factors eIF4A and eIF4B, as well as a part of eIF4F, interact with the mRNA and disrupt secondary/tertiary structures in the 5' untranslated region (UTR) through the helicase activity of eIF4A [17].

eIF4A consists of two domains, an amino- and a carboxyl-terminal domain which are connected by an extended, 11-residue linker [22]. The C-terminal domain alone is sufficient to bind eIF4G.

The structure of both domains has been reported. [22-24]. The amino-terminal domain (Figure 2) has an αβα fold. The protein contains a central, seven-stranded, twisted β sheet, flanked by nine α-helices. [23]
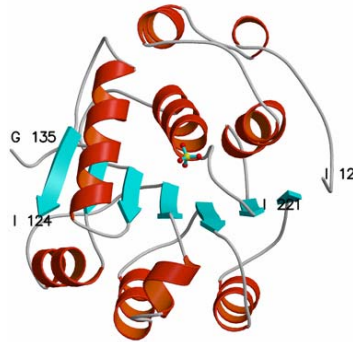
Figure 2: Structure of the ATPase domain of eIf4A. Ribbon diagram of the structure. The blue arrows represent β strands and the red ribbons represent α helices [23].

The carboxyl-terminal domain (Figure 3) has a parallel α-β structure. The conserved core is composed of six parallel β-strands plus helices 1, 2, 4, and 5.
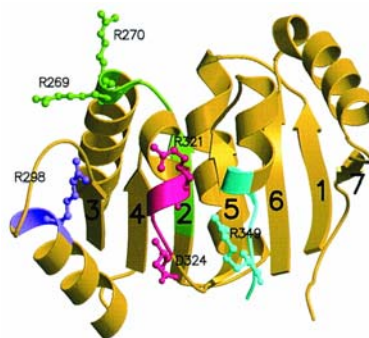
Figure 3: Ribbon drawing of the structure of the carboxyl-terminal domain of eIF4A. Conserved motifs are coloured as follows: motif IV, VIFCNTRR, residues 263-270, green; "conserved R" motif, residue Arg-298, purple; motif V, RGID, residues 321-324, magenta; motif VI, HRIGRGGR, residues 345-352, cyan. The strands of the β-sheet are labelled sequentially [22].

Motif IV participates in oligonucleotide binding. The side chain of Arg-298 belongs to a "QXXR" motif that is specific to the DEA(D/H)-box helicases. In the DNA helicases, residues from the motif V "bridge" the ATP-binding site and the the oligonucleotide-binding site [22].

We attempted to solve the three-dimensional structure of the complex between eIF4A and eIF4G in order to analyze the interaction between these proteins and gain a deeper understanding of the molecular details of eukaryotic translation initiation.

Materials and Methods

Plasmid Construction, mutagenesis

The *Saccharomyces cerevisiae TIF4631* (eIF4G constructs) and *TIF1* (eIF4A constructs) were subcloned by PCR methods into the *E. coli* expression vector pET28 (Novagen) using the *NdeI* and the *XhoI* sites. This construct possesses a hexahistidine tag and a thrombin cleavage site at the N-terminus.

The mutants of pET28a-eIF4A226-395 were produced by the QuickChange (Stratagene) method using appropriate mutagenic oligonucleotides. The pET28a-eIF4A226-395 plasmid was used as template. Diagnostic restriction sites were inserted to verify the mutations.

Protein expression and purification

Escherichia coli strain BL21-CodonPlus[TM](DE3)–RIL (Stratagene) in the case of eIF4G and the *Escherichia coli* strain Star[TM] (DE3) (Invitrogen) for eIF4A were used for expression. Bacteria were cultivated at 20°C in LB broth and induced for protein expression with 1.0 mM isopropyl-beta-D-thiogalactopyranoside at an OD_{600} of 1.0. Cell growth was continued over night. Cells were harvested by centrifugation (4500g) at 4°C and resuspended in lysis-buffer (20 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0). Cells were disrupted by a French press (twice) and subsequent sonication for 30 seconds on ice. The debris was pelleted by centrifugation at 4°C (4000g) for 45 min. The supernatant was applied to a 10 ml metal-chelate affinity column (Ni-NTA, Qiagen) and the column was washed with 20 mM imidazole until the baseline at 280 nm had stabilized. eIF4A or eIF4G were eluted with 250 mM imidazole. 1 U of thrombin (Pharmacia) per milligram of protein was added and the eluate was dialysed against 20 mM Tris-HCl pH 7.5 at 18°C over night. Phenyl-methylsulfonyl fluoride was added to a final concentration of 0.5 mM for thrombin inhibition. Each protein was purified further separately by gel filtration using a Superdex75 (Amersham

Bioscience) column equilibrated in 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.01 % NaAc, pH 7.5.

Typical yields were 40 mg/l pure protein for the eIF4A constructs and 10 mg/l for the eIF4G constructs.

Selenomethionine labelled proteins were prepared by the method of methionine biosynthesis inhibition [25] and purified as described above with yields of 40 mg/l and 10 mg/l for eIF4A and eIF4G, respectively.

Formation of the complexes eIF4G(542-883)/eIF4A(1-395) and eIF4G(572-853)/eIF4A(1-395)

After determination of protein concentration by UV spectroscopy at 280 nm using extinction coefficients of 0.951 and 1.086 for eIF4G(542-883) and eIF4G(572-853) and 0.345 for eIF4A, respectively, the two proteins were mixed in 1:1 molar ratio and concentrated. The complex was purified by gel filtration. The same column and the same buffer were used as described before.

Complex 4G(572-853)/4A(9-395) and complex 4G(572-853)/4A(226-395)

Instead of gel filtration an ion-exchange column was performed as a second purification step for the individual components eIF4G and eIF4A. The sample was loaded onto a 1 ml Resource Q^{TM} column (Amersham Biosciences), equilibrated with 20 mM Tris-HCl , pH 8.0. The bound protein was eluted with a linear NaCl gradient (0-0.5 M) in the same buffer. After protein concentration determination (described above) the proteins were mixed in 1:1 molar ratio. The protein-complex was purified by gel-filtration using the Superdex S75 column and buffer conditions described above (Amersham Bioscience).

Analysis of the fragments

All protein fragments were analysed by SDS-PAGE and mass spectroscopy. DNA constructs were checked by DNA sequencing. Mass spectrometric analyses were carried out by the analytical research services of the Chemistry Departement, University of Berne.

Dynamic light scattering (DLS)

Dynamic light scattering was performed on a single wavelength, fixed angle DynaPro system (Protein Solutions). Protein samples were centrifuged at 14000 g for 10 min and measured in UVette ®220 – 1600 nm Eppendorf cuvettes. Size distribution analysis was performed with the program Dynamics 5.0.

Crystallisation

Initial crystallisation experiments were based on the sparse-matrix sampling method [26] using Crystal Screen I and II, Natrix- and PEG/ION-Screen from Hampton Research. For complexes eIF4G(572-853)/eIF4A(9-395) and eIF4G(572-853)/eIF4A(226-395) Index Screen I and II from Hampton Research and Wizard Screen I and II from Emerald BioStructures were carried out additionally. Trials were duplicated at 277 and 291K. The droplets consisted of 1 μl protein solution (10-20 mg/ml, 5 mM Tris, pH 7.5) mixed with an equal volume of reservoir solution. Promising crystallisation conditions were optimised by fine grid screenings of pH, drop ratio, protein-, precipitant- and salt concentrations as well as by Additive Screen I- III and Detergent Screen I-III from Hampton Research.

Data collection and analysis

Diffraction data were collected in-house at 291 K and 110 K on a RaxisIV image-plate detector mounted on a Rigaku RU300 rotating-anode generator operating at 100mA and 50

kV and at the Swiss Light Source at 100 K on beamline X06SA with a MAR CCD detector (Marresearch GmbH, Hamburg, Germany). Data were processed with *XDS* [27-29].

Structure solution

The molecular replacement method was used for structure determination. The N-terminal domain (RCSB PDB code 1QDE) [23] and the C-terminal domain (RCSB PDB code 1FUK) [24] of eIF4A from *S. cerevisiae* and the middle domain (RCSB PDB code 1HU3) [14] of eIF4G from *Homo sapiens* (33 % sequence identity) were used as search models. The molecular replacement was performed with the program MOLREP [30] Version 7.3 as implemented in the CCP4 suite. The solutions were inspected using the graphics program O [31].

Results

Overview

The following table 1 gives an overview of all the cloned, expressed and purified fragments of eIF4A and eIF4G. The starting point was the fragments eIF4G (542-883) and eIF4A(1-395). The eIF4A binding site was mapped to amino acids 542-883 of yeast eIF4GI [10].

Table 1: *Overview of the cloned eIF4G- and eIF4A-fragments*

Species	Protein	Fragment size (AA)	Plasmid
S. cerevisiae	eIF4G	542-883	pET28a
S. cerevisiae	eIF4G	552-883	pET28a
S. cerevisiae	eIF4G	562-883	pET28a
S. cerevisiae	eIF4G	572-883	pET28a
S. cerevisiae	eIF4G	542-862	pET28a
S. cerevisiae	eIF4G	552-862	pET28a
S. cerevisiae	eIF4G	562-862	pET28a
S. cerevisiae	eIF4G	572-862	pET28a
S. cerevisiae	eIF4G	542-853	pET28a
S. cerevisiae	eIF4G	552-853	pET28a
S. cerevisiae	eIF4G	562-853	pET28a
S. cerevisiae	eIF4G	572-853[b]	pET28a
S. cerevisiae	eIF4G	582-853	pET28a
Homo sapiens	eIF4G[a]	676-970	pGEX6P3
S. cerevisiae	eIF4A	1-395	pET28a/pET22b
S. cerevisiae	eIF4A	9-395	pET28a
S. cerevisiae	eIF4A	226-395	pET28a/pET22b
S. cerevisiae	eIF4A	230-395	pET28a
Homo sapiens	eIF4A	1-406	pET28a

[a]Cloned and expressed by Siamak Djafarzadeh

[b]Pointmutation (K591A)

First, the purification and crystallisation of the complex eIF4G(542-883)/eIF4A(1-395) was initiated, because the eIF4A binding site was mapped to amino acids 542-883 of yeast eIF4GI [10]. One question concerned the minimal length of the eIF4G fragment with full binding activity for eIF4A. The smallest fragment able to form a complex with eIF4A was the one from amino acid 572 to 853. The next shorter fragment, eIF4G(582-853), did not form a stable complex with full-length eIF4A. Figure 4 shows the results of the binding studies of N-terminal truncated eIF4G-fragments

							Binding to eIF4A
1	188	299	441	490	542	853	952

Domains: PABP, eIF4E, eIF4A

Fragment	Binding to eIF4A
441 – 539	-
883 – 952	-
592 – 862	-
542 – 883	+
542 – 853	+
552 – 853	+
562 – 853	+
572 – 853	+
582 – 853	-
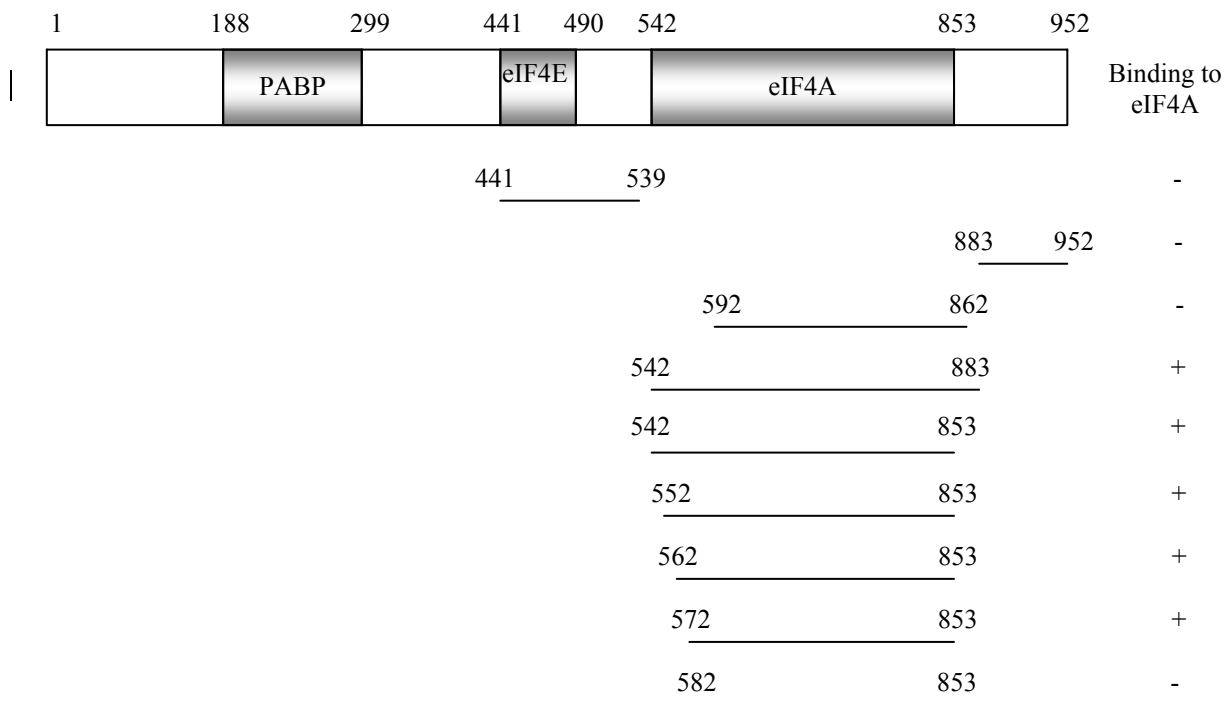
Figure 4: *Results of the binding studies of N-terminal truncated eIF4G-fragments.*

In eIF4A, the C-terminal fragment eIF4A(226-395) suffices to bind eIF4G. eIF4A(230-395), in contrast, shows a drastically reduced binding activity. The four missing amino acids, which are located in the flexible 11-residue linker between the two domains of eIF4A, therefore, seem to play a role in eIF4G binding.

Crystallisation

Table 2 shows all the complexes of eIF4G and eIF4A that have been crystallised. Dynamic light-scattering analysis of these complexes showed a monomodal size distribution except for the human complex.

Table 2: *Overview over the crystallised complexes of eIF4G/eIF4A*

eIF4G	eIF4A	Crystallisation	measurable crystals
542-883[a]	1-395[a]	needles	√
542-883[a]	226-395[a]	needles	
552-883[a]	1-395[a]	needles	
572-883[a]	1-395[a]	no crystals	
572-883[a]	226-395[a]	no crystals	
552-862[a]	1-395[a]	needles	
552-853[a]	1-395[a]	no crystals	
552-853[a]	226-395[a]	no crystals	
562-853[a]	1-395[a]	no crystals	
562-853[a]	226-395[a]	no crystals/spherulites	
572-853[a]	1-395[a]	crystals	√
572-853[a]	9-395[a]	plates	√
572-853[a]	226-395[a]	crystals	√
676-970[b]	1-406[b]	no crystals	

[a]*S. cerevisiae*

[b]*Homo sapiens*

The complex eIF4G(542-883)/eIF4A(1-395)

Crystallisation was performed using the sitting-drop vapour-diffusion method at 277 K. Initial crystallisation trials were conducted using commercially available screening kits from Hampton Research (Crystal Screen I and PEG/Ion Screen). Needles were obtained under various conditions (Conditions 5, 7, 16 and 20) from PEG/Ion Screen. Condition 7 (0.2 M calcium chloride, pH 5.1, 20 % (w/v) PEG 3350) was refined and the best crystals were obtained using the vapour-diffusion method at 277 K with a hanging drop consisting of 0.2 M calcium chloride, pH 5.1 20 % (w/v) PEG 3350, 2 % (v/v) MPD. Figure 5 shows crystals of the complex eIF4G(542-883)/eIF4A(1-395).
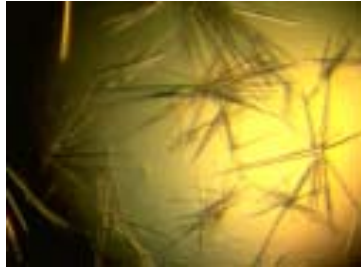
Figure 5: *improved crystals of the complex eIF4G(542-883)/eIF4A(1-395)*

To verify whether the needle-shaped crystals ($100 \times 4 \times 4\ \mu m^3$) are salt or protein crystals the complex were capillary-mounted and measured in-house at 291 K. No diffraction pattern has been observed. Most likely, this diffraction behaviour is due to insufficient internal crystal order or by the size of the crystals.

The complex eIF4G(572-853)/eIF4A(1-395)

Sequencing of pET28a-eEI4G572-853 revealed a conflict with the published sequence at residue 591, where the lysine of the published sequence is replaced by an arginine. Residue 591 is not conserved among different species. This mutation might be helpful for Crystallisation because of reduced flexibility of the Arg sidechain compared with Lys.

eIF4G(572-853)/eIF4A(1-395) crystals (Figure 6) were obtained by hanging drop vapour diffusion at 277 K against a reservoir of 0.2 M potassium sodium tartrate tetrahydrate, pH 7.2, 20 % (w/v) PEG 3350 and 10 mM AMP.

Figure 6: *Improved crystals from the eIF4G(572-853)/eIF4A(1-395)*

Crystal cryoprotection prior to quick freezing by immersion in liquid nitrogen was achieved by transfer to 0.2 M potassium sodium tartrate tetrahydrate, pH 7.2, 20 % (w/v) PEG 3350 and 20 % (v/v) glycerol. Although many different cryoprotectants were tried, e.g. 5-20 % glycerol (direct and serial transfer), 5-20 % glycerol with 10 mM AMP (direct and serial transfer), 5-35 % MPD with 10 mM AMP (direct and serial transfer), 5-40 % PEG 400 with 10 mM AMP (direct and serial transfer), it was not possible to freeze more than one crystal without any serious damages like cracks or disorder of crystal lattice. One data set was collected in-house. The crystal belongs to the space group P2 or P2$_1$ (unit cell: a=68.84 Å, b=111.13 Å, c=112.04 Å, α=γ=90°, β=99.55°). The data collection statistics are reported in table 3.

Table 3: X-ray diffraction data statistics.

	eIF4G(572-853)/ eIF4A(1-395)
Space group	P2$_1$
Cell dimensions	a=68.84 Å, b=111.13 Å, c=112.04 Å, β=99.55°
Beamline	In-house
Wavelength (Å)	1.5418
Resolution range[a] (Å)	20-3.05 (3.34 -3.05)
No. measurements	113559
Unique reflections	32059
R$_{sym}$[b] (%)	9.1 (33.5)
I/σ(I)	11.2(3.74)
Completeness (%)	99.9 (100)

[a] The values in parentheses of resolution range, R$_{sym}$, I/σ(I) and completeness correspond to the outermost resolution shell
[b] Rsym = $\Sigma_{hkl}\Sigma_j$|I(*hkl;j*) - ⟨I(*hkl*)⟩|/($\Sigma_{hkl}\Sigma_j$⟨I(*hkl*)⟩) where I(*hkl;j*) is the *j*th measurement of the intensity of the unique reflection (*hkl*) and ⟨I(*hkl*)⟩ is the mean over all symmetry-related measurements.

The data quality was checked by using the program SFCHECK[32]. Molecular replacement was performed with the program MOLREP [30] using data in the resolution range 20 – 4 Å. The N-terminal domain [23] and the C-terminal domain [24] of eIF4A from *S. cerevisiae* and

the middle domain [14] of eIF4G from *Homo sapiens* (33 % sequence identity) were used as search models.

From the estimated V_M value (Matthew's coefficient of 2.7) and a solvent content of 54.2 % it was expected that the asymmetric unit contained two complexes. A native Patterson function calculated in the resolution range 20 to 4 Å showed a prominent pseudo-origin peak at u,v,w= 0.5,0.5,0.25 with a height of 45.3 % of the origin peak. This hints at a local twofold parallel to the crystallographic 2_1 axis. Molecular replacement was carried out on the assumption that two molecules should exist in the asymmetric unit. The rotation function and the translation function were calculated at 4 Å resolution. The first solution with two molecules per asymmetric unit was found as the highest peaks in the rotation function and the translation function, yielding a correlation coefficient of 0.552 and an *R* factor of 0.518. The N-terminal domain of eIF4A could be located, but neither the C-terminal domain of eIF4A nor the eIF4G domain were found in the search procedure.

To verify that both components of the complex were in the crystal, a SDS-PAGE analysis was carried out (Figure 8).
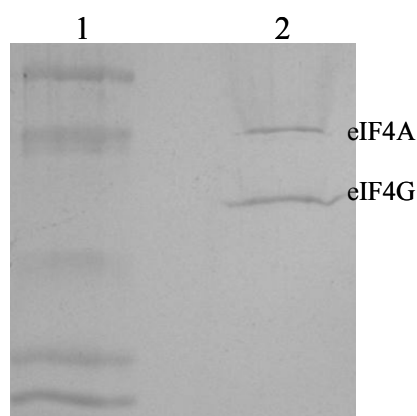
Figure 8: *SDS-Page of the complex eIF4G(572-853)/eIF4A(1-395). 4 crystals were washed and analysed to identify the crystal content. Both proteins, eIF4A and eIF4G were found. 1: Molecular weight markers (66, 45, 24.6, 18.7, 14.3 kDa); 2: eIF4A (44.5 kDa), eIF4G (32 kDa).*

By using the graphic program O [31] it was shown that the two molecules in asymmetric unit cannot form a crystal lattice by itself. Because of the failure of molecular replacement selenomethionine-labelled eIF4G(572-853) and eIF4A(1-395) were produced in order to perform a MAD experiment. Three different complexes were crystallised incorporating (i) only labelled eIF4A, (ii) only labelled eIF4G, and (iii) labelled eIF4A and eIF4G. All of the different complexes crystallised under the same conditions as the native complex but none of them could be frozen. Several cryoprotectants were tested, crystals were transferred directly and serially. The following cryoprotectants were used: 5-20 % glycerol, 5-20 % glycerol with 10 mM AMP, 5-35 % MPD with 10 mM AMP, 5-40% PEG 400 with 10 mM AMP, Paratone N, silicone oil and paraffin oil.

The complex eIF4G(572-853)/eIF4A(9-395)

The eIF4A in this complex lacks the first eight N-terminal residues which are not conserved and not visible in the electron density of the N-terminal domain alone [23]. Removing of these flexible residues could lead to better ordered crystals.

Figure 8:*Crystals of the complex eIF4G(572-853)/eIF4A(9-395)*

eIF4G(572-853)/eIF4A(9-395) crystals (Figure 8) were obtained by hanging drop vapour diffusion at 277 K against a reservoir of 0.2 M potassium sodium tartrate tetrahydrate, pH 7.2, 20 % (w/v) PEG 3350 3 % (v/v) MPD.

All crystals were flash-frozen for data collection by soaking in cryoprotectant (mother liquor plus 35 % MPD) immediately prior to placement in a stream of cold N_2 gas (110 K). Crystals were measured at the X06SA beamline at the Swiss Light Source. They belong to space group $P4_{(1)}22$ or its enantiomorph $P4_{(3)}22$. Data-statistics are presented in Table 4.

Table 4: *Data-processing statistics merged from three different data sets.*

	eIF4G(572-853)/ eIF4A(9-395)
Space group	$P4_{(1)}22$ or $P4_{(3)}22$
Cell dimensions	a=b=67.70 Å, c=316.18 Å
Beamline	X06SA
Wavelength (Å)	0.9999
Resolution range[a] (Å)	20-3.77 (3.99-3.77)
No. measurements	46675
Unique reflections	13220
R_{sym}[b] (%)	9.9 (31.9)
$I/\sigma(I)$	11.4 (4.6)
Completeness (%)	92.3 (74.5)

[a] The values in parentheses of resolution range, R_{sym}, $I/\sigma(I)$ and completeness correspond to the outermost resolution shell
[b] $Rsym = \Sigma_{hkl}\Sigma_j|I(hkl;j) - \langle I(hkl)\rangle|/(\Sigma_{hkl}\Sigma_j\langle I(hkl)\rangle)$ where $I(hkl;j)$ is the jth measurement of the intensity of the unique reflection (hkl) and $\langle I(hkl)\rangle$ is the mean over all symmetry-related measurements.

Radiation damage was a serious problem during data collection. Either it was not possible to collect a complete data set from one crystal or we had to shift the crystal to another position to complete the data set.

The molecular replacement experiment failed. Neither the eIF4G nor the two domains of eIF4A could be located.

The complex eIF4G(572-853)/eIF4A(226-395)

The full-length structure of eIF4A consists of two compact domains connected by an extended 11-residue linker which is highly flexible [22]. We found that the C-terminal domain alone is sufficient to bind to eIF4G. By crystallizing the complex of eIF4G(572-853) and the C-terminal domain of eIF4A the flexibility of the complex might be reduced. We therefore crystallized the complex eIF4G(572-853)/eIF4A(226-395).
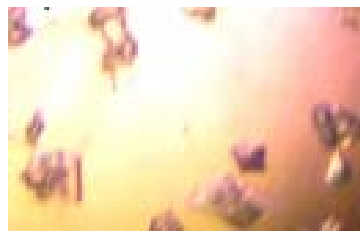
Figure 9: *Crystals of the complex eIF4G(572-853)/eIF4A(226-395) The size of the crystals were about 10 x 10 x 5 μm³.*

The eIF4G(572-853)/eIF4A(226-395) crystals (Figure 9) were obtained by hanging drop vapour diffusion at 277 K against a reservoir of 0.2 M tri-ammonium citrate, pH 7.0, 20 % (w/v) PEG 3350, 50 mM β-ME. We obtained those crystals in the presence of β-ME and other reducing agents (Detergent Screen I-III from Hampton Research). Reducing agents can suppress the formation of multimers via oxidative S-S-bond formation.

Prior to freezing in liquid nitrogen, the crystals were soaked in mother liquor containing 20 % glycerol. Crystals diffracted to 5 Å (in-house at 110 K). At the X06SA beamline at the Swiss Light Source crystals diffracted to 2 Å. The diffraction pattern indicated multiple crystals and could not be indexed.

To improve the crystal quality, we started to mutate a cysteine (C249) on the surface of eIF4A. This might prevent aggregation of eIF4A(226-395) possibly caused by the formation of disulfide bridges of solvent exposed cysteins.

Discussion

Four different complexes between eIF4A and eIF4G were crystallised and analysed by X-ray crystallography, but none of them was sufficient to solve structure. Only the amino-terminal part of eIF4A has been localized by molecular replacement. An explanation of failure might be that the C-terminal domain of eIF4A and/or the middle domain of eIF4GII can undergo a conformational change during the complex formation.

The question remains whether 33% of sequence identity between the middle domain of human eIF4GII and the yeast eIF4G is sufficient to serve as a search model. Molecular replacement becomes more difficult as the model becomes less complete or shares less sequence identity.

We had enormous difficulty freezing the crystals of the complex eIF4G(572-853)/eIF4A(1-395) without any serious damages like cracks or disorder of crystal lattice. One of the most important factors affecting crystal and diffraction quality is growth solution purity [33-35] Impurities can have profound effects on crystal growth, producing reduced or increased solubility, suppressed or enhanced nucleation, and changes in growth habit and morphology, and are often responsible for the irreproducibility of Crystallisation experiments [36-48]. They can significantly degrade crystal quality, causing dislocations and cracks, formation of twins and polycrystalline and amorphous aggregates, and increased mosaicity and decreased diffraction resolution [37, 40, 49].

The crystals of the complex eIF4G(572-853)/eIF4A(226-395) diffracted to 2 Å, but the diffraction pattern could not be indexed. We introduced mutants (cysteine 249 was changed to alanine or serine) to prevent aggregations through solvent exposed cystein residues. Possibly we can improve the crystal quality in the future.

In addition, we will attempt to improve the search models for molecular replacement one can use the program ElNémo, a Web-interface to the *Elastic Network Model* that generates a large

number of different starting models. ElNémo is a interface for NMA (normal mode analysis) which is a tool for predicting the possible movements of a given macromolecule [50].

References

1. Dever, T.E., Gene-specific regulation by general translation factors. *Cell*, 2002. **108**(4): p. 545-56.
2. Browning, K.S., The plant translational apparatus. *Plant Mol Biol*, 1996. **32**(1-2): p. 107-44.
3. Gradi, A., et al., A novel functional human eukaryotic translation initiation factor 4G. *Mol Cell Biol*, 1998. **18**(1): p. 334-42.
4. Goyer, C., et al., TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. *Mol Cell Biol*, 1993. **13**(8): p. 4860-74.
5. Tarun, S.Z., Jr. and A.B. Sachs, Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *Embo J*, 1996. **15**(24): p. 7168-77.
6. Imataka, H., A. Gradi, and N. Sonenberg, A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *Embo J*, 1998. **17**(24): p. 7480-9.
7. Mader, S., et al., The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4E-binding proteins. *Mol Cell Biol*, 1995. **15**(9): p. 4990-7.
8. Imataka, H. and N. Sonenberg, Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A. *Mol Cell Biol*, 1997. **17**(12): p. 6940-7.
9. Neff, C.L. and A.B. Sachs, Eukaryotic translation initiation factors 4G and 4A from Saccharomyces cerevisiae interact physically and functionally. *Mol Cell Biol*, 1999. **19**(8): p. 5557-64.
10. Dominguez, D., Altmann, M., Benz, J., Baumann, U. & Trachsel, H., Interaction of translation initiation factor eIF4G with eIF4A in the yeast Saccharomyces cerevisiae. *J Biol Chem*, 1999. **274**: p. 26720-26726.
11. Morino, S., et al., Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol Cell Biol*, 2000. **20**(2): p. 468-77.
12. Lamphear, B.J., et al., Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation. *J Biol Chem*, 1995. **270**(37): p. 21975-83.
13. Pyronnet, S., et al., Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *Embo J*, 1999. **18**(1): p. 270-9.
14. Marcotrigiano, J., et al., A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell*, 2001. **7**(1): p. 193-203.

15.   Andrade, M.A. and P. Bork, HEAT repeats in the Huntington's disease protein. *Nat Genet*, 1995. **11**(2): p. 115-6.

16.   Walker, J.E., et al., Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J*, 1982. **1**(8): p. 945-51.

17.   Linder, P., et al., Birth of the D-E-A-D box. *Nature*, 1989. **337**(6203): p. 121-2.

18.   Pause, A., N. Methot, and N. Sonenberg, The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol*, 1993. **13**(11): p. 6789-98.

19.   Pause, A. and N. Sonenberg, Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *Embo J*, 1992. **11**(7): p. 2643-54.

20.   Schmid, S.R. and P. Linder, Translation initiation factor 4A from Saccharomyces cerevisiae: analysis of residues conserved in the D-E-A-D family of RNA helicases. *Mol Cell Biol*, 1991. **11**(7): p. 3463-71.

21.   Pause, A.S., N, Helicases and RNA unwinding in translation. *Curr. Opin. Struct. Biol.*, 1993. **3**: p. 953-959.

22.   Caruthers, J.M., E.R. Johnson, and D.B. McKay, Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc Natl Acad Sci U S A*, 2000. **97**(24): p. 13080-5.

23.   Benz, J., H. Trachsel, and U. Baumann, Crystal structure of the ATPase domain of translation initiation factor 4A from Saccharomyces cerevisiae--the prototype of the DEAD box protein family. *Structure Fold Des*, 1999. **7**(6): p. 671-9.

24.   Johnson, E.R. and D.B. McKay, Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. *Rna*, 1999. **5**(12): p. 1526-34.

25.   Van Duyne, G.D., et al., Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*, 1993. **229**(1): p. 105-24.

26.   Kim, J.J.a.S.-H., Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.*, 1991. **24**: p. 409-411.

27.   Kabsch, W.e.a., Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.*, 1988. **21**: p. 67-71.

28.   Kabsch, W., Evaluation of single x-ray diffraction data from a position sensitive detector. *J. Appl. Cryst.*, 1988. **22**: p. 916-924.

29.   Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, 1993. **26**: p. 795-800.

30.   Vagin, A., Teplyakov, A., MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, 1997. **30**: p. 1022-1025.

31.   Jones, T.A., et al., Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, 1991. **47 (Pt 2)**: p. 110-9.

32.   Vaguine, A.A., J. Richelle, and S.J. Wodak, SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr*, 1999. **55 (Pt 1)**: p. 191-205.

33.   McPherson, A., Preparation and analysis of protein crystals. 1982: Malabar: Krieger. 372 p.

34.   Ducruix A, G.R., Crystallization of nucleic acids and proteins. 1992: Oxford: IRL. 331 p.

35.   Giegé R, L.B., Theobald-Dietrich A, Crystallogenesis of biological macromolecules - facts and perspectives. *Acta Crystallogr D*, 1994. **50**: p. 339-350.

36.   Wilson LJ, S.F., Control of solvent evaporation in hen egg-white lysozyme crystallization. *J Cryst Growth*, 1992. **116**: p. 414-420.

37.   Lorber B, S.M., Munch J-P, Giegé R, The influence of impurities on protein crystallization: the case of lysozyme. *J Cryst Growth*, 1993. **128**: p. 1203-1211.

38.   Skouri M, L.B., Giege R, Munch J-P, Candau JS, Effect of macromolecular impurities on lysozyme solubility and crystallizability: dynamic light scattering, phase diagram, and crystal growth studies. *J Cryst Growth*, 1995. **152**: p. 209-220.

39.   Ewing FL, F.E., van der Woerd M, Pusey ML, Effects of purification on the crystallization of lysozyme. *J Cryst Growth*, 1996. **160**: p. 389-397.

40.   Thomas BR, V.P., Rosenberger F, Heterogeneity determination and purification of commercial hen egg-white lysozyme. *Acta Crystallogr D*, 1996. **52**: p. 776-784.

41.   Abergel C, N.M., Fontecilla-Camps JC, The effect of protein contaminants on the crystallization of turkey egg white lysozyme. *J Cryst Growth*, 1991. **110**: p. 11-19.

42.   Forsythe E, E.F., Pusey ML, Studies of tetragonal lysozyme crystal-growth rates. *Acta Crystallogr D*, 1994. **50**: p. 614-619.

43.   Forsythe E, P.M., The effects of temperature and NaCl concentration on tetragonal lysozyme face growth-rates. *J Cryst Growth*, 1994. **139**: p. 89-94.

44.   Provost K, R.M., Crystal growth of lysozymes in media contaminated by parent molecules: influence of gelled media. *J Cryst Growth*, 1995. **156**: p. 112-120.

45.   Vekilov PG, M.L., Rosenberger F, Facet morphology response to nonuniformities in nutrient and impurity supply: I. Experiments and interpretation. *J Cryst Growth*, 1995. **156**: p. 267-278.

46.   Hirschler J, F.-C.J., Contaminant effects on protein crystal morphology in different growth environments. *Acta Crystallogr D*, 1996. **52**: p. 806-812.

47.   Hirschler J, F.-C.J., Protein crystal growth rates are face-specifically modified by structurally related contaminants. *J Cryst Growth*, 1997. **171**: p. 559-565.

48.   Vekilov PG, R.F., Dependence of lysozyme growth kinetics on step sources and impurities. *J Cryst Growth*, 1996. **158**: p. 540-551.

49.   Hirschler J, H.F., Forest E, Fontecilla-Camps JC, Contaminant inclusion into protein crystals analyzed by electrospray mass spectrometry and X-ray crystallography. *Protein Sci*, 1998. **7**: p. 185-192.

50.   Suhre, K. and Y.H. Sanejouand, On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallogr D Biol Crystallogr*, 2004. **60**(Pt 4): p. 796-9.

Chapter II (Manuscript submitted to *J. Biol. Chem.*)

Crystal structure of yeast Ypr118w, a methylthioribose-1-phosphate isomerase related to regulatory eIF2B subunits

Mario Bumann[1,2§] , Siamak Djafarzadeh[2§], Anselm Erich Oberholzer[1§], Peter Bigler[1], Michael Altmann[2], Hans Trachsel[2] & Ulrich Baumann[1*]

Affiliation:

[1]Department of Chemistry and Biochemistry, University of Berne, Freiestrasse 3, CH-3012 Switzerland

[2]Institute for Molecular Biology and Biochemistry, University of Berne Bühlstrasse 28, CH-3012 Switzerland

§ These three authors contributed equally to the work and are listed in alphabetical order.

* Corresponding author, e-mail: ulrich.baumann@ibc.unibe.ch, phone + 41 31 6314320, fax + 41 31 631 4887

Running Title: Structure and Function of Ypr118w

Key Words: eIF2B, methionine salvage pathway, translation initiation, crystal structure

Summary

Ypr118w is a non-essential, low-copy number gene product from *Saccharomyces cerevisiae*. It belongs to the PFAM family PF01008, which contains the α−, β-, and δ-subunits of eukaryotic translation initation factor eIF2B, as well as proteins of unknown function from all three kingdoms. Recently, one of those latter proteins from *Bacillus subtilis* has been characterized as a 5-methylthioribose-1-phosphate isomerase, an enzyme of the methionine salvage pathway. We report here the crystal structure of Ypr118w which reveals a dimeric protein with two domains and a putative active site cleft. The C-terminal domain resembles ribose-5-phosphate isomerase (RpiA) from *Escherichia coli* with a similar location of the active site. In vivo, Ypr118w protein is required for yeast cells to grow on methylthioadenosine in the absence of methionine showing that Ypr118w is involved in the methionine salvage pathway. The crystal structure of Ypr118w reveals for the first time the fold of a PF01008 member and allows a deeper discussion of an enzyme of the methionine salvage pathway, which has in the past attracted interest due to tumor suppression and as target of aniprotozoal drugs.

Introduction

The *Saccharomyces cerevisiae* gene *YPR118W* is a non-essential gene on chromosome 16 encoding an acidic protein (pI 4.89) of 411 amino acids, Ypr118w, whose function is unknown (1) and which is of rather low abundance (922 molecules per cell (2). Ypr118w belongs to the PFAM family PF01008 and the TIGR 00512 and 00524 families, the latter also being called eIF2B-related (eIF2B_rel) (3-5) . Members of PF01008 contain the α, ß and δ – but not the catalytically active ε - subunit of eukaryotic translation initiation factor 2B (eIF2B) from yeast and mammals (4,5) (Figure 1). eIF2B is an important regulator of translation initiation. In eukaryotic translation initiation, a ternary complex consisting of Met-tRNA$_i$, GTP and the heterotrimeric initiation factor eIF2 (6,7) associates with the 40S ribosomal subunit together with other initiation factors to form the 43S pre-initiation complex. This complex binds close to the 5'-end of mRNA and scans it in the 5' to 3' direction to localize the AUG initiation codon. AUG-recognition is mediated by codon-anticodon interaction and involves GTP hydrolysis stimulated by eIF5. This is a prerequisite for 80S initiation complex formation, *i.e.* the joining of the large ribosomal subunit and the release of initiation factors bound to the 40S ribosomal subunit. After its release the eIF2-GDP complex is bound by eIF2B which catalyzes the exchange of GDP for GTP. The recycling of eIF2 is modulated by posttranslational modifications of eIF2B (8,9) and by phosphorylation of the alpha subunit of eIF2 (10,11). Phosphorylated eIF2 binds 150 times more strongly to eIF2B than the non-phosphorylated form and therefore acts as a competitive inhibitor in the nucleotide exchange reaction, which consequently leads to inhibition of translation initiation (12).

Beside the well-characterized eIF2B subunits, the PF01008 family contains a subfamily of proteins (eIF2_rel) of mainly unknown function in eukaryotes, archaeae (e.g., *Pyrococcus*

furiosus) or eubacteria like *Bacillus subtilis* and *Thermotoga maritima*. Many of these proteins were initially annotated as putative translation initiation factors, despite the fact that there is no evidence for the requirement of an IF2 recycling factor in prokaryotic translation initiation. Recently, one of these proteins from B*acillus subtilis*, Swiss-Prot entry MTNA_BACSU, has been functionally characterized as a 5-methylthioribose-1-phosphate isomerase (MTNA) (13). This enzyme participates in the methionine salvage pathway catalyzing the isomerisation of 5-methylthioribose-1-phosphate to 5-methylthioribulose-1-phosphate (14). The methionine salvage pathway leads to the synthesis of methionine from methylthioadenosin (for a scheme; see Figure 2A), the end-product of the spermidine and spermine anabolism in many species including mammals (15), *Trypanosoma brucei* (15), *Saccharomyces cerevisiae* (16), *Klebsiella pneumoniae* (17) and *Bacillus subtilis* (14).

The methionine salvage pathway has attracted much attention due to the finding that methylthioadenosine phosphorylase (MTAP), the first enzyme in the pathway, is deleted in a variety of human tumors (15) and acts as a tumor suppressor gene (18). This was attributed to the fact that products of the methionine salvage pathway regulate polyamine biosynthesis by inhibiting the activity of ornithine decarboxylase, an enzyme found to be upregulated in a variety of human and animal tumors (19). Furthermore, the methionine salvage pathway is a target for antimalarian drugs, as *Plasmodium falciparum* was shown to harbour a 5-methylthioadenosine phosphorylase which is inhibited by methylthioadenosine analogs in μM concentrations leading to growth inhibition (20).

Ypr118w shares 37 % sequence identity with the established 5-methylthioribose-1-phosphate isomerase from *Bacillus subtilis* and 26 - 28 % identity with the well-characterized eIF2B α-subunits from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. These findings raise the question of whether Ypr118w is a 5-methylthioribose-1-phosphate isomerase or a translation initiation factor. We solved the crystal structure of Ypr118w and present genetic

evidence showing that Ypr118w is the yeast *Saccharomyces cerevisiae* 5-methylthioribose-1-phosphate isomerase.

Experimental Procedures

Yeast strain

Yeast BY4741 wild type and derivative strain Δypr118:: kanX[R] (Euroscarf collection) were used. The genotype of these strains is: MATa Δhis3 Δleu2 Δmet15 Δura3.

Plasmid construct

The YPR118w open reading frame was PCR-amplified from the start ATG to the stop codon from total yeast genomic DNA using primers to introduce a 5' SmaI and a 3' SacI site and subcloned into the multiple cloning site of yeast shuttle vector p301HIS3 under the control of the *GAL1/10* promoter resulting in the construct p301HIS3-YPR118.

Cell growth on MTA

Yeast cells were transformed with p301HIS3-YPR118 or p301HIS3 vector (negative control). Precultures of transformed cells were grown in SD or SDGal in the presence of all auxotrophic markers. Cells were washed with water and used to inoculate SD- or SDGal-cultures containing leucine, uracil and methionine (all at 20 μg/ml) or 5 mM MTA. Growth was monitored by measuring the cell density (A_{600}) after 40 hours of incubation at 30°C.

Protein expression and purification

DNA encoding full length *YPR118W* was amplified by PCR from total yeast genomic DNA using primers designed to introduce a 5' *SalI* site and a 3' *Not1* site. The PCR product was digested and inserted into pGEX-6P3 (Amersham Biosciences) expression vector. The

glutathione S-transferase (GST)-fusion protein was expressed in BL21-CodonPlus™(DE3)-RIL cells (Stratagene). Cells were grown to an A_{600} of 0.9 in minimal media at 37°C and expression was induced by 1mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 3hr induction, cells were harvested and resuspended in Tris-buffered saline (100mM NaCl and 20mM Tris-HCl, pH 7.5) and disrupted using a French Press. After centrifugation, batch purification of GST-Ypr118w was performed by incubating the extract with glutathione Sepharose 4B beads (Amersham Bioscienses) as described by the manufacturer, followed by three washes with TBS. The protein was eluted from the beads using reduced glutathione and the GST moiety was cleaved with Precission protease (Amersham Biosciences) at 4°C for 16hr. The protein was purified by gel filtration using a Superdex75 (Amersham) column equilibrated in 100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 1mM EDTA and 1 mM DTT. The protein fraction was pooled, reincubated with glutathione Sepharose 4B beads and loaded onto a 1 ml Resource Q™ column (Amersham Biosciences), equilibrated with 20 mM Tris-HCl, pH 8.0. Bound protein was eluted with a linear NaCl gradient (0-0.5 M) in the same buffer. All purification steps were monitored by SDS-PAGE analysis (21). The protein was desalted by ultrafiltration using a Centriprep-10 (Millipore) with 5 mM Tris-HCl, pH 7.5 and concentrated to a final concentration of 10 mg/ml. Mass spectrometric analysis was carried out by the Analytical Research Services of the Chemistry Departement, University of Berne. Selenomethionine (SeMet)-labeled protein was prepared by the method of methionine biosynthesis inhibition (22) and purified following the same protocol as for wild-type protein.

Crystallization

Crystals of Ypr118w and (SeMet)-Ypr118w were obtained within 4 days by the sitting-drop vapour-diffusion method. Drops were set up by mixing 2 μl of protein solution (10 mg/ml) with 2 μl of reservoir solution (200 mM ammonium sulfate, 25 % PEG 3350, Bis-Tris. pH 5.5) and equilibrated against 120 μl reservoir solution at 291 K. Typical crystals were needle-

shaped with an average size of $30 \times 30 \times 300 \ \mu m^3$. They belong to the orthorhombic spacegroup $P2_12_12_1$ with cell parameters of a=60.7 Å, b= 105.6 Å, c=263.3 Å and contain four monomers per asymmetric unit.

Data collection

Before data collection crystals were flash-cooled in a nitrogen stream at 110 K after raising the glycerol concentration of the crystallization solution to 15 % (v/v). Data were collected in a MAD experiment on beamline BW7A at the EMBL Hamburg Outstation at the DESY at 100 K, employing a MAR-165 CCD (MAR X-ray-research, Hamburg, Germany) detector. A second MAD experiment was performed at the Grenoble ESRF MAD beamline BM14 at 100 K, employing a Mosaic 225 CCD (MAR X-ray-research, Hamburg, Germany) detector. Due to the alignment of the crystals, only 0.2 degree/frame could be collected. Exposure times varied from about 30 seconds (Hamburg) to about 5 seconds (Grenoble). All datasets were integrated and scaled with XDS (23,24).

Structure solution, refinement and analysis

The structure was solved by MAD using selenomethionine-labeled protein. The Hamburg as well as the Grenoble MAD datasets were sufficient to solve the structure, both yielding clean electron density maps. However, the Grenoble dataset extended to higher resolution due to the brighter beam and larger detector and was therefore later used to obtain an experimental density map suitable for automated model building.

Sixteen out of twenty expected selenium positions were determined by SHELXD (25). Phases were computed using SOLVE (26) and RESOLVE (27). Further density modification, phase extension and automatic model building was done by ArpWarp (28) version 6.0. Refinement was effected using REFMAC (29) version 5.1.24, for model rebuilding the program O (30) version 9.0.3 was used.

Data collection and refinement statistics are given in Table I.

Sequence alignment was done using CLUSTALW (31), Figure 1 was prepared using ESPRIPT (32). Comparison of three-dimensional structures was effected using DALI (33). Structure figures were created using the program PYMOL (www.pymol.org).

Results

Ypr118w is a 5-methylthioribose-1-phosphate isomerase

The sequence similarity between Ypr118w and 5-methylthioribose-1-phosphate isomerase from *Bacillus subtilis* led us to test whether Ypr118 w is involved in the methionine salvage pathway. In this pathway, methylthioadenosine (MTA), a byproduct of the polyamine biosynthetic pathway, is recycled in cells by reconversion to methionine. This requires several enzymatic steps. The first step is the conversion of MTA to 5-methylthioribose-1-phosphate (MTR) by MTA-phosphorylase followed by its isomerization to methylthioribulose-1-P (Figure 2A). Recent studies led to the identification of some genes involved in the methionine salvage pathway in *Saccharomyces cerevisiae*. Among those, *Meu1* was identified as encoding the enzyme MTA-phosphorylase (16), whereas the gene encoding the MTR-isomerase could so far not be identified.

In order to find out whether Ypr118w is the MTR isomerase we tested whether a Δypr118w-knockout yeast strain auxotrophic for methionine was able to grow in a synthetic medium lacking methionine when supplemented with MTA. As opposed to the isogenic wild-type strain BY4741 (not shown) the Δypr118-knockout strain was not able to grow on MTA, neither in the presence of glucose nor of galactose (2%) as carbon source, but it grew well in the presence of methionine (Figure 2B). We then transformed the Δypr118-knockout strain with a plasmid carrying the gene *YPR118W* under the control of the galactose-inducible

GAL1/10 promoter. Cells transformed with this construct (but not with the control vector) were able to grow on MTA in the presence of galactose (Figure 2B) but not of glucose, clearly demonstrating that expression of Ypr118w was required to convert MTA into methionine. This indicates that *YPR118W* encodes for an enzyme of the methionine salvage pathway.

We also wanted to proof directly the isomerization of methylthioribose-1-phosphate to methylthioribulose-1-phosphate by Ypr118w. As methylthioribose-1-phosphate is not comercially available, we performed a two-step enzymatic reaction using recombinant Meu1p and 5-thiomethyladenosine to produce methylthioribose-1-phosphate. In a second step Ypr118w was added to the reaction mixture. We could readily detect the formation of a compound that was positive in an assay for reducing sugars (34). The NMR spectra of the expected compounds 5-methylthioribose-1-phosphate and 5-methylthioribulose-1-phsophate were identified (17), though the coupled reaction leading to 5-methylthioribulose-1-phosphate was not very efficient (not shown).

Monomer structure

The current model contains residues 1 to 407 without the amino acids 212 to 220 which are located in a disordered surface loop in all four monomers. This segment contains a unique and large insertion in the PF01008 family. The four crystallographically independent monomers do not show any significant deviation from each other, the RMS deviation of the C^{α}-positions is about 0.3 Å while single exposed surface residues show deviations of up to 1.4 Å.

The Ypr118w monomer structure has dimensions of 55 x 55 x 73 Å and can be formally divided into two domains (Figure 3A). The N-terminal 138 amino acids fold into a three-stranded antiparallel β-sheet followed by 6 α-helices with helices 1, 2, 4 and 5 building the core of this domain. Helix 5 is 38 amino acids long and bridges the N- and C-terminal

domains. This helix is kinked in the middle by about 15 degrees. The pollen allergen phl p 6 (PDB entry 1NLX) and the ATP synthase subunit C (PDB entry 1C17) were identified as closest structural homologs of the N-terminal domain with a DALI Z-score of 5.0, where 77 or 91 amino acids where aligned with a RMS deviation of 3.2 and 4.5 Å, respectively. The C-terminal domain possesses an αβα-fold resembling ribose-5-phosphate isomerase (RpiA) from *Escherichia coli* (PDB entry 1LKZ, DALI Z-score 11.6, 158 amino acids aligned with an RMS deviation of 3.2 Å) (35), and glutaconate coenzyme A-transferase from *Acidaminococcus fermentans* (PDB entry 1POI, DALI-score 8.4) (36). This domain contains a central six-stranded, mostly parallel, β sheet, surrounded by four α-helices and the C-terminal part of helix 6. The core resembles the well-known Rossman-type fold with three βαβ-units.

An "arm", consisting of residues 341 -367 wraps around the molecule with a short alpha-helix and two beta strands connected by a loop forming the tip ("hand") of this segment (residues 338 to 372). This arm is rather unique in the whole family PF01008.

Ypr118w is a dimer

Gel filtration experiments indicate that Ypr118w is a dimer in solution (not shown). This was further corroborated by the crystal structure: the asymmetric unit contains two dimers which are identical despite rather different crystal contacts. A total of 5,379 $Å^2$ (30%) of accessible surface gets buried upon dimer formation, a surface much larger than a usual crystal contact. The dimer has a "butterfly" shape (Figure 3B). The interactions between the monomers are mediated by the C-terminal domains, especially via the "arm" and "hand" and neighbouring segments (residues 335 to 383). Residues 231 to 260 including helices 7 and 8 also contribute to the interface.

Putative active site

A sulfate ion originating from the crystallization solution is deeply bound inside the cleft between the N- and C-terminal domains. It is coordinated via hydrogen bonds to the sidechains of Arg[51], Arg[106], Gln[237] and Lys[291] (Figure 4). The sidechain of Arg[51] is anchored via hydrogen bonds to the mainchain carbonyl oxygen and to the sidechain of Asn[182]. These amino acids are highly, though not absolutely conserved. They vary especially in the case of bona fide translation initiation factors which supports the hypothesis that Ypr118w has a function different from translation initiation. The sulfate ion is located solvent-inaccessible deeply inside of the protein in a rather large cavity (volume 110 Å^3). We interpret the observation of such a tightly bound sulfate ion in all four crystallographically independent monomers as a putative substrate binding site, where a phosphate would occupy the position of the sulfate, as has been observed frequently. The cavity surrounding the sulfate binding site is in principle large enough to harbour a phosphorylated sugar molecule. Interestingly, after superposition the putative phosphate binding site is close to the active centre of ribose-5-phosphate isomerase RpiA (Figure 5) which further supports the hypothesis that these enzymes have evolved from a common ancestor.

Conserved sequence motifs

Kyrpides and Woese, who first described the family PF01008 (3), defined six conserved sequence motifs within this family (Figure 1). Nearly all of them are located close to the putative active site assigned to the sulfate binding cavity (Figure1, Figure 3B and Figure 4).

Motif 1 includes amino acids 51 to 57. Arg[51] is located at the start of helix 2 and binds probably to the phosphate moiety of the substrate.

Motif 2, which is located mainly in the turn between helix 3 and helix 4, comprises residues 105 -110. Again, Arg[106] is likely to be involved in binding to a phosphate moiety.

Motif 3 includes amino acids 181 to 189 and possesses a cysteine (Cys[181]) which is absolutely conserved in the MNTA (eIF2_rel) subfamily of PF01008, but is absent in eukaryotic eIF2B subunits. A cysteine is also present in the active site of some non-homologous ribose-5-phosphate isomerases, e.g. in *Escherichia coli* RpiB, where it may act as catalytic base (37).

Motif 4 includes residues 231 to 240 and is located between ß-strand 5 and the beginning of helix 7. This motif contains Gln[237] which coordinates to the sulfate ion and is replaced by serine or glutamic acid in most eIF2B subunits. Pro[234] is in the *cis*-conformation and absolutely invariant as is Gly[238], as even an alanine residue in this position would lead to steric hinderance. Motif 4 is also involved in the dimer interface. Two inter-subunit salt bridges related by the dimer twofold axis are formed by the Glu[231] – Arg[233] and Arg[233] – Glu[231] pairs. These two amino acids are also mostly conserved in the eIF2B subunits, although they are present as monomers in the pentameric eIF2-B holo-complex. The absence of the "arm" (residues 341-367), which is crucially participating in dimer formation, may explain the monomeric nature of eIF2B subunits.

Motif 5 covers amino acids 275 to 293. Lys[291] is one of the putative phosphate ligands and is conservatively substituted by arginine in the mouse and human eIF2-δ subunit but is a leucine in the α-subunit from yeast. The invariant glycine residues exhibit backbone dihedral angles which are energetically unfavourable for other amino acids. Thr[294] forms hydrogen bonds with the sidechain of Asn[290] and the mainchain carbonyl oxygen of Lys[291]. This motif also consists of Asp[280], a residue not conserved in eIF2B subunits where it is replaced by glutamate or histidine. Asp[280] is completely buried and lines the putative active site pocket where it forms hydrogen bonds to a buried water molecule but does not contact the sulfate ion (Figure 4). *Escherichia coli* RpiA carries at this position a conserved aspartic acid (Asp[84]) which binds to the 3'-OH of the arabinose moiety of an inhibitor bound to the active site (38) . Replacement of Asp[84] reduces activity by a factor of 250 with only little change of the K_M

value. This aspartic acid is located in a five-amino acid consensus sequence G A **D** R/E I/V which is conserved in RpiA and eIF2B_rel and seems to be the only residue in the active centre which is invariant in RpiA and the eIF2B_rel family (Figure 5).

Motif 6 comprises amino acids 377 to 396. Phe^{379} lines the active site cavity (Figure 4) and is close to the SO_4^{2-} ion and to Arg^{106} from motif 2. The absolutely conserved Asp^{380} is completely buried and forms two hydrogen bonds with the mainchain amide groups from Ile^{293} and Leu^{295}. The C-terminal end of the motif is located on the surface of the molecule. The reason for the absolute conservation of Thr^{392} is not clear.

Discussion

Our genetic experiments (Figure 2) together with the structural data indicate that the gene *YPR118W* encodes the yeast *Saccharomyces cerevisiae* 5-methylthioribose-1-phosphate isomerase, a protein required for the methionine salvage pathway : deletion of the gene *YPR118W* results in the inability of yeast cells to grow on methylthioadenosine (MTA) in the absence of methionine. Transformation of these cells with a plasmid carrying *YPR118W* restores growth on MTA in the absence of methionine. Additionally, we could detect *in vitro* the formation of 5-methylthioribulose-1-phosphate, the isomerization product catalysed by Ypr118w.

With regard to the sequence homology (37% identity) between Ypr118w and the recently characterized *Bacillus subtilis* 5-thiomethylribose-1-phosphate isomerase (13) we conclude that Ypr118w is indeed the yeast ortholog. There is no sequence similarity to any other enzyme of the methionine salvage pathway. Interestingly, the threedimensional structure

resembles ribose-1-phosphate isomerase A, which shows a hitherto unnoted lineage between different aldo-keto isomerases.

Also at the level of primary sequence Ypr118w resembles more the MTNA-subfamily of PF01008 than the eIF2B-αβδ subfamily. Especially pronounced are the differences at position 181 which is for all known and putative MTNAses a potentially catalytically active cysteine, while this residue is missing in all established eIF2B subunits. Similarly, Asp^{280}, a catalytically quite important residue is absolutely conserved in RpiA and MTNAses but not in the eIF2B subunits.

The three-dimensional structure of Ypr118w reveals in its C-terminal domain similarity to ribose-5-phosphate isomerase RpiA, including the location of the putative active site and the conservation of a catalytically important aspartic acid. This is in agreement with the function of an aldose-ketose isomerase.

It is currently not clear whether all members of the eIF2B_rel family, *i.e.* those PF01008 members which are not translation initiation factor subunits, are indeed methylthioribose-1-phosphate isomerases. The Swiss-Prot entry E2B_AQUAE, a homolog from *Aquifex aeolicus*, has an N-terminal extension which codes for a thioesterase domain (PFAM PF03061). As mentioned above, the fold of Ypr118w is also related to glutaconate coenzyme a-transferase, an enzyme which catalyses the transfer of CoA from acetyl-CoA to the 1-carboxylate of glutaconate. This could hint at another function since a thioesterase-activity is not obviously required in the methionine salvage pathway. However, from the sequence alignment E2B_AQUAE fits seamlessly into the methylthioribose-1-phosphate isomerase branch of PF01008. Striking is the similarity between Ypr118w and the TREMBL entry Q9BV20. The latter is a human protein which is annotated as GTP-binding and involved in translation initiation. We assume that it is actually the Ypr118w ortholog. This protein is of potential

interest in cancer research: as mentioned above, ornithine decarboxylaseactivity is downregulated by the production of downstream metabolites of the methionine salvage pathway (16), and MTA, the first enzyme in the pathway, is a tumor suppressor. It will be interesting to examine the effect of the downstream enzymes on tumorgenesis.

Another open question is also whether the two homologous subfamilies (eIF2B-αβδ and eIF2B_rel) of PF01008 share some biological function. Binding of GTP and GDP to *Saccharomyces cerevisiae* eIF2B was established with binding constants of about 1 μM (39), however, the catalytic activity resides in the non-homologous ε-subunit. Therefore it seems to be likely that GTP or GDP binds to the ε-subunit, and not to the αβδ subunits. There are two other phosphate moieties involved in the biological function of eIF2B. First, eIF2B binds strongly to eIF2-α(P-51), the phosphorylated form of eIF2α. It was shown that in *Saccharomyces cerevisiae* the GCD2-GCD7-GCN3 subcomplex (*i.e.* the regulatory αβδ-subcomplex) mediates regulation of eIF2B activity in response to eIF2-α phosphorylation (40,41) while the γε-complex promoted release of GDP from eIF2 at a higher rate than the ε-subunit alone, but stable binding to eIF2 was not affected by eIF2-α phosphorylation (41). Hence the sulfate binding site determined by us in the Ypr118w structure could represent the eIF2-α-(Ser-51)-phosphate binding site in the eIF2B regulatory subunits. Second, it was reported that glucose-6-phosphate is required to maintain the activity of eIF2B by a hitherto unknown mechanism (42). Regarding the homology to RpiA and MTNA it is tempting to speculate that α, ß and δ subunits of eIF2B bind glucose-6-phosphate.

Apparently the ancestral protein possessed a sugar-phosphate binding/processing activity from which the two subfamilies of PF01008 and the RpiAs have emerged.

Acknowledgements

References

1. Bussey, H., Storms, R. K., Ahmed, A., Albermann, K., Allen, E., Ansorge, W., Araujo, R., Aparicio, A., Barrell, B., Badcock, K., Benes, V., Botstein, D., Bowman, S., Bruckner, M., Carpenter, J., Cherry, J. M., Chung, E., Churcher, C., Coster, F., Davis, K., Davis, R. W., Dietrich, F. S., Delius, H., DiPaolo, T., Hani, J., and et al. (1997) *Nature* **387**, 103-105

2. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) *Nature* **425**, 737-741

3. Kyrpides, N. C., and Woese, C. R. (1998) *Proc Natl Acad Sci U S A* **95**, 3726-3730

4. Price, N. T., Francia, G., Hall, L., and Proud, C. G. (1994) *Biochim Biophys Acta* **1217**, 207-210

5. Cigan, A. M., Bushman, J. L., Boal, T. R., and Hinnebusch, A. G. (1993) *Proc Natl Acad Sci U S A* **90**, 5350-5354

6. Kimball, S. R. (1999) *Int J Biochem Cell Biol* **31**, 25-29

7. Merrick, W. C. (1992) *Microbiol Rev* **56**, 291-315

8. Wang, X., Paulin, F. E., Campbell, L. E., Gomez, E., O'Brien, K., Morrice, N., and Proud, C. G. (2001) *Embo J* **20**, 4349-4359

9. Proud, C. G. (2001) *Prog Mol Subcell Biol* **26**, 95-114

10. Hinnebusch, A. G. (2000) in *Translational Control of Gene Expression* (Mathews, M. B., ed), pp. 185-243, Cold Spring Harbor Laboratory Press

11. Farrell, P. J., Balkow, K., Hunt, T., Jackson, R. J., and Trachsel, H. (1977) *Cell* **11**, 187-200

12. Dever, T. E., Yang, W., Astrom, S., Bystrom, A. S., and Hinnebusch, A. G. (1995) *Mol Cell Biol* **15**, 6351-6363

13. Ashida, H., Saito, Y., Kojima, C., Kobayashi, K., Ogasawara, N., and Yokota, A. (2003) *Science* **302**, 286-290

14. Sekowska, A., and Danchin, A. (2002) *BMC Microbiol* **2**, 8

15. Riscoe, M. K., Ferro, A. J., and Fitchen, J. H. (1988) *Antimicrob Agents Chemother* **32**, 1904-1906

16. Subhi, A. L., Diegelman, P., Porter, C. W., Tang, B., Lu, Z. J., Markham, G. D., and Kruger, W. D. (2003) *J Biol Chem* **278**, 49868-49873

17. Furfine, E. S., and Abeles, R. H. (1988) *J Biol Chem* **263**, 9598-9606
18. Christopher, S. A., Diegelman, P., Porter, C. W., and Kruger, W. D. (2002) *Cancer Res* **62**, 6639-6644
19. Auvinen, M., Paasinen, A., Andersson, L. C., and Holtta, E. (1992) *Nature* **360**, 355-358
20. Sufrin, J. R., Meshnick, S. R., Spiess, A. J., Garofalo-Hannan, J., Pan, X. Q., and Bacchi, C. J. (1995) *Antimicrob Agents Chemother* **39**, 2511-2515
21. Laemmli, U. K. (1970) *Nature* **227**, 680-685
22. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J Mol Biol* **229**, 105-124
23. Kabsch, W. (1988) *J. Appl. Cryst.* **22**, 916-924
24. Kabsch, W. (1993) *J. Appl. Cryst.* **26**, 795-800
25. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779
26. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55** (**Pt 4**), 849-861
27. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55** (**Pt 2**), 501-505
28. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat Struct Biol* **6**, 458-463
29. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallog. sect. D* **53**, 240-255
30. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr A* **47** (**Pt 2**), 110-119
31. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res* **22**, 4673-4680
32. Gouet, P., Robert, X., and Courcelle, E. (2003) *Nucleic Acids Res* **31**, 3320-3323
33. Holm, L., and Sander, C. (1993) *J Mol Biol* **233**, 123-138
34. Avigad, G. (1975) *Methods Enzymol* **41**, 27-29
35. Rangarajan, E. S., Sivaraman, J., Matte, A., and Cygler, M. (2002) *Proteins* **48**, 737-740
36. Jacob, U., Mack, M., Clausen, T., Huber, R., Buckel, W., and Messerschmidt, A. (1997) *Structure* **5**, 415-426
37. Zhang, R. G., Andersson, C. E., Skarina, T., Evdokimova, E., Edwards, A. M., Joachimiak, A., Savchenko, A., and Mowbray, S. L. (2003) *J Mol Biol* **332**, 1083-1094
38. Zhang, R., Andersson, C. E., Savchenko, A., Skarina, T., Evdokimova, E., Beasley, S., Arrowsmith, C. H., Edwards, A. M., Joachimiak, A., and Mowbray, S. L. (2003) *Structure (Camb)* **11**, 31-42
39. Nika, J., Yang, W., Pavitt, G. D., Hinnebusch, A. G., and Hannig, E. M. (2000) *J Biol Chem* **275**, 26011-26017
40. Yang, W., and Hinnebusch, A. G. (1996) *Mol Cell Biol* **16**, 6603-6616
41. Pavitt, G. D., Ramaiah, K. V., Kimball, S. R., and Hinnebusch, A. G. (1998) *Genes Dev* **12**, 514-526
42. Gross, M., Rubino, M. S., and Starn, T. K. (1988) *J Biol Chem* **263**, 12486-12492
43. Kleywegt, G. J., and Jones, T. A. (1997) *Methods Enzymol.* **277**, 525-545

Figures and figure legends


```
                                             β1          β2            β3        α1
|Q06489|YP18_YEAST              TT        →       TT     →           →   0000000000      0000000000
                                1        10        20          30          40        50          60

1 |Q06489|YP18_YEAST     ..................MSLEAIVFDRSEPENVSVKVLDQLLLPYTTKYVPIHTIDDGYSVIKSMQVRGAPATAIVG
  |Q8T5K0                ..................MTLEAIKYKSG.....QLQILDQLLLPAESKYIPVAGVKDGWSAIHKMQVRGAPAIAIVG
  |Q9BV20                ..................MTLEAIRYSRG.....SLQILDQLLLPKQSRYEAVGSVHQAWEAIRAMKVRGAPAIALVG
  |O31662|MTNA_BACSU     ..............MTHSFAVPRSVEWKETAITILNQQKLPDETEYLELTTKEDVFDAIVTLKVRGAPAIGITA
  |Q819F2               ................MSTIVTVPRSVSWKGDAIAVLNQTKLPHSTEYKTLTTIEEVWKSIIMLEVRGAPAIGIVA
  |O67879|E2B_AQUAE_tru_ .GEDKVFEGEFFTVFLIDMFWRGEKMEVKPFYWKGDKLLLLDQRKLPHQEVWLELKTYEEVAKAIKEMAVRGAPAIGCTA
  |Q9X013|MTNA_THEMA    ..................MKLKTKTMEWSGNSLKLLDQRKLPFIEEYVECKTHEEVAHAIKEMIVRGAPAIGVAA
  |Q87A92|MTNA_XYLFT    ...........MHHPLPPDDTLYDQVRPILWTGHFLKLLDQRKLPFVVEYVECHSSEDVTQAIRTLIVRGAPAIGIVA
  |O58433|E2B1_PYRHO    .............MEIRYTPKELTKLPRTVEYKNKSVYMINQRLLPKEFKVEKFSKVEEVAEAIKNMTVRGAPAIGAAA
  |P74497|MTNA_SYNY3    ...........MTQSTNFDDVYPVQWRDDQVILIDQTRLPLEYKEVAIADYEAMGHAIKSMVVRGAPAIGVAA
2 |Q99LC8|E2BA_MOUSE    ...............................MEDGELIEYFKSQMKGDPKMASAVAAIQ
  |Q14232|E2BA_HUMAN    ...............................MDDKELIEYFKSQMKEDPDMASAVAAIR
  |Q64270|E2BA_RAT      ...............................MEDGELIKYFKSQMKGDPNMASAVAAIQ
  |P14741|E2BA_YEAST    ...........................MSEFNITETYLRFLEEDTEMTMPIAAIE
  |Q9UI10|E2BD_HUMAN    .LLLRRLVKKPERQQVPTRKDYGSKVSLFSHLPQYSRQNSLTQFMSIPSSVIHPAMVRLGLQYSQGLVSGSNARCIALLR
  |Q63186|E2BD_RAT      PTLLRRLLRKPDRQQVPTRKDYGSKVSLFSHLPQYSRQSSLTQYMSIPSSVIHPAMVRLGLQYSQGLVSGSNARCIALLH
  |Q61749|E2BD_MOUSE    .TLLRRLLRKPDRQQVPTRKDYGSKVSLFSHLPQYSRQSSLTQYMSIPSSVIHPAMVRLGLQYSQGLISGSNARCIALLH
```

```
                              α2              η1          α3                  α4
|Q06489|YP18_YEAST      0000000000000    000    00000000000000000     000000000000     0000000000
                              70        80        90          100         110        120          130

1 |Q06489|YP18_YEAST     SLSVLTEVQLIKHNPTSDVATLYSLVNWESTKTVLNKRLDFLLSSRPTAVNLSNSLVEIKNILK..SSSDLKAFDGSLYN
  |Q8T5K0                CLSLVVEIYD.KQYET..KAALANEIG........EHLQYLVTSRPTAVNLKLAADDVKGQVESLLANETVTVDGMKQE
  |Q9BV20                CLSLAVELQAGAGGPG..LAALVAFVR........DKLSFLVTARPTAVNMARAARDLADVAAREAEREGATEEAVRER
  |O31662|MTNA_BACSU     AFGLALAAKDIETDNV...TEFRRRLE........DIKQYLNSSRPTAINLSWALERLSHSVE...N..AISVNEAKTN
  |Q819F2               AFGLALAAKKYNTLHI...EEFQKKFN........RDCNYLGTSRPTAVNLFWAIDRMRESIR...E..ITTIKEAQKI
  |O67879|E2B_AQUAE_tru_ AYGFVLGVKVQKED..........PE........KVYETLKNTRPTAYNLFWALDRMMKALK..EGKD........
  |Q9X013|MTNA_THEMA    AFGYVLGLRDYKTGS.....LTDWMK........QVKETLARTRPTAVNLFWALNRMEKVFF...ENADR..ENLFEI
  |Q87A92|MTNA_XYLFT    GWGAVLAAREIEAVDG...IEALCKLE........PALQRLHAARPTAVNLAWVLARMRRTLS...AAHAD....WRQV
  |O58433|E2B1_PYRHO    GFGLALYAETSKAKTK..EEFLDGFE........KAYEILKNTRPTAVNLFWALNRIKKLVE...EHSEDPLDEIKRL
  |P74497|MTNA_SYNY3    AYGMYLGARRIQTTEL...REFVVQLE........FVADQLRQTRPTAVNLFWAIDQMLNVAY...HSGEN.VEQIKAN
2 |Q99LC8|E2BA_MOUSE    TLLEFLKRDKGETLQG.LRANLTYAIK........TLCG..VDSSVAVSSGGELFLRFISLTSL.....EYSDYSKCKKI
  |Q14232|E2BA_HUMAN    TLLEFLKRDKGETIQG.LRANLTSAIE........TLCG..VDSSVAVSSGGELFLRFISLASL.....EYSDYSKCKKI
  |Q64270|E2BA_RAT      TLLEFLKRDKGETIQG.LRAHLTKAIE........TLCA..VDSSVAVSSGGELFLRFISLTSL.....EYSDYSKCKKI
  |P14741|E2BA_YEAST    ALVTLLRIKTPETAAE.MINTIKSSTE........ELIKSIPNSVSLRAGCDIFMRFVLRNLH.....LYGDWENCKQH
  |Q9UI10|E2BD_HUMAN    ALQQVIQDYTTPPNEE..LSRDLVNKLK........PYMSFLTQCRPLSASMHNAIKFLNKEIT.....SVGSSSKREEEA
  |Q63186|E2BD_RAT      ALQQVIQDYTTPPNEE..LSRDLVNKLK........PYISFLTQCRPMSASMCNAIKFFNKEVT.....GMSSSKREEEA
  |Q61749|E2BD_MOUSE    ALQQVIQDYTTPPSEE..LSRDLVNKLK........PYISFLTQCRPMSASMCNAIKFLTKEVT.....GMSSSKREEEA
```

```
                           α5                            β4         η2              α6
|Q06489|YP18_YEAST      000  0000000000000000000000000000     00000000     0000000000000000
                          140        150          160         170         180         190         200        210

1 |Q06489|YP18_YEAST     YVC.......ELIDEDLANNMKMGDNGAKYLIDVLQKDGFKDEFAVLTICNTGSLATSGYGTALGVIRSLWKDSLAKTDK
  |Q8T5K0                AI.......EYMLEKDISDNRAIGDNGANVLVKGVDR.....PLKLLTHCNTGSLATAGYGTALGVIRSVNER.......
  |Q9BV20                VICC....TEDMLEKDLRDNRSIGDLGARHLLLERVAPS..GGKVTVLTHCNTGALATAGYGTALGVIRSLHSL.......
  |O31662|MTNA_BACSU     LVHE....AIQIQVEDEETCRLIGQNALQLFKKG......DRIMTICNAGSIATSRYGTALAPFYLAKQK.......
  |Q819F2               LEEE....ALRIQQEDEEVCRNIGEYALTCFKDG......DNILTICNAGSIATARYGTALAPFYIGKEK.......
  |O67879|E2B_AQUAE_tru_ IEEE....AKEIEKEDYEANKRMGEIGSEIVPLN......AKVLTHCNTGALATAGWGTALGVIRSAHYG.......
  |Q9X013|MTNA_THEMA    LENE....ALKMAYEDIEVNKAIGKNGAQLIKDG......STILTHCNAGALATVDYGTALGVIRAAVES.......
  |Q87A92|MTNA_XYLFT    MECE....AESIAREDLTANRCMGAYGAALIPIG......SGVLTHCNTGSLATAGFGTALGVIRDGIAQ.......
  |O58433|E2B1_PYRHO    IVQE....AYKIADEDVEANLRMGHYGAEVLPEG......NILTHCNAGSLATVHLGTVGSVVRVMHKD.......
  |P74497|MTNA_SYNY3    LLSQ....AQQIQLDDLRTCQAIGAAGLEVLPREPHQ.....LTILTHCNAGALATAGYGTAIGVIRAAWSA......
2 |Q99LC8|E2BA_MOUSE    MIER....GELFLRRISLSRNKIANLCHTFIKDG......ARILTHAYSRVVLRVLEEAVAA.......
  |Q14232|E2BA_HUMAN    MIER....GELFLRRISLSRNKIADLCHTFIKDG......ATILTHAYSRVVLRVLEEAVAA.......
  |Q64270|E2BA_RAT      MIER....GELFLSRISLSRTKIASLCHAFIKDG......ARILTHAYSRVVLRVLEEAVAA.......
  |P14741|E2BA_YEAST    LIEN....GQLFVSRAKKSRNKIAEIGVHFIADD......DILLVHGYSRAVFSLLNHAANK.......
  |Q9UI10|E2BD_HUMAN    KSELRAAIDRYVQEKIVLAAQAISRFAYQKISNG......DVILVYGCSSLVSRILQEAWTE.......
  |Q63186|E2BD_RAT      KSELKEAIDRYVQEKIVLASQAISRFASKKISDG......DVILVYGCSSLVSRILQEAWVE.......
  |Q61749|E2BD_MOUSE    KSELKEALDRYVQEKIVLAAQAISRFASTKISDG......DVILVYGCSSLVSRILQEARVE.......
```

```
                              β5           α7        α8              η3    α9          β6       β7       β8
|Q06489|YP18_YEAST      →              →     0000   00000000      0000000000      →        →      TT   →
                          220        230        240        250          260         270        280        290

1 |Q06489|YP18_YEAST     ADSGLDNEKCPRMGHVFPLETRPYNQGSRLTAYELVYDKIPSTLITDSSTAYRIRTSPIPIKAAFVGADRIVRNGDTANK
  |Q8T5K0                .........NLLEHVYCTETRPYNQGARLTAYELVHDKLPATLVTDSMVAALLNSR..KIDAIIVGADRVAANGDTANK
  |Q9BV20                .........GRLEHAFCTETRPYNQGARLTAFELVYEQIPATLITDSMVAAAMAHR..GVSAVVVGADRVVANGDTANK
  |O31662|MTNA_BACSU     .........DLGLHIYACETRPVLQGSRLTAWELMQGGIDVTLITDSMAAHTMKEK..QISAVIVGADRIAKNGDTANK
  |Q819F2               .........GVRLHAYACETRPYLQGRLTTWELKQANIDVTLITDNTAAHAIQTK..EINAIIVGADRIVANGDTANK
  |O67879|E2B_AQUAE_tru_ .........GKNIFVWVDETRPYLQGARLTAWELVKESIPHKIITDSTAGFLMKKG..MVDLVIVGADRITAKGDVANK
  |Q9X013|MTNA_THEMA    .........GKRIRVFADETRPYLQGARLTAWELMKDGIEVYVITDNMAGWLMKRG..LIDAVVVGADRIALNGDTANK
  |Q87A92|MTNA_XYLFT    .........GRIARVFVGETRPWLQGARLTVWELQQDGIDATLIADSAAAHLMKSG..QVQWVIVGADRICANGDTANK
  |O58433|E2B1_PYRHO    .........GSLKLLWLDETRPVLQGARLSAWEYSYDGLNVKLIADNAAAFVMQQG..FVDAIIVGADRIVANGDFANK
  |P74497|MTNA_SYNY3    .........GRLARVVADETRPRLQGAKLTVWECVQAGIPVTLITDNMAAHCIMQQQ..RIDAVVVGADRIARNGDIANK
2 |Q99LC8|E2BA_MOUSE    .........KKRFSVYITESQPDLSGKKMAKALSHLNVPVTVVLDAAVGYIMEK....ADLVIVGAEGVVENGGIINK
  |Q14232|E2BA_HUMAN    .........KKRFSVYVTESQPDLSGKKMAKALCHLNVPVTVVLDAAVGYIMEK....ADLVIVGAEGVVENGGIINK
  |Q64270|E2BA_RAT      .........KKRFSVYITESQPDLSGKKMAKALCHLNVPVTVVLDAAVGYIMEK....VDLVIVGAEGVVENGGIINK
  |P14741|E2BA_YEAST    .........FIRFRCVVTESRPSKQGNQLYTLLEQKGIPVTLIVDSAVGAVIDK....VDKVFVGAGHALLANGSGGIINL
  |Q9UI10|E2BD_HUMAN    .........GRRFRVVVVDSRPWLEGRHTLRSLVHAGVPASYLLIPAASYVLPE....VSKVLLGAHALLANGSVMSR
  |Q63186|E2BD_RAT      .........GRRFRVVVVDSRPRLEGRHMLHCLVRAGVPTSYLLIPAASYVLPE....VSKVLLGAHALLANGSVMSR
  |Q61749|E2BD_MOUSE    .........GRRFRVVVVDSRPRLEGRHMLHSLVRAGVPTSYLLIPAASYVLPE....VSKVLLGAHALLANGSVMSR
```

```
                              α10          β9     η4           η5            α11   β10        TTT          TT          β11
|Q06489|YP18_YEAST      0000000000    →     000     000         000          0000                                              TT
                          300        310        320        330          340        350        360          370

1 |Q06489|YP18_YEAST     IGTLQLAVICKQFGIKFFVVAPKTTIDNVTETGDDIIVEERNPEEFKVVTGTVINPENGSLILNESGEPITGKVGIAPLE
  |Q8T5K0                IGTYQMAVVAKHHGVFFYVAAPFTSIDVAIEDGSHIKIEERPEHELTHIGG..................QRIAAPG
  |Q9BV20                VGTYQLAIVAKHHGIPFYVAAPSSSCDLRLETGKEIIIEERPGQELTDVNG...................VRIAAPG
  |O31662|MTNA_BACSU     IGTYGLAILANAFDIPFFVAAPLSTFDTKVKCGADIPIEERPEEVRQISG..................VRTAPSN
  |Q819F2               IGTMNLAILAKYFNIPFYVAAPLSTFDVTKETGAEIIIEERDETEVTKIFG..................KQVAPIG
  |O67879|E2B_AQUAE_tru_ IGTYTLAVLCKEHNVPFYVAAPLSTDSNIKSGEEIIIEERSPEEVKVCGG..................CAIAPKE
  |Q9X013|MTNA_THEMA    IGTYSLAVLAKRNNIPFYVAAPVSTIDPTIRSGEEIPIEERRPEEVTHCGG..................NRIAPEG
  |Q87A92|MTNA_XYLFT    IGTYQLAITARHHGVKFMVVSANTVDMDTVAGEAIEIQERDPEELLGVSG..................VRTVAEG
  |O58433|E2B1_PYRHO    IGTYMLAVLAREHGIPFFAVAPLSSIDMELKSGKDIPIEERSPEEVLTCGG..................CRIAP.D
  |P74497|MTNA_SYNY3    IGTYGLAVLAKMHAIPFFVAVAPLSTVDFALDDGSQIPIEERDPVEIYQPEG..................NRITPEG
2 |Q99LC8|E2BA_MOUSE    IGTNQMAVCAKAQNKPFYVVAESFKFVR......LFPLNQEDVPDKFK.YKA....D...........TLKSVQTGQDLK
  |Q14232|E2BA_HUMAN    IGTNQMAVCAKAQNKPFYVVAESFKFVR......LFPLNQQDVPDKFK.YKA....D...........TLKVAQTGQDLK
  |Q64270|E2BA_RAT      IGTNQMAVCAKAQNKPFYVVAESFKFVR......LFPLNQQDVPDKFK.YKA....D...........TLKSVQAGQDLK
  |P14741|E2BA_YEAST    VGTYSVGVLAHNARKPFYVVTESHKFVR......MFPLSSDDLP....MAG....P...........PLDFTRRTDDLE
  |Q9UI10|E2BD_HUMAN    VGTAQLALVARAHNVPVLVCCETYKFCERVQTDAFVSNELDDPDDLQCKRG....E...........HVALANWQNHAS
  |Q63186|E2BD_RAT      VGTAQLALVARAHNVPVLVCCETYKFCERVQTDAFVSNELDDPDDLQCKRG....D...........QVTLANWQNNSS
  |Q61749|E2BD_MOUSE    VGTAQLALVARAHNVPVLVCCETYKFCERVQTDAFVSNELDDPDDLQCKRG....D...........QVALANWQSHPS
```

```
                              β12    η6    β13  β14                        η7
|Q06489|YP18_YEAST      →      000     →   TT    →              TT            000
                          380        390          400          410

1 |Q06489|YP18_YEAST     INVWNPAFDITPHELIDG.IITEEGVFTK...NSSGEFQLESLF............
  |Q8T5K0                IGCWNPAFDVTPAELITG.IITERGVLKP...CELAEKVRQK............
  |Q9BV20                IGVWNPAFDVTPHDLITGGIITELGVFAP...EELRTALTTTISSRDGTLDGPQM
  |O31662|MTNA_BACSU     VPVFNPAFDITPHDLISG.IITEKGIMTGNYEEEIEQLFKGEKVH............
  |Q819F2               TDVYNPAFDITPNELITG.IITEKGIIRGDYKREIASLFEKQANT...........
  |O67879|E2B_AQUAE_tru_ SDALHLAFDITPAELITG.IITEKGIER.ERITEALKD............
  |Q9X013|MTNA_THEMA    VKVLNPAFDVTENTLITA.IITEKGVIRPPFEENIKKILEV............
  |Q87A92|MTNA_XYLFT    IAAWNPVFDVTPGALIDA.IVTEKGVIQSPDAAQMRATFGN............
  |O58433|E2B1_PYRHO    VPVYNPAFDVTPHKYLTG.IITDRGVVWPPFKRNLKKLFEVNKSGGDEAV.....
  |P74497|MTNA_SYNY3    AEFYNPAFDVTPASLITA.IITEQGAIAP..EKLADLQA............
2 |Q99LC8|E2BA_MOUSE    EEH..PWVDYTSPSLITL.LFTDLGVLTP...SAVSDELIKLYL............
  |Q14232|E2BA_HUMAN    EEH..PWVDYTAPSLITL.LFTDLGVLTP...SAVSDELIKLYL............
  |Q64270|E2BA_RAT      EEH..PWVDYTS..LITL.LFTDLGVLTP...SAVSDELIKLYL............
  |P14741|E2BA_YEAST    DALRGPTIDYTAQEYITA.LITDLGVLTP...SAVSEELIKMWYD...........
  |Q9UI10|E2BD_HUMAN    LRLLNLVYDVTPPELVDL.VITELGMIPC...SSVPVVLRVKSSDQ...........
  |Q63186|E2BD_RAT      LRLLNLVYDVTPPELVDL.VITELGMIPC...SSVPVVLRVKSSDQ...........
  |Q61749|E2BD_MOUSE    LRLLNLVYDVTPPELVDL.VITELGMIPC...SSVPVVLRVKSSDQ...........
```
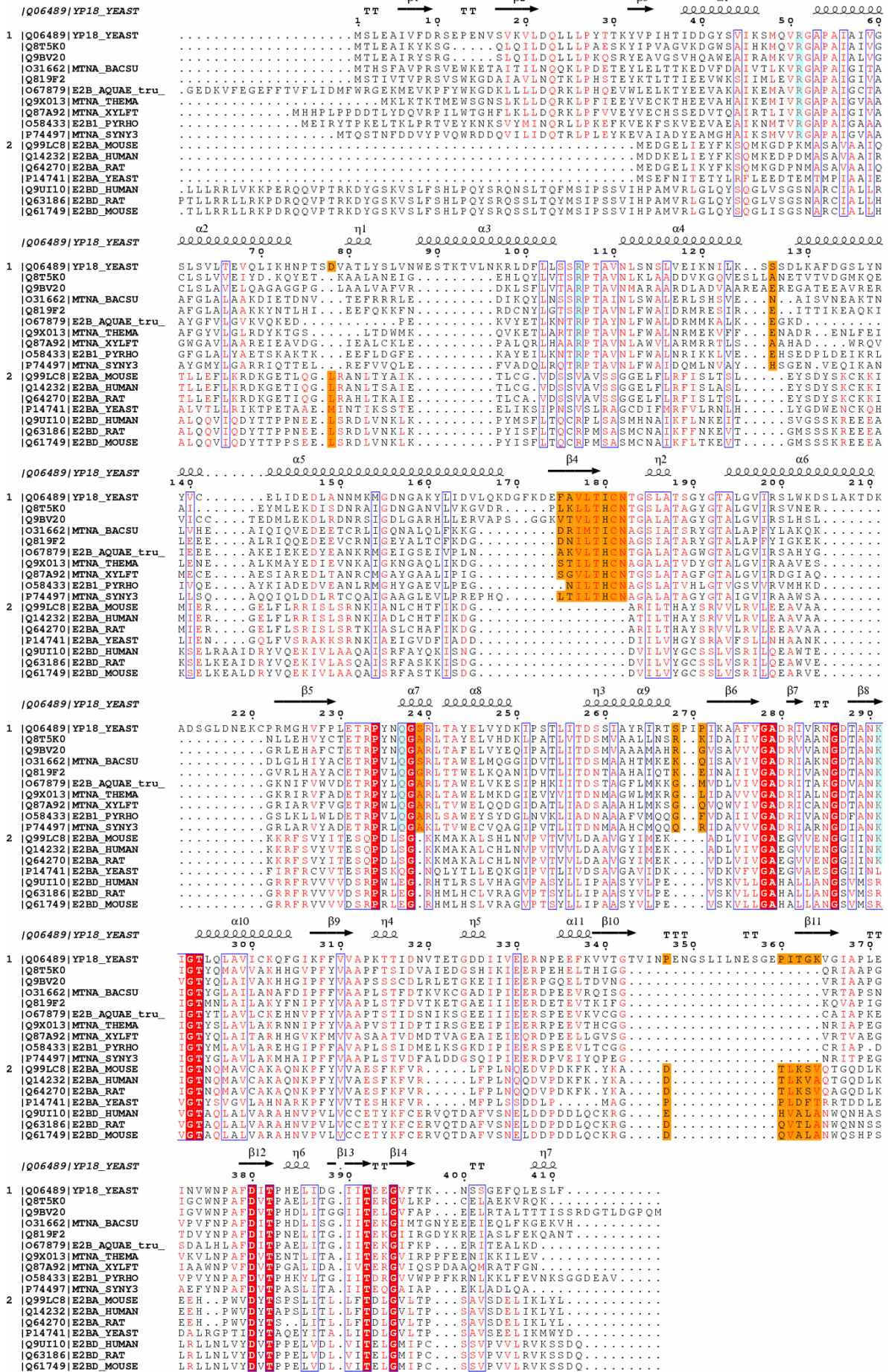
56

Figure 1: Sequence alignment. Coloring is according to similarity. Sequences were divided into two groups, one containing the putative MTNA (eIF2_rel) members of PF01008 (sequences 1-10) and the second comprising established eIF2B subunits (sequences 11 to 17). A red box with a white character means strict identity, a red character indicates similarity within a group. Red characters with a blue frame symbolize similarity between the two groups while an orange box indicates differences between the two conserved groups. The sulfate ion ligands have a cyan box. Sequence numbering is according to Ypr118w (YP18_YEAST). Q8T5K0 is annotated as translation initiation factor from *Anopheles gambiae*, Q9BV20 is a similar human protein sharing 41 % identity with Ypr118w. E2B_AQUAE and E2B1_PYRHO are annotated as translation initiation factors from *Aquifex aeolicus* and *Pyrococcus horikoshii*, respectively. About 120 to 140 amino acid long N-terminal extensions occurring in the eIF2B-δ subunits and in E2B-AQUAE have been omitted from the alignment.
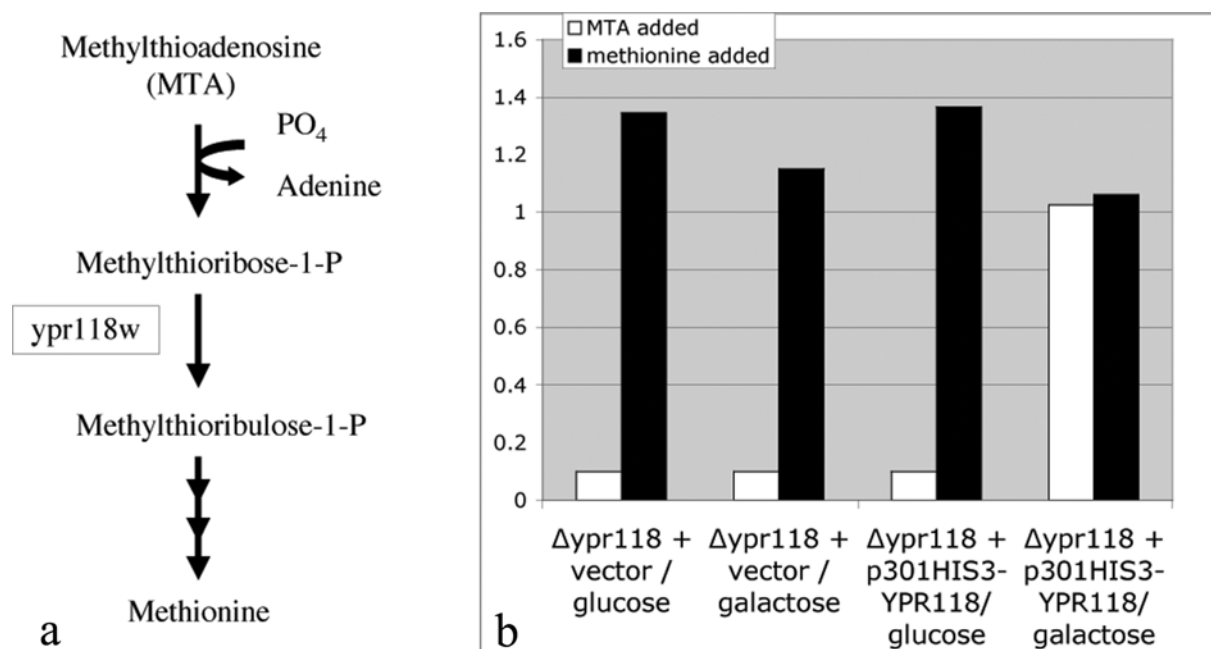
Figure 2

A. A scheme of the methionine salvage pathway is presented. Only the first steps leading to the production of 5-methylthioribulose-1-phosphate and the final product, methionine are presented.

B. Growth of a ypr118w-knockout strain on MTA. Cells transformed with construct p301HIS3-YPR118 (or control vector) were cultivated for 40 hours in minimal medium containing MTA or methionine in the presence of galactose or glucose as carbon source (for further details, see Materials and Methods)
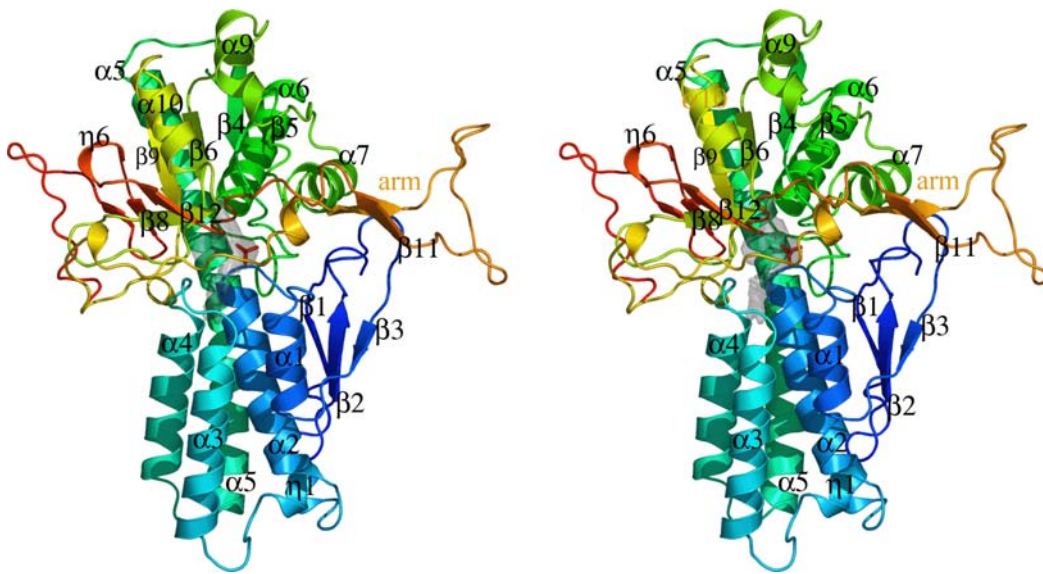
Figure 3A:

Stereo cartoon representation of the monomer. The rainbow coloring runs from blue (N-terminus) to red (C-terminus). The sulfate ion is shown as sticks, the cavity in grey.
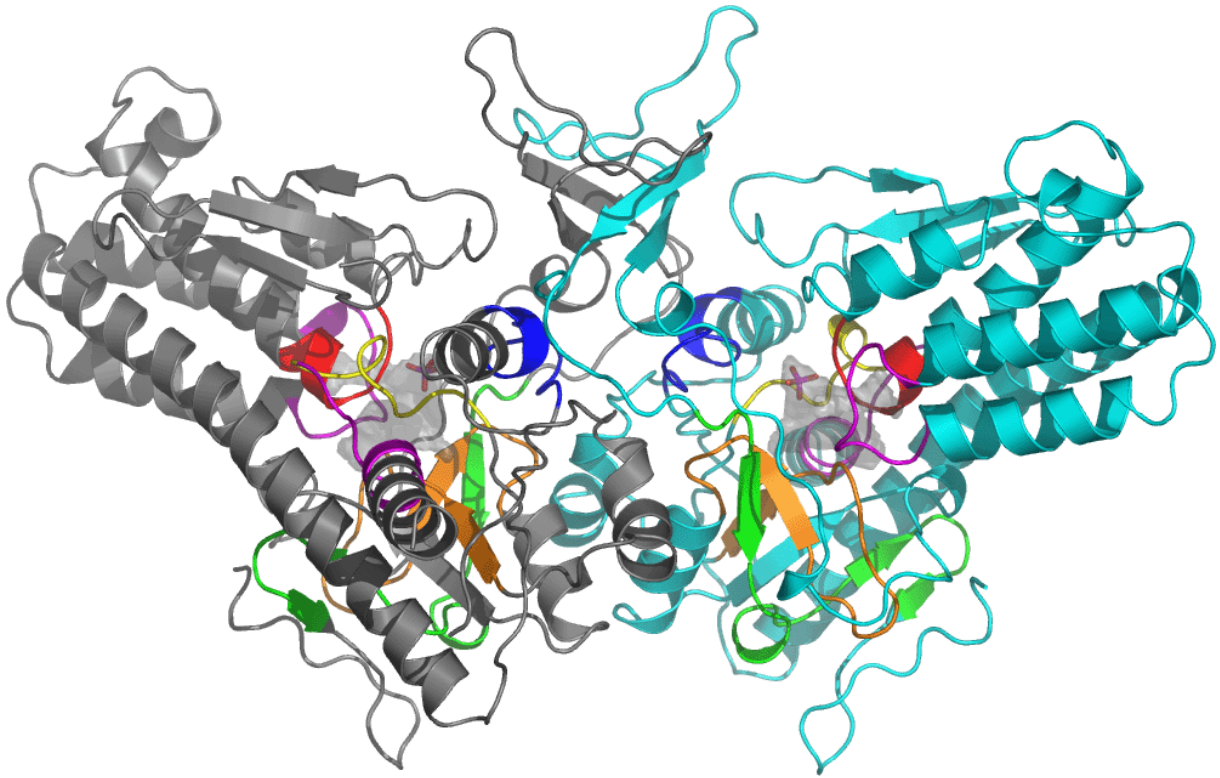
Figure 3B:

Cartoon representation of the dimer. One monomer is cyan, the other grey. The positions of the six conserved sequence motifs are indicated: motif 1 (red), motif 2 (magenta), motif 3 (yellow), motif 4 (blue), motif 5 (orange), motif 6 (green).
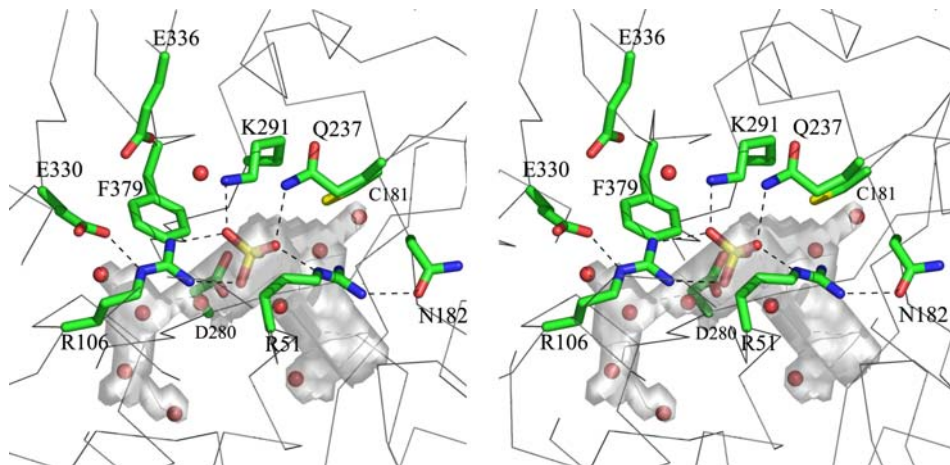
Figure 4: Close-up of the putative active site including the sulfate ion and the amino acids discussed in the text. Waters are shown as red spheres.
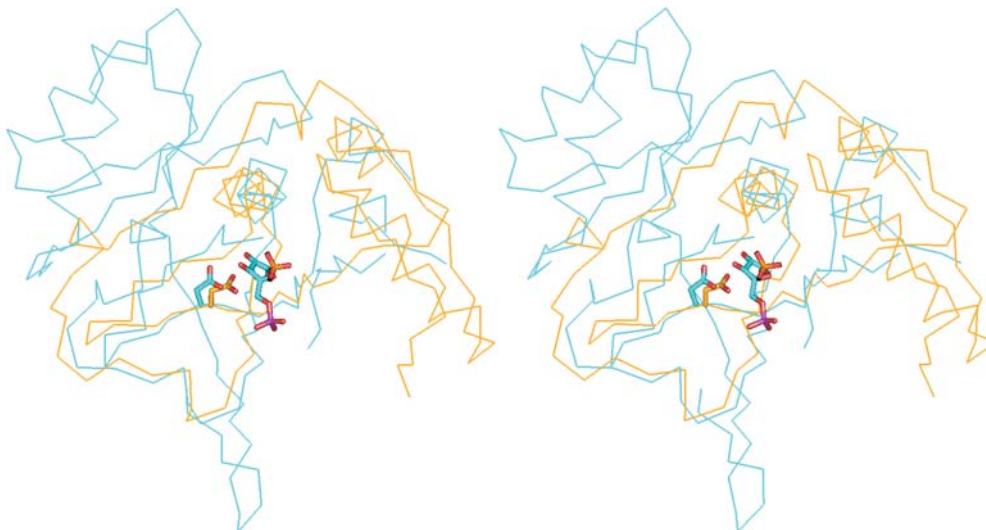
Figure 5: Overlay of RpiA from E*scherichia .coli* (cyan, PDB code 1OB8) and Ypr118w (orange). The structures were superimposed using the program LSQMAN (43). Represented by sticks is the RpiA competitive inhibitor arabinose-5-phosphate and the conserved aspartic acid residues 84/280, as well as the sulfate ion found in the Ypr118w structure.

Tables

Table I: Data collection and Refinement

Hamburg BW7A	Peak	Inflection	Remote
Data collection (XDS)			
Wavelength (Å)	0.9781	0.9785	0.9747
Resolution range[a]	40-2.78 (2.95-2.78)	40-2.78 (2.95-2.78)	40-2.78 (2.95-2.78)
No. observations	467781[b]	330284[b]	313278[b]
No. unique reflections	81766[b]	81570[b]	82182[b]
Completeness (%)	99.6 (98.9)	100 (100)	97.4 (89.8)
R_{sym}[c] (%)	8.5 (24.6)	8.0 (22.9)	8.4 (23.8)
$I/\sigma(I)$	17.6 (6.8)	15.0 (5.8)	13.9 (5.5)
FOM[d] MAD (overall)		0.41	
FOM[e] solvent flattened		0.72	

Grenoble BM14	Low remote	Peak	Inflection	High remote
Data collection (XDS)				
Wavelength (Å)	0.9879	0.9788	0.9790	0.8856
Resolution range[c]	40-1.75	40-2.40	40-2.40	40-2.40
(outer shell)	(1.87-1.75)	(2.56-2.40)	(2.56-2.40)	(2.56-2.40)
No. observations	674837	509232[b]	250563[b]	512565[b]
No. unique reflections	169966	127394[b]	124154[b]	127394[b]
Completeness (%)	99.6 (98.9)	100 (100)	97.4 (89.8)	100 (100)
R_{sym}[c] (%)	7.3 (29.7)	7.6 (19.0)	5.3 (13.4)	9.3 (25.1)
$I/\sigma(I)$	11.6 (4.1)	13.6 (6.5)	10.6 (5.3)	11.6 (5.1)
FOM[d] MAD (40 – 2.4 Å)	0.51			
FOM[e] solvent flattened	0.76			
Refinement (REFMAC)				
Resolution range (Å)	30-1.75			
No. reflections working set	167734			
No. reflections test set	2208			
No. non hydrogen atoms	13362			
Solvent water molecules	934			
Sulfate ions	4			
R/R_{free} (%)	17.1/19.8			
RMSD bond length (Å)	0.008			
RMSD bond angles (deg.)	1.4			

[a] The values in parentheses of resolution range, completeness, R_{sym} and $I/\sigma(I)$ correspond to the outermost resolution shell

[b] Friedel pairs were treated as different reflections.

[c] $R_{sym} = \Sigma_{hkl}\Sigma_j |I(hkl;j) - \langle I(hkl) \rangle| / (\Sigma_{hkl}\Sigma_j \langle I(hkl) \rangle)$ where $I(hkl;j)$ is the *j*th measurement of the intensity of the unique reflection (*hkl*) and $\langle I(hkl) \rangle$ is the mean over all symmetry-related measurements.

[d] Figure-of-merit as computed by SOLVE

[e] Figure-of-merit as computed by RESOLVE

Chapter III

Crystal structure of the phosphoenolpyruvate-binding enzyme I-domain from the *Thermoanaerobacter tengcongensis* PEP: sugar phosphotransferase system (PTS)*

Anselm Erich Oberholzer[‡§], Mario Bumann[‡§], Philipp Schneider[‡], Christoph Bächler[‡], Christian Siebold[⊥], Bernhard Erni[‡*] and Ulrich Baumann[‡*]

[‡]Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland

[⊥]Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Rosevelt Drive, Oxford OX3 7BN, UK

*Authors to whom correspondence should be addressed:

Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland; Tel. ++4131/631 43 46 (43 20); Fax ++41 (0)31/631 48 87; E-mail: bernhard.erni@ibc.unibe.ch, ulrich.baumann@ibc.unibe.ch

§ These two authors contributed equally to the work.

Running title: Crystal structure of the *T. tengcongensis* EIC domain of the PTS

Summary

The phosphoenolpyruvate: sugar phosphotransferase system (PTS) is a carbohydrate import mechanism unique to bacteria which tightly couples transport to substrate phosphorylation. The high-energy phosphate originates from PEP and is transferred sequentially to the sugar *via* various phosphoprotein intermediates. The first phosphate acceptor in this chain is enzyme I (EI), which in turn transfers the phosphate to HPr. Enzyme I is a multi-domain 64 kDa protein which is strongly conserved in bacteria. Therefore, it represents an attractive target for the development of novel antibiotics. In order to supplement structural information we present the crystal structure of the C-terminal domain of enzyme I of the thermophilic bacterium *Thermoanaerobacter tengcongensis* at 1.82 Å resolution. The dimeric protein folds in to a $(\beta\alpha)_8$ barrel with some large extrusions in the loops connecting $\beta2$ and $\alpha2$, $\beta3$ and $\alpha3$ and $\beta6$ and $\alpha6$. The dimer interface is very large burying 3749 Å^2 of accessible surface per monomer. A comparison to the homologous pyruvatephosphate dikinase reveals the conservation of the conformation of all the active site residues. The structure provides a rational explanation for the kinetic effects displayed in some recent inhibitor studies.

Introduction

The bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS) plays a dual role in transport and phosphorylation of sugars and hexitols on one hand and regulation of the cellular metabolism in response to the availability of these carbohydrates on the other. The molecular basis of its multiple functions is a protein phosphorylation cascade comprising four phosphoprotein units which sequentially transfer phosphoryl groups from phosphoenolpyruvate to the transported carbohydrates (for a review see [1]). Enzyme I (EI) is the first protein in this chain. It transfers phosphate from PEP to HPr, the general high energy phosphate carrier protein of the PTS (reviewed in [2]). HPr itself then serves as phosphoryl donor to the PTS transporters (EII, enzymes II) of different and sometimes overlapping substrate specificity. These transporters consist of three protein subunits or domains. The EIIA and EIIB domains serve as phosphate transfer relay to the carbohydrate substrate which is being translocated by the membrane spanning EIIC domain [3]. An increasing rate of sugar transport leads to a decrease of the phosphorylation of EII subunits which serves as the "input" for the PTS-mediated signal transduction. Phosphoryl transfer occurs through phosphohistidine and phosphocysteine intermediates. With PEP as substrate, EI links the protein phosphorylation relay of the PTS with glycolysis. PTS occur in eubacteria ([4-8] but not in eucaryotes and archaebacteria. EI of the PTS is one of the best conserved bacterial proteins with only minimal similarity to animal proteins [9].

EI is a 64 kDa two-domain protein [10-12]. The amino-terminal domain (EIN) contains a histidine which by double displacement transfers phosphate from PEP to HPr. The three dimensional crystal structure of EIN from *E. coli* has been elucidated, and its mode of interaction with HPr characterized by NMR spectroscopy [13, 14]. EIN is composed of the HPr-binding α-helical subdomain and an α/β subdomain which is structurally similar to the

phospho-histidine swivel domain of pyruvate phosphate dikinase (PPDK [15]), the latter bearing the phospho-histidine, which is His[189] in *E.coli*. The carboxy-terminal domain (EIC) contains the PEP binding site [16]. EIC of *E. coli* plays a crucial role in reversible dimerization of EI [17, 18]. It is proteolytically unstable. The unfolding transition temperature of EI varies between 41°C in wild-type EI in the absence of substrates and 60°C in the H189A mutant in the presence of PEP and Mg^{2+} [19, 20]. Attempts to express the *E. coli* EIC resulted in only partial success so far. The EIC (residues 240-572) fragment could be purified [21], but a twenty residues shorter variant could be purified only as fusion protein with the maltose binding protein MalE from which EIC was then released with a sequence specific protease (genenase) [22]. EIC of *Mycoplasma capricolum* was purified in an active form [18].

The mode of action of *E. coli* EI and how its activity is controlled is not yet fully understood. EIC is responsible for dimer formation which is strongly temperature-dependent [23]. The dissociation constant of *E. coli* EI has been reported to shift from 20 μM at 6°C to 0.9 μM at 30°C. The presence of divalent cations (Mg^{2+} or Mn^{2+}) and PEP also affects the equilibrium and the association and dissociation rate constants. The association rate constant is 2-3 orders of magnitude slower than in other dimeric proteins, suggesting that dimerization is accompanied by major conformational rearrangements of the interacting EIC domains [24-27]. The activity of *E. coli* EI at rate-limiting concentrations was enhanced by the addition of an inactive mutant [28, 29]. Changes of activity and fluorescence of pyrene labeled EI were correlated which was interpreted to reflect a correlation between function and dimerization [26, 27]. EI of *M. capricolum* has a 10^3 to 10^5 times larger association constant, suggesting that reversible dimer-monomer transitions are less likely to be part of the catalytic mechanism [18]. The cellular concentration of EI is around 5 μM [30].

Whereas the structure of the N-terminal domain of *E. coli* EI has been determined [13, 14], the structures of EIC or full-length protein are not yet known. *E. coli* EIC has 28% sequence identity with the PEP-binding domain of pyruvate phosphate dikinase (PPDK), an enzyme from the C4 pathway of plants which catalyses the transfer of phosphate from ATP to pyruvate. The X-ray structure of the PPDK-substrate complex has been solved, and the C-terminal domain been shown to adopt a $(\beta\alpha)_8$ barrel fold [15, 31-33]. Here we present the X-ray structure of the EIC domain from *Thermoanaerobacter tengcongensis*, a gram-negative, saccharolytic bacterium which optimally grows at 75°C [34, 35]. This EIC displays 54% amino acid sequence identity with the *E. coli* EIC domain.

Materials and Methods

Protein expression and purification

T. tengcongensis full-length EI and the N-terminal (His)$_6$-tagged construct of the C-terminal domain of Enzyme I (EIC(251-573)) were made by cloning the PCR product derived from a *Thermoanaerobacter tengcongensis* genomic DNA library into the *NdeI* and *BamHI* sites of the pET28a vector (Novagen) using the primers 5'-ACGTACATATGGAAGGATTAAAGCAGTTAAAAG-3' and 5'-ACGTAGGATCCTTAGCCAATATCTTTTATCACG-3'. EIC was expressed in Rosetta^{TM}(DE3) cells (Novagen). Cells were grown at 37°C to an A$_{550}$ of 0.8 and the expression was induced by 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 12 hr induction, cells were harvested and resuspended in 20 mM Tris-HCl, pH 8.0, 400 mM NaCl, 10 mM β-mercaptoethanol, 0.2 mM PMSF and disrupted using a French press. EIC was purified on 20 ml Ni-NTA SUPERFLOW (Qiagen) according the manufacturer's instructions. Typical yields were 30 mg/l pure EIC. After purification on Ni-NTA the eluate was

concentrated by Centriprep-30 (Millipore) to a volume of 2 ml. The protein was further purified by gel filtration using a Superdex 75 (Amersham Bioscience) column equilibrated in 10 mM HEPES, pH 7.5, 400 mM NaCl, 2 mM DTT. The purification steps were monitored by SDS-PAGE analysis [36]. The peak fractions were pooled and dialysed against 5 mM HEPES, pH 7.5, 2 mM DTT, 5 mM $CaCl_2$ for 2 hr at 4°C and concentrated to a final concentration of 10 mg/ml. Selenomethionine (SeMet)-labelled protein was prepared by the method of methionine biosynthesis inhibition [37]. Cells were grown at 37°C to an A550 of 0.6 and and the expression was induced by by 0.2 mM IPTG. After 12 hr of induction at 4°C, cells were harvested and the protein purified as the wild type.

Crystallisation

Trigonal crystals of EIC were obtained within 4 days by the sitting-drop vapour-diffusion method. Drops were set up by mixing 3 μl of protein solution (14 mg/ml) with 3 μl of reservoir solution (0.1 M sodium acetate trihydrate, pH 4.6, 12 % PEG 4000) and equilibrated against 120 μl reservoir solution at 291 K.

Crystals of selenomethionine-EIC were obtained within 4 days by mixing 2 μl of protein solution (10 mg/ml) with 2 μl of reservoir solution (0.1 M sodium acetate trihydrate, pH 4.6, 12 % PEG 4000, 7 % MPD). They belong to the orthorhombic space group $P2_12_12_1$ with cell parameters of a=82.81 Å, b=91.85 Å, c=185.96 Å and contain four monomers per asymmetric unit. Typical crystals have an average size of $100 \times 100 \times 110$ μm^3. Only the selenomethionine-labelled crystals were suitable for structure determination.

Data collection

Before data collection crystals were flash-cooled in a nitrogen stream at 110 K after raising the MPD concentration of the crystallization solution to 35 % (v/v). Data were collected in a MAD experiment at the Grenoble ESRF MAD beamline BM14 at 100 K, employing a Mosaic

225 CCD (MAR X-ray-research, Hamburg, Germany) detector. Due to the alignment of the crystals in the cryoloop only 0.2 degree/frame could be collected. All datasets were integrated and scaled with XDS [38, 39]. Peak, inflection and high-remote data were collected using one crystal. For refinement purposes a high-resolution low-remote data set was collected on a different crystal which was larger in size. The two crystals turned out to be non-isomorphous. Data statistics is given in Table 1.

Structure solution, refinement and analysis

Sixty out of sixtyfour expected selenium positions were determined by SHELXD [40] using F_A-values computed by XPREP (G. Sheldrick, unpublished program) employing peak, inflection and high-remote wavelength. Phases were computed using SOLVE [41] and improved by RESOLVE [42]. Automatic model building was done by RESOLVE. 142 out of 324 residues were build by the program. Further attempts to use the low-remote high-resolution data for automated model-building failed, even after transporting the electron density into the unit cell of the high-resolution crystal using MOLREP [43]. Therefore, the remaining residues were placed manually into the electron density using the program O [44] version 9.0.3. This initial model was the placed into the unit cell of the low-remote data set followed by density modification, phase extension and automatic model building using ArpWarp [45] version 6.0. Refinement was effected using REFMAC [46] version 5.1.24. Refinement statistics are given in Table 1.

Structure-based sequence alignment was done by Indonesia [47]. Figure 3 was prepared using ESPript [48]. Comparison of structures was effected using DALI [49] and LSQMAN [50]. The structures of EIC of *E. coli* and *M. capricolum* were modelled using the SWISS MODEL server [51-53] using EIC of *T. tengcongensis* as template. Three-dimensional structures of E-

Cl-PEP and Z-CL-PEP were created using The Dundee PRODRG2 Server [54] and were superimposed on P-pyr, an inhibitor of pyruvate phosphate dikinase [33] using O .

Structure figures were created using the program PYMOL (www.pymol.org).

Results

Cloning, purification and crystallization of EIC from T. tengcongensis

The known dynamic behaviour of *E. coli* EI may compromise the formation of well-ordered three-dimensional crystals. In the past we subcloned the genes for EI of *S. aureus, B. stearothermophilus, B. burgdorferi*, *E. coli, E. faecalis, H. influenzae* and *T. tengcongensis* and overexpressed the proteins in *E. coli*. All EIs complemented PTS-activity in an *E. coli ΔptsI* mutant, and after purification displayed *in vitro* phosphotransferase activity, indicating that the proteins were active. EI^{ntr} of *E. coli* was expressed in inclusion bodies only. The recombinant EI-like domain of the multiphosphoryl protein FruB of *P. aeruginosa* could not be expressed in *E. coli*. At first EI of *S. aureus* appeared the most suitable of all. To prevent the protein from aggregation and to improve homogeneity, all surface-exposed cysteines were exchanged against alanine and serine. One out of 12 mutant proteins afforded crystals which diffracted to a resolution of 3.5 Å but crystal mosaicity and disorder of the crystal lattice prevented interpretation (results not shown, C. Siebold PhD thesis, University of Berne, 2002).

Thermophilic orthologues have frequently proven their use in structural biology in the past. Consequently, we focussed our efforts on *Thermoanaerobacter tengcongensis*, one of very few thermophilic organisms possessing a PTS. Ni^{2+}-NTA purification of *T. tengcongensis* EI

expressed in *E.coli* afforded only fragments and almost no full-length protein. Four major fragments were sequenced and all were found to start with a methionine which hinted at internal translation initiation after rare codons. The most abundant fragment (>50%) turned out to be the complete C-terminal domain, EIC. The codons 252 to 573 of EI were PCR-amplified and the recombinant EIC domain expressed in an *E. coli* Rosetta strain which supplements tRNAs for rare codons. EIC could be purified to homogeneity by Ni^{2+}-NTA and gel-filtration chromatography. EIC is a stable homodimer with an unfolding transition temperature of 90°C (V. Navdaeva and B. Erni, unpublished results). While wildtype EIC yielded crystals of mediocre quality, selenomethionine-labelled EIC containing 16 Met (0.45 Met/kD) was prepared and crystallized in an orthorhombic spacegroup. These crystals were used for MAD experiments and structure refinement.

Crystal structure determination

Crystals of EIC belonged to the orthorhombic spacegroup $P2_12_12_1$ (a = 82.8 Å, b = 91.8 Å, c = 185.9 Å) with two physiological dimers in the asymmetric unit. Only reflections with l=4n of the (00*l*) reflections were strong and a native Patterson map at 4.5 Å resolution showed strong peaks (20 % of the origin peak) on the sections w = 0.25 and w = 0.5. The crystals diffracted to a maximum resolution of 1.82 Å (Table 1). The structure of EIC was solved by multiple anomalous dispersion (MAD) at 2.8 Å resolution. Using the peak wavelength alone resulted in better correlation coefficients in spacegroup $P2_12_12$ than in the true space group, but no interpretable electron density was obtained. Using the right space group and F_A-values yielded eventually a solution with a correlation coefficient of 70% and an interpretable electron density.

Refinement of the model against 1.82 Å data resulted in reasonable *R*-factors and satisfying model geometry (Table 1). 1272 residues (92.7 %) are in the most favoured regions and 90

residues (6.6%) in the additional allowed regions of the Ramachandran plot. The 4 residues (0.3 %) in disallowed regions are well defined in the electron density. Residues 251-572 of the EIC monomer are well ordered in all monomers, only the starting methionine 250 in one of the four monomers is disordered.

Overall structure

EIC features a $(\beta\alpha)_8$ barrel fold (Figure 1A). The four crystallographically independent monomers of the asymmetric unit are very similar. The RMS deviation of the C^α-positions over the entire length of a molecule is 0.35 Å. Only two solvent exposed loops (302-308 and 337-351) have a more than average RMS deviations of 0.7 Å and 1.5 Å, respectively (Figure 1A). These loops are engaged in crystal contacts in 2 of the four monomers. The $(\beta\alpha)_8$ barrel has an extra N-terminal α-helix and three extensions on the C-terminal face of the barrel. The first extension (296-309) which is inserted between $\beta2$ and $\alpha2$ has an irregular structure. The second extension (333-365) between $\beta3$ and $\alpha3$ comprises two short α-helices, and the third (452-479) between $\beta6$ and $\alpha6$ comprises one extra α-helix. The three extensions protrude approximately 30 Å above the core of the barrel. The active site cysteine (see below) is located in the $\beta\alpha7$ turn and accessible over the short $\beta\alpha8$ and $\beta\alpha1$ turns at the rim of the barrel. The shape of EIC can be compared with an "easy chair", the three extensions forming the backrest and the C-terminal end of the barrel the cushion (Figure 1B).

EI of *E. coli* is a homodimer (Figure 1C) and EIC was identified as the dimerization domain [18]. The asymmetric unit of *T. tengcongensis* EIC contains two identical dimers. There is an extensive contact area comprising 45 residues which cover 3'749 $Å^2$ per monomer between the two monomers. 15 % of the EIC subunit surface becomes buried upon dimer formation. This large interface is formed mostly by back to back contacts between the $\beta3/\alpha3$ and $\beta6/\alpha6$

extensions (the backrest, Figure 1C). Additional contacts are provided by the β4/α4 and β5/α5 turns and the first half of helix 6. The interface is composed mainly of hydrophobic residues. The contact area of more than 3'700 Å^2 is exceptionally large for a dimerization domain which usually are of the order of 1200-2400 Å^2 [55].

Relationship to pyruvate phosphate dikinase

A comparison of the amino acid sequence of the maize and *Bacillus symbiosus* pyruvate phosphate dikinase (PPDK) with other protein sequences revealed homology to the mechanistically related enzymes PEP synthase and EI of the PTS [56]. PPDK catalyzes the reversible conversion of ATP, inorganic phosphate (P$_i$) and pyruvate to AMP, pyrophosphate and phosphoenolpyruvate (PEP) [15]. The sequence similarity extends from the α/β fold (swivel domain) which contains the active site histidine to the C-terminal PEP/pyruvate binding domain [15].

EIC and the C-terminal PEP/pyruvate binding domains of PPDK (PDB entry 1DIK) display 25% sequence identity and are the closest structural homologues in the protein data bank with a DALI Z-score of 31.0. Both possess an (β/α)$_8$ fold and 279 amino acids out of the 360 of PPDK and 320 of EIC can be aligned with a RMS deviation of less than 2.2 Å.

Using INDONESIA, 268 residues of EIC can be superimposed on PPDK with bound phosphonopyruvate (PDB entry 1KC7) with an RMS deviation of 1.6 Å for the paired C^α-atoms (Figure 2). Figure 3 shows a structure-based sequence alignment between EIC and the pyruvate/PEP binding domain of PPDK.

EIC shares with the PEP/pyruvate binding domain of pyruvate phosphate dikinase the three extensions and the large intersubunit contact area. The PPDK extensions are however longer. The β/α2 and extensions features an α helical hairpin. β/α3 and β/α6 also have more α-helical structure. 32 out of the 45 residues making up the dimer interfaces of EIC and PPDK

are structurally conserved (Figure 3). The PEP binding site at the C-terminal end of the β/α barrel are almost completely superimposable (see below).

Substrate binding site

The PEP-binding site of E. coli EI has been identified by affinity labelling with the mechanism-based inhibitor Z-chloro-PEP [29] which labelled Cys-502 covalently during catalytic turnover. In PPDK, the structural details of the binding site have been characterized with crystals of PPDK in complex with the substrate analogue 3-phosphonopyruvate (P-pyr). In the phosphonopyruvate the C=O group isosterically replaces C=CH2, while the -CH2-PO3 replaces the -O-PO3 of PEP. A rich network of interactions ensures the precise positioning of the P-pyr ligand.in the PPDK binding site [33]. A superimposition of the PPDK/P-Pyr structure with EIC shows that all the residues that form the charge network with the substrate analogue and Mg^{2+} are also conserved in EIC (Figure 4). P-pyr can be accommodated at the C-terminal end of the β/α-barrel of the *T. tengcongensis* EIC structure without sterical clashes and constraints. The side chain conformations of the active site residues are remarkably well conserved, and this independently of whether a substrate is bound as in PPDK or not as in EIC. This may be taken as evidence for a rather rigid and inflexible structure of the PEP binding sites of PPDK and *T. tengcongensis* EIC.

The reaction mechanism at the PEP binding site of E. coli EI has previously been characterized with C-3 modified PEP-analogues [29, 57]. The K_m of *E. coli* EI for Z-ClPEP (Z: COOH and Cl trans to C=C double bond) and E-ClPEP were found to be similar (0.04 mM and 0.12 mM) to the K_m for PEP (0.14 mM), while the k_{cat} values were 2 and 3 orders of magnitude lower for the *Z*- and *E*-isomer, respectively. The Cl-PEP isomers were modelled in the EIC binding site with the carboxylate, C2 and phosphate in the same position as those of

P-Pyr in the PPDK model. Under these conditions the orientation of the bulky and electronegative chloride atoms at C3 are different for both isomers. Due to sterical clashes, the chlorine atom of Z-ClPEP (Figure 5B) may change the position of the Z-ClPEP nearer to the surface whereas the E-ClPEP (Figure 5A) still fits in the binding site similar to PEP. The C-3 of Z-ClPEP is turned away from the general acid/base Cys-502 which may lead to a much slower protonation rate and dissociation of the enolate from the binding pocket before protonation can take place. This would provide an explanation for the non-stereospecific protonation of the Z-ClPEP-enolate by water after its release into the aqueous medium.

Thermostability and salt bridges

Electrostatic interactions are thought to act as an important factor conferring thermostability to proteins [58, 59]. This opinion is supported by the increased number of salt bridges found in many structures of thermostable proteins. On the basis of the known structure of EIC from *T. tengcongensis* the structures of EIC from *E. coli* and *M. capricoloum* were modelled. The thermostable EIC from *T. tengcongensis* exhibits 17 salt bridges. Eight charged residues are forming an ion-pair network (Figure 6). In EIC of *E. coli* and *M. capricolum* are 12 and 9 salt bridges present. In both structures, six charged residues are forming an ion-pair network. The higher denaturation temperature of EIC from *T. tengcongensis* contrary to EIC from the mesophilic organisms *E. coli* and *M. capricolum* could be a result of the increased number of salt bridges and the larger ion-pair network on the surface.

Discussion

Enzyme I of *E. coli*, the key enzyme of the PTS, has been the object of numerous physiological, biochemical and biophysical studies. The structure of the HPr-binding N-

terminal domain was solved by X-ray [13] and multidimensional NMR [14] eight years ago. The structure of the full-length protein, however, has not been solved, possibly because the dynamic behaviour of EI and the marginal stability of the recombinant *E. coli* EIC disfavoured crystallisation, and because the structure of EIC could be predicted based on the sequence similarity with the PEP binding domain of PPDK. Hyperthermophilic bacteria such as *Thermotoga maritima, Thermus thermophilus* and archaebacteria do not have a PTS. *T. tengcongensis*, a saccharolytic bacterium which optimally grows at 70°C turned out to have a PTS very similar to *E. coli*. EIC from *T. tengcongensis* is thermally stable to 90°C and more stable than the EIN domain, unlike in *E. coli* where EIN is significantly more stable than EIC. EIC of *E. coli* and *T. tengcongensis* have almost identical amino acid composition and pI. The contents of glycine, and basic amino acids are identical and only the Lys/Arg ratios are inversed. The sums of carboxylate and amide containing residues are identical but the carboxlyate to amide ratio is shifted in favour of more Asn and less Gln in the thermotolerant EIC. The only Asn-Gly sequence, a linkage thought to be weak and hydrolysis-sensitive, is conserved in both proteins. EIC of *T. tengcongensis* contains 5 extra prolines in non-conserved sequences. There are thus no obvious alterations in the amino acid sequence and composition which would provide an explanation for the remarkably different thermostability. The increased number of surface exposed salt bridges in EIC of *T. tengcongensis* in comparison with EIC from *E. coli* and *M. capricolum* might one of the major reasons for the different thermostability.

The purified PTS proteins of *T. anaerobacter* displays maximum *in vitro* phosphotransferase acitivity at 70°C (Navdaeva and Erni, unpublished). Although this value may reflect the optimum of another PTS protein than EI, it is 20°C below the thermal unfolding temperature. This is more than the difference for *E. coli* EI (Tm 41°C to 60°C depending on the experimental conditions) and it raises the question whether thermotolerant EI can display the same complex dynamical behavior at 70°C that EI of *E. coli* displays at 37°C. EI plays the

role of a gate-keeper by controlling the flux of phosphorylgroups into the PTS and it also provides the interface between PTS and chemotaxis [60, 61]. Moreover EI can utilize not only PEP but also acetyl phosphate kinase as source of high energy phosphate [62]. The genome of *T. tencongensis* contains the genes for all these proteins and therefore EI is likely to function similarly in *T. tengcongensis* and *E. coli*. But there are differences downstream along the protein phosphorylation cascade. *T. tencongensis* contains no *crr* gene for the EIIA^{Glc} subunit which in *E. coli* mediates inducer exclusion and activation of adenylcyclase, and also no gene for an adenylcyclase. On the other hand it does contain two *crp* genes for catabolite gene activators (cAMP receptor proteins) which are controlled by cAMP. How Crp is activated in the absence of an adenylcylase is not clear but it points to differences downstream in the PTS system which might be compensated by differences upstream. Control and catalysis of EI in *T. tencongensis* may be less complex, more robust and ultimately better adjusted to the environment. The contact areas in the EIC dimer of *T. tencongensis* and in the modelled dimeric structures of *E. coli* and *M. capricolum* are comparable with respect to both, area and disribution of apolar, polar and charged residues (Figure 7) and thus also do not offer an explanation for the different stabilities.

Footnotes

References

1. Postma, P.W., Lengeler, J. W., and Jacobson, G. R. (1996). *Escherichia coli and Salmonella: cellular and molecular biology*. Washington, DC.
2. Ginsburg, A. & Peterkofsky, A. (2002). Enzyme I: the gateway to the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *Arch Biochem Biophys* **397**, pp. 273-278.
3. Robillard, G.T. & Broos, J. (1999). Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim Biophys Acta* **1422**, pp. 73-104.
4. Reizer, J., Reizer, A. & Saier, M.H., Jr. (1995). Novel phosphotransferase system genes revealed by bacterial genome analysis--a gene cluster encoding a unique Enzyme I and the proteins of a fructose-like permease system. *Microbiology* **141 (Pt 4)**, pp. 961-971.
5. Reizer J, C.A., Reizer A, Saier MH (1996a). Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar-specific permease domains of the phosphotransferase system and pentose catabolic enzymes. *Genome Sci Technol* **1**, pp. 53-57.
6. Reizer J, R.A., Merrick MJ, Plunkett G, Rose DJ, Saier MH (1996b). Novel phosphotransferase-encoding genes revealed by analysis of the Escherichia coli genome: a chimeric gene encoding an Enzyme I homologue that possesses a putative sensory transduction domain. *Gene* **181**, pp. 103-108.
7. Reizer, J., Bachem, S., Reizer, A., Arnaud, M., Saier, M.H., Jr. & Stulke, J. (1999). Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of Bacillus subtilis. *Microbiology* **145 (Pt 12)**, pp. 3419-3429.
8. Tchieu, J.H., Norris, V., Edwards, J.S. & Saier, M.H., Jr. (2001). The complete phosphotranferase system in Escherichia coli. *J Mol Microbiol Biotechnol* **3**, pp. 329-346.
9. Mukhija, S. & Erni, B. (1997). Phage display selection of peptides against enzyme I of the phosphoenolpyruvate-sugar phosphotransferase system (PTS). *Mol Microbiol* **25**, pp. 1159-1166.
10. Saier, M.H., Jr., Simoni, R.D. & Roseman, S. (1970). The physiological behavior of enzyme I and heat-stable protein mutants of a bacterial phosphotransferase system. *J Biol Chem* **245**, pp. 5870-5873.
11. Cordaro, J.C. & Roseman, S. (1972). Deletion mapping of the genes coding for HPr and enzyme I of the phosphoenolpyruvate: sugar phosphotransferase system in Salmonella typhimurium. *J Bacteriol* **112**, pp. 17-29.
12. LiCalsi, C., Crocenzi, T.S., Freire, E. & Roseman, S. (1991). Sugar transport by the bacterial phosphotransferase system. Structural and thermodynamic domains of enzyme I of Salmonella typhimurium. *J Biol Chem* **266**, pp. 19519-19527.
13. Liao, D.I., Silverton, E., Seok, Y.J., Lee, B.R., Peterkofsky, A. & Davies, D.R. (1996). The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the E. coli PEP: sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure* **4**, pp. 861-872.
14. Garrett, D.S., Seok, Y.J., Liao, D.I., Peterkofsky, A., Gronenborn, A.M. & Clore, G.M. (1997). Solution structure of the 30 kDa N-terminal domain of enzyme I of the

Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry* **36**, pp. 2517-2530.

15. Herzberg, O., Chen, C.C., Kapadia, G., McGuire, M., Carroll, L.J., Noh, S.J. & Dunaway-Mariano, D. (1996). Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc Natl Acad Sci U S A* **93**, pp. 2652-2657.

16. Seok, Y.J., Lee, B.R., Gazdar, C., Svenson, I., Yadla, N. & Peterkofsky, A. (1996). Importance of the region around glycine-338 for the activity of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry* **35**, pp. 236-242.

17. Chauvin F, F.A., Johnson CR, Roseman S (1996b). The N-terminal domain of Escherichia coli enzyme I of the phosphoenolpyruvate/glucose phosphotransferase system: Molecular cloning and characterisation. *Proc Natl Acad Sci U S A* **93**, pp. 7028-7030.

18. Zhu, P.P., Szczepanowski, R.H., Nosworthy, N.J., Ginsburg, A. & Peterkofsky, A. (1999). Reconstitution studies using the helical and carboxy-terminal domains of enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry* **38**, pp. 15470-15479.

19. Dimitrova, M.N., Szczepanowski, R.H., Ruvinov, S.B., Peterkofsky, A. & Ginsburg, A. (2002). Interdomain interaction and substrate coupling effects on dimerization and conformational stability of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry* **41**, pp. 906-913.

20. Dimitrova, M.N., Peterkofsky, A. & Ginsburg, A. (2003). Opposing effects of phosphoenolpyruvate and pyruvate with Mg(2+) on the conformational stability and dimerization of phosphotransferase enzyme I from Escherichia coli. *Protein Sci* **12**, pp. 2047-2056.

21. Brokx, S.J., Talbot, J., Georges, F. & Waygood, E.B. (2000). Enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system. In vitro intragenic complementation: the roles of Arg126 in phosphoryl transfer and the C-terminal domain in dimerization. *Biochemistry* **39**, pp. 3624-3635.

22. Fomenkov, A., Valiakhmetov, A., Brand, L. & Roseman, S. (1998). In vivo and in vitro complementation of the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system by the cloned C-terminal domain. *Proc Natl Acad Sci U S A* **95**, pp. 8491-8495.

23. Kukuruzinska, M.A., Turner, B.W., Ackers, G.K. & Roseman, S. (1984). Subunit association of enzyme I of the Salmonella typhimurium phosphoenolpyruvate: glycose phosphotransferase system. Temperature dependence and thermodynamic properties. *J Biol Chem* **259**, pp. 11679-11681.

24. Han, M.K., Knutson, J.R., Roseman, S. & Brand, L. (1990). Sugar transport by the bacterial phosphotransferase system. Fluorescence studies of subunit interactions of enzyme I. *J Biol Chem* **265**, pp. 1996-2003.

25. Chauvin, F., Toptygin, D., Roseman, S. & Brand, L. (1992). Time-resolved intrinsic fluorescence of Enzyme I. The monomer/dimer transition. *Biophys Chem* **44**, pp. 163-173.

26. Chauvin, F., Brand, L. & Roseman, S. (1994). Sugar transport by the bacterial phosphotransferase system. Characterization of the Escherichia coli enzyme I monomer/dimer equilibrium by fluorescence anisotropy. *J Biol Chem* **269**, pp. 20263-20269.

27. Chauvin F, B.L., Roseman S (1996a). Sugar transport by the bacterial phosphotransferase system. Characterization of the Escherichia coli enzyme I

monomer/dimer transition kinetics by fluorescence anisotropy. *J Biol Chem* **269**, pp. 20270-20274.

28. Seok, Y.J., Zhu, P.P., Koo, B.M. & Peterkofsky, A. (1998). Autophosphorylation of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system requires dimerization. *Biochem Biophys Res Commun* **250**, pp. 381-384.

29. Garcia-Alles, L.F., Alfonso, I. & Erni, B. (2003). Enzyme I of the phosphotransferase system: induced-fit protonation of the reaction transition state by Cys-502. *Biochemistry* **42**, pp. 4744-4750.

30. Rohwer, J.M., Meadow, N.D., Roseman, S., Westerhoff, H.V. & Postma, P.W. (2000). Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements in vitro. *J Biol Chem* **275**, pp. 34909-34921.

31. Xu, Y., Yankie, L., Shen, L., Jung, Y.S., Mariano, P.S., Dunaway-Mariano, D. & Martin, B.M. (1995). Location of the catalytic site for phosphoenolpyruvate formation within the primary structure of Clostridium symbiosum pyruvate phosphate dikinase. 1. Identification of an essential cysteine by chemical modification with [1-14C]bromopyruvate and site-directed mutagenesis. *Biochemistry* **34**, pp. 2181-2187.

32. Wei, M., Li, Z., Ye, D., Herzberg, O. & Dunaway-Mariano, D. (2000). Identification of domain-domain docking sites within Clostridium symbiosum pyruvate phosphate dikinase by amino acid replacement. *J Biol Chem* **275**, pp. 41156-41165.

33. Herzberg, O., Chen, C.C., Liu, S., Tempczyk, A., Howard, A., Wei, M., Ye, D. & Dunaway-Mariano, D. (2002). Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis. *Biochemistry* **41**, pp. 780-787.

34. Xue, Y., Xu, Y., Liu, Y., Ma, Y. & Zhou, P. (2001). Thermoanaerobacter tengcongensis sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol* **51**, pp. 1335-1341.

35. Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J. & Yang, H. (2002). A complete sequence of the T. tengcongensis genome. *Genome Res* **12**, pp. 689-700.

36. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, pp. 680-685.

37. Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**, pp. 105-124.

38. Kabsch, W.e.a. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21**, pp. 67-71.

39. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, pp. 795-800.

40. Schneider, T.R. & Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, pp. 1772-1779.

41. Terwilliger, T.C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55** (**Pt 4**), pp. 849-861.

42. Terwilliger, T.C. & Berendzen, J. (1999). Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods. *Acta Crystallogr D Biol Crystallogr* **55** (**Pt 2**), pp. 501-505.

43. Vagin, A., Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, pp. 1022-1025.

44. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, pp. 110-119.

45. Perrakis, A., Morris, R. & Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, pp. 458-463.

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

47. D. Madsen, P.J., and G. J. Kleywegt (2002). Indonesia: An integrated sequence analysis system. *(in prep.)*.

48. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, pp. 3320-3323.

49. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, pp. 123-138.

50. Kleywegt, G.J.J., T.A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, pp. 525-545.

51. Schwede, T., Kopp, J., Guex, N. & Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31**, pp. 3381-3385.

52. Guex, N. & Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, pp. 2714-2723.

53. Peitsch, M.C. (1995). Protein modeling by E-mail. *Bio/Technology* **13**, pp. 658-660.

54. van Aalten, D.M., Bywater, R., Findlay, J.B., Hendlich, M., Hooft, R.W. & Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des* **10**, pp. 255-262.

55. Ponstingl, H., Henrick, K. & Thornton, J.M. (2000). Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins* **41**, pp. 47-57.

56. Pocalyko, D.J., Carroll, L.J., Martin, B.M., Babbitt, P.C. & Dunaway-Mariano, D. (1990). Analysis of sequence homologies in plant and bacterial pyruvate phosphate dikinase, enzyme I of the bacterial phosphoenolpyruvate: sugar phosphotransferase system and other PEP-utilizing enzymes. Identification of potential catalytic and regulatory motifs. *Biochemistry* **29**, pp. 10757-10765.

57. Garcia-Alles, L.F., Flukiger, K., Hewel, J., Gutknecht, R., Siebold, C., Schurch, S. & Erni, B. (2002). Mechanism-based inhibition of enzyme I of the Escherichia coli phosphotransferase system. Cysteine 502 is an essential residue. *J Biol Chem* **277**, pp. 6934-6942.

58. Perutz, M.F. & Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**, pp. 256-259.

59. Karshikoff, A. & Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem Sci* **26**, pp. 550-556.

60. Lux, R., Jahreis, K., Bettenbrock, K., Parkinson, J.S. & Lengeler, J.W. (1995). Coupling the phosphotransferase system and the methyl-accepting chemotaxis protein-dependent chemotaxis signaling pathways of Escherichia coli. *Proc Natl Acad Sci U S A* **92**, pp. 11583-11587.

61. Lux, R., Munasinghe, V.R., Castellano, F., Lengeler, J.W., Corrie, J.E. & Khan, S. (1999). Elucidation of a PTS-carbohydrate chemotactic signal pathway in Escherichia coli using a time-resolved behavioral assay. *Mol Biol Cell* **10**, pp. 1133-1146.

62.   Fox, D.K. & Roseman, S. (1986). Isolation and characterization of homogeneous acetate kinase from Salmonella typhimurium and Escherichia coli. *J Biol Chem* **261**, pp. 13487-13497.

Figures and figure legends
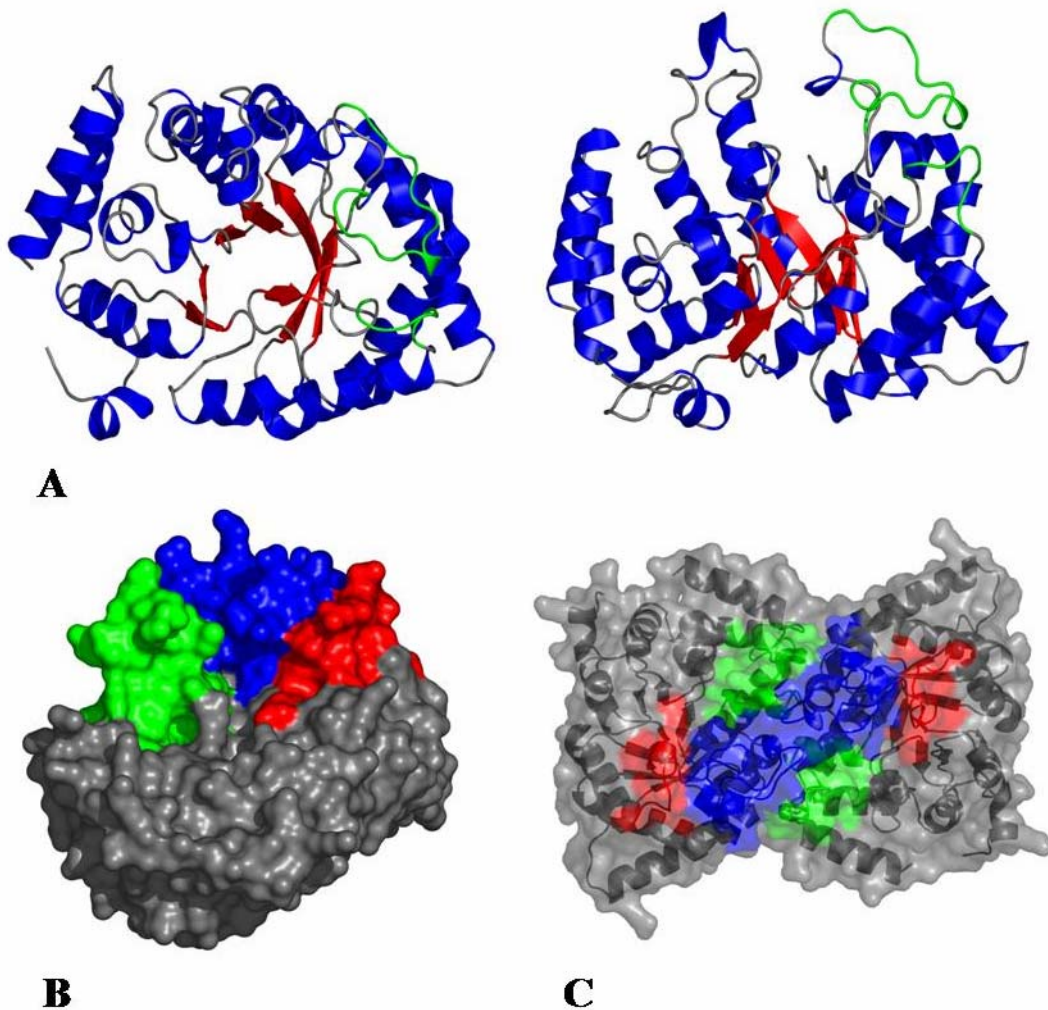
Figure 1:

A: Cartoon representation of the monomer structure of EIC: α-helices are coloured in blue, β-sheets are coloured in red and loops are coloured in grey. Residues 302 to 308 and 337 to 351 are coloured in green.

B: Surface representation of the monomer of EIC. Residues 296 to 309 are coloured in red, residues 333-365 are coloured in blue and residues 452-479 are coloured in red.

C: Cartoon representation of the dimmer of EIC. Residues 296 to 309 are coloured in red, residues 333-365 are coloured in blue and residues 452-479 are coloured in red. The surface representation is superposed in semi-transparent.
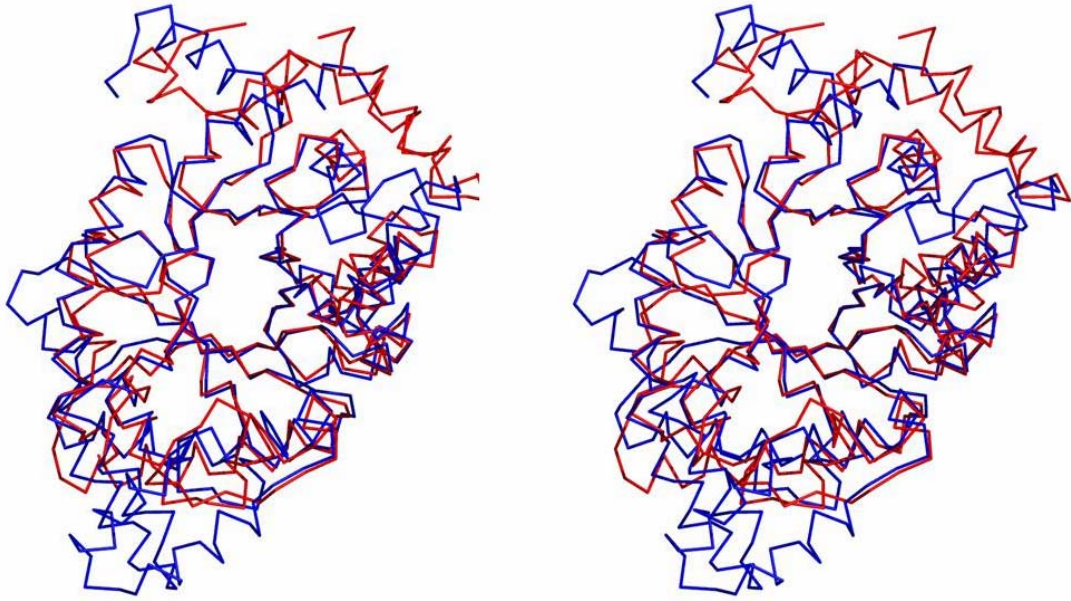
Figure 2: Stereoview of a superimposition of EIC (red) and the PEP/pyruvate binding site of PPDK (blue).

EIC_Q8R7R4

α0 TT β1 η1 α1
250   260   270   280   290

EIC_Q8R7R4   . . . . M E . G L K Q L K D L P A E T P D G K K V M L A A N I G T P K D V A S A L A N G A E G V
PPDK_P22983 K T F E L G G H T F A E G D Y I S L D G S T G K I Y K G D I E T Q E A S V S G S F E R I M . . . V W A D K F R T L K V R T N A D T P E D T L N A V K L G A E G I

490   500   510   520   530   540   550
PPDK_P22983
β21' β22' TT β23' β24' β25' α0 β1 α1

β2 η2 α2 β3 T . . . .

EIC_Q8R7R4

300   310   320   330

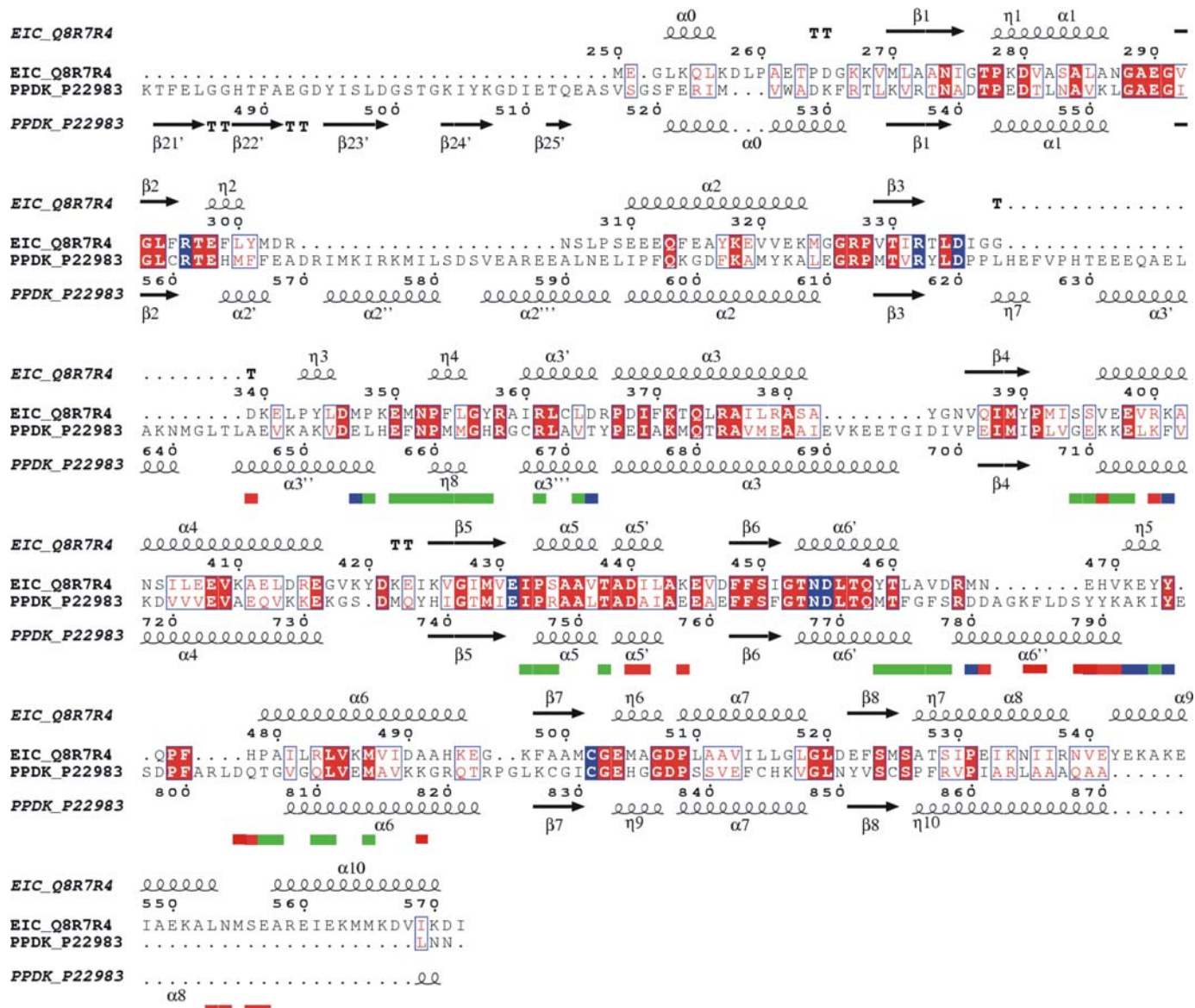EIC_Q8R7R4   G L F R T E F L Y M D R . . . . N S L P S E E E Q F E A Y K E V V E K M G G R P V T I R T L D I G G . . . .
PPDK_P22983 G L C R T E H M F F E A D R I M K I R K M I L S D S V E A R E E A L N E L I P F Q K G D F K A M Y K A L E G R P M T V R Y L D P P L H E F V P H T E E E Q A E L

560   570   580   590   600   610   620   630
PPDK_P22983
β2 α2' α2'' α2''' α2 β3 η7 α3'

. . . . T η3 η4 α3' α3 β4

EIC_Q8R7R4

340   350   360   370   380   390   400

EIC_Q8R7R4   . . . . D K E L P Y L D M P K E M N P F L G Y R A I R L C L D R P D I F K T Q L R A I L R A S A . . . . Y G N V Q I M Y P M I S S V E E V R K A
PPDK_P22983 A K N M G L T L A E V K A K V D E L H E F N P M M G H R G C R L A V T Y P E I A K M Q T R A V M E A A I E V K E E T G I D I V P E I M I P L V G E K K E L K F V

640   650   660   670   680   690   700   710
PPDK_P22983
α3'' η8 α3''' α3 β4

α4 β5 α5 α5' β6 α6' η5

EIC_Q8R7R4

410   420   430   440   450   460   470

EIC_Q8R7R4   N S I L E E V K A E L D R E G V K Y D K E I K V G I M V E I P S A A V T A D I L A K E V D F F S I G T N D L T Q Y T L A V D R M N . . . . E H V K E Y Y .
PPDK_P22983 K D V V V E V A E Q V K K E K G S . D M Q Y H I G T M I E I P R A A L T A D A I A E E A E F F S F G T N D L T Q M T F G F S R D D A G K F L D S Y Y K A K I Y E

720   730   740   750   760   770   780   790
PPDK_P22983
α4 β5 α5 α5' β6 α6' α6''

α6 β7 η6 α7 β8 η7 α8 α9

EIC_Q8R7R4

480   490   500   510   520   530   540

EIC_Q8R7R4   . Q P F . . . . H P A I L R L V K M V I D A A H K E G . . K F A A M C G E M A G D P L A A V I L L G L G L D E F S M S A T S I P E I K N I I R N V E Y E K A K E
PPDK_P22983 S D P F A R L D Q T G V G Q L V E M A V K K G R Q T R P G L K C G I C G E H G G D P S S V E F C H K V G L N Y V S C S P F R V P I A R L A A A Q A A . . . . . .

800   810   820   830   840   850   860   870
PPDK_P22983
α6 β7 η9 α7 β8 η10

α10

EIC_Q8R7R4

550   560   570

EIC_Q8R7R4   I A E K A L N M S E A R E I E K M M K D V I K D I
PPDK_P22983 . . . . L N N .

PPDK_P22983   . . . .
α8

Figure 3: Structure-based sequence alignment of EIC and PPDK. Conserved residues involved in PEP/pyruvate and Mg^{2+} binding are coloured in blue. The coloured bar below the sequence shows the dimer interface. A red bar symbolises residues only involved in the EIC dimer interface, a blue bar residues of the PPDK dimer interface and a green bar residues involved in both dimer interfaces.
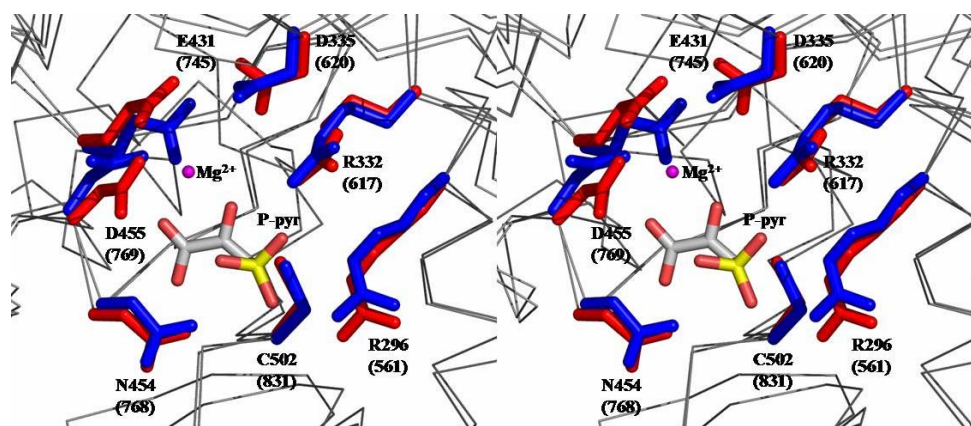
Figure 4: Stereo close up view of a superimposition of the putative active site of EIC with the PEP/pyruvate binding site of PPDK containing P-pyr and Mg^{2+}. Residues of EIC are colored in red, residues of PPDK are colored in blue The numbers in parentheses correspond with the PPDK residues. P-pyr is shown as a stick model. Atomic colours are as follows: oxygen red, carbon white, phosphorus yellow and Mg^{2+} magenta.
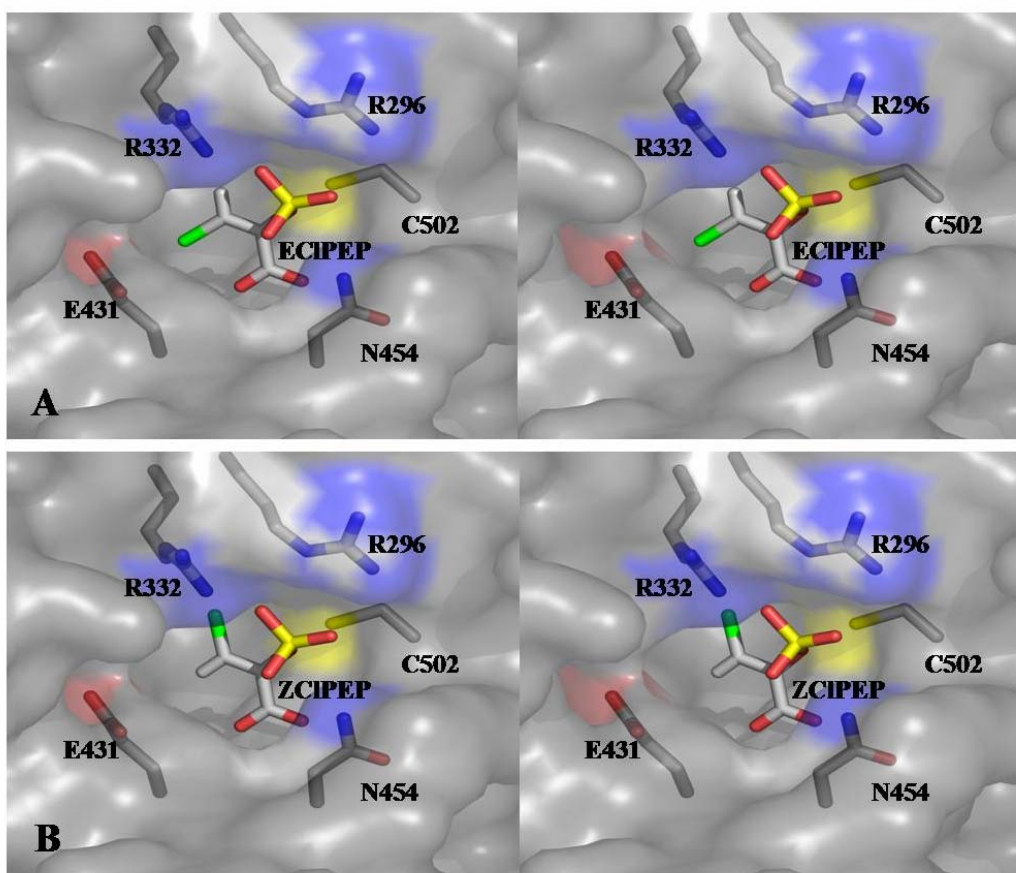
Figure 5:

A: Stereoview of the active site of EIC with modelled EClPEP. EClPEP and the side chains of Arg296, Arg332, Glu431, Asn454 and Cys502 are shown as a stick model. Atomic colors are as follows: oxygen, red; carbon, white; and phosphorus, yellow.

B: Stereoview of the active site of EIC with modelled ZClPEP. ZClPEP and the side chains of Arg296, Arg332, Glu431, Asn454 and Cys502 are shown as a stick model. Atomic colors are as follows: oxygen is shown in red; carbon in white, phosphorus and sulphur yellow.
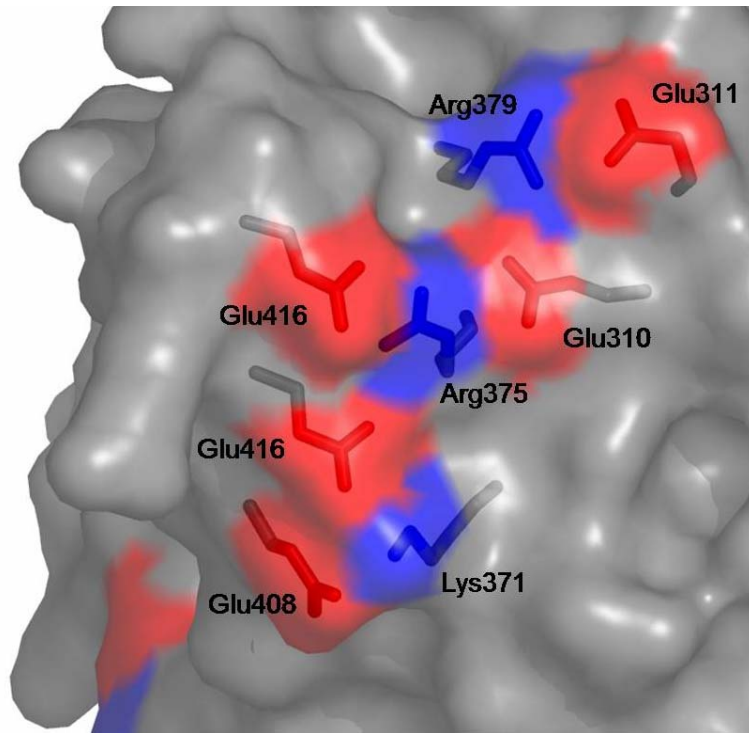
Figure 6: Representation of the 8-member ion-pair network on the surface of EIC from *T. tengcongensis*. Basic residues are coloured in blue, acidic residues are coloured in red.
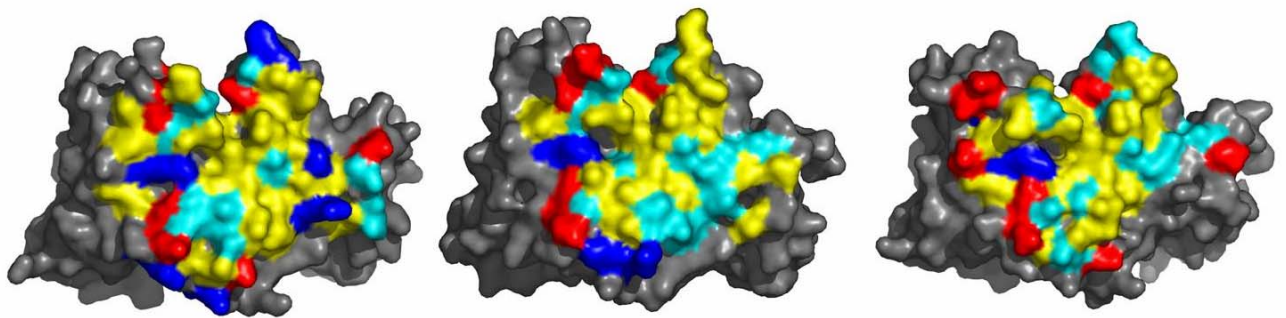

Figure 7: Surface comparison of the dimer interface of EIC of *T. tengcongensis* (left), EIC of *E. coli* (middle) and *M. capricolum* (right). Hydrophobic residues (Ala, Val, Cys, Leu, Ile, Phe, Tyr, Trp, Pro, Met) are coloured in yellow, basic residues (Arg, Lys, His) are coloured in blue, acidic amino acids (Glu and Asp) are coloured in red, polar residues (Asn, Gln, Ser, Thr, Gly) are shown in cyan.

Table 1**:** Data collection and refinement statistics

	Low Remote	Peak	Inflection	Remote
Crystal parameters	P2$_1$2$_1$2$_1$; a = 82.10 Å, b = 91.43 Å, c = 181.86 Å four molecules/ a.s.u.[a]		P2$_1$2$_1$2$_1$; a = 82.82 Å, b = 91.85 Å, c = 185.96 Å four molecules/ a.s.u.[a]	
Data collection (XDS)				
Beamline	BM14	BM14	BM14	BM14
Wavelength (Å)	0.9878	0.9788	0.9791	0.8856
Resolution range[c]	40-1.82 (1.94-1.82)	40-2.8 (2.97-2.80)	40-2.8 (2.97-2.80)	40-2.8 (2.97-2.80)
No. observations	820735	275471[b]	277294[b]	275169[b]
No. unique reflections	123055	66731[b]	67336[b]	66882[b]
Completeness (%)	99.9 (99.7)	98.8 (92.9)	99.7 (98.4)	99.0 (94.3)
R$_{sym}$[d] (%)	6.7 (24.4)	6.1 (18.4)	6.3 (20.3)	7.1 (23.4)
I/σI	16.5 (4.6)	13.8 (5.1)	13.5 (4.8)	12.2 (4.2)
FOM[e] MAD (overall)	0.57			
FOM[f] solvent flattened	0.77			
Refinement (REFMAC)				
Resolution range (Å)	30-1.82			
No. reflections working set	116738			
No. reflections test set	2245			
No. non hydrogen atoms	10958			
Solvent water molecules	837			
R/R$_{free}$ (%)	20.5/24.4			
RMSD bond length (Å)	0.015			
RMSD bond angles (deg.)	1.870			

[a] Asymmetric unit

[b] Friedel pairs were treated as different reflections.

[c] The values in parentheses of resolution range, completeness, R$_{sym}$ and I/σ(I) correspond to the outermost resolution shell

[d] Rsym = $\Sigma_{hkl}\Sigma_j|I(hkl;j) - \langle I(hkl)\rangle|/(\Sigma_{hkl}\Sigma_j\langle I(hkl)\rangle)$ where I(*hkl;j*) is the *j*th measurement of the intensity of the unique reflection (*hkl*) and $\langle I(hkl)\rangle$ is the mean over all symmetry-related measurements.

[e] Figure-of-merit as computed by SOLVE

[f] Figure-of-merit as computed by RESOLVE

Danksagung

Bei meinen Eltern und meinen Schwestern Monya und Melanie möchte ich mich zuerst bedanken, da sie mir durch ihre Unterstützung meinen Werdegang erst ermöglicht haben.

Mein besonderer Dank gilt Herrn Prof. Ulrich Baumann für die freundschaftliche Zusammenarbeit, die hervorragende Betreuung und seine ständige Diskussionsbereitschaft.

Herrn Prof. Hans Trachsel und Herrn Prof. Michael Altmann danke ich für die Unterstützung sowie für viele wertvolle Anregungen.

Herrn Prof. Bernhard Erni möchte ich danken für die Möglichkeit am Enzym I-Projekt mitzuarbeiten.

Mein spezieller Dank gilt Anselm Oberholzer, Siamak Djafarzadeh und Philipp Schneider für die sehr erfolgreiche Zusammenarbeit bei den verschiedenen Projekten.

Bei Anselm Oberholzer, Santina Russo, Reto Meier, Christoph Bieniossek, Dr. Stefan Hörer, Dr. Jörg Benz, Dr. Achim Stocker, Dr. Christian Siebold und Michael Locher bedanke ich mich für die mehr als angenehme Atmosphäre im Labor.

Bei Herrn Dr. Schaller und Urs Kämpfer möchte ich mich für den hervorragenden Service für Massenspektrometrie bzw. Sequenzierung bedanken.

Für die angenehme Arbeitsatmosphäre und die ständige Hilfsbereitschaft danke ich allen Mitarbeiterinnen und Mitarbeitern des Arbeitskreises.

es chunt, so wies chunt und so wies chunt, chunts äbu güät!

Curriculum Vitae

Mario Bumann

Geburtsdatum:	29. April 1969
Heimatort:	Saas-Fee/VS

Ausbildung

1976-1982	Primarschule in Visp
1982-1984	Sekundarschule in Visp
1984-1987	Gymnasium „Spiritus Sanctus" in Brig / Typus B
1987-1990	Handelsmittelschule „Spiritus Sanctus" in Brig: Abschluss mit Handelsdiplom
1990	Rekrutenschule
11/1990-3/1993	Privatschule Ecole Theler in Visp: Abschluss mit eidg. Maturität / Typus E
1993-1994	Studium der Medizin an der Universität Freiburg/FR
1994-1995	Sprachaufenthalte in Frankreich und Neuseeland
1995-2000	Studium der Biologie, biochemischer Richtung an der Universität Bern
11/1998-02/2000	Diplomarbeit am Departement für Chemie und Biochemie der Universität Bern mit dem Titel: „Strukturelle und funktionelle Untersuchungen an Serinhydroxymethyltransferase" Leiter der Arbeit: Prof. Dr. Ulrich Baumann
02/2000	Abschluss als diplomierter Naturwissenschaftler
03/2000-06/2004	Dissertation am Departement für Chemie und Biochemie der Universität Bern mit dem Titel: „From translation initiation factors to enzymes – a crystallographic study" Leiter der Arbeit: Prof. Dr. Ulrich Baumann