FEBS Letters 504 (2001) 104–111 FEBS 25114

Minireview

Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS)

Christian Siebold, Karin Flükiger, Rudolf Beutler, Bernhard Erni*

Departement für Chemie und Biochemie, Universität Bern, CH-3012 Bern, Switzerland

Received 11 June 2001; accepted 26 June 2001

First published online 26 July 2001

Edited by Andreas Engel and Giorgio Semenza

Abstract The glucose transporter of *Escherichia coli* couples translocation with phosphorylation of glucose. The IICB^{Glc} subunit spans the membrane eight times. Split, circularly permuted and cyclized forms of IICB^{Glc} are described. The split variant was 30 times more active when the two proteins were encoded by a dicistronic mRNA than by two genes. The stability and activity of circularly permuted forms was improved when they were expressed as fusion proteins with alkaline phosphatase. Cyclized IICB^{Glc} and IIA^{Glc} were produced in vivo by RecA intein-mediated *trans*-splicing. Purified, cyclized IIA^{Glc} and IICB^{Glc} had 100% and 30% of wild-type glucose phosphotransferase activity, respectively. Cyclized IIA^{Glc} displayed increased stability against temperature and GuHCl-induced unfolding. **© 2001 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.**

Key words: Bacterial phosphotransferase system; Cyclic permutation; Glucose permease; Intein; Membrane protein; Sugar phosphotransferase system; *Escherichia coli*

1. Introduction

1.1. Membrane proteins

Membrane proteins are estimated to account for 10–40% of all proteins in bacteria. The membrane lipid bilayer, however, occupies less than 4% of the cell volume. Therefore, the concentration of individual membrane proteins is low and the space accessible for overexpression is limited. Exceptions are a few proteins in convoluted membranes of increased surface area and volume. Nevertheless, membrane proteins can be overexpressed in large amounts and in a functional form. The membrane bilayer apparently accommodates an excess of proteins by folding up. Insertion of membrane proteins into the lipid bilayer is not spontaneous but is catalyzed by an ATP driven secretion machinery [1]. Folding begins with the insertion of a signal sequence presumably the most N-terminal membrane spanning segment [2]. The sequential insertion of the more distal hydrophobic stop transfer and internal signal sequence then determines the overall topology of the protein.

For purification, membrane proteins and lipids must be separated by solubilization with detergents [3]. In this process, a large proportion of the protein surface becomes buried in the detergent micelle, and only a small surface remains accessible for protein–protein contacts. The scarcity of polar surfaces makes crystallization of membrane proteins difficult. Ostermeier and Michel used conformation-specific antibodies against membrane proteins to increase the polar surface and successfully cocrystallized cytochrome *c* oxidase with a native conformation-specific Fv fragment [4]. More recently, Carola Hunte used the same approach to solve to 2.3 Å the structure of the bc1 complex ([5] and this issue [68]).

Many membrane spanning proteins have extramembranous domains, which can be expressed as soluble recombinant proteins and crystallized. These domains might provide contact surfaces similar to the Fv fragments. However, if too flexible, such a contact might fail to produce a rigid orientation of the membrane domain by the soluble one.

Prompted by the spectacular successes achieved by Ostermeier and Hunte, we wondered whether the rigid protein–antibody interaction could be mimicked by a two-point anchoring of a soluble to a membrane domain. As the first step to strengthen the connection, the C-terminus of a cytoplasmic domain was fused with the N-terminus of a membrane domain by circular permutation and cyclization of the polypeptide chain. The glucose transporter of the phosphoenolpyruvate: (PEP) sugar phosphotransferase system (PTS) of *Escherichia coli* was utilized, because it can be overexpressed and purified efficiently, and because phosphotransferase activity can be assayed more conveniently than vectorial transport.

1.2. The bacterial PTS

Bacteria utilize different transport mechanisms for the uptake of solutes: facilitated diffusion, active transport driven by ATP or ion-gradients and group translocation (Fig. 1A). Group translocation of carbohydrates is mediated by the bacterial PEP: PTS [6]. The PTS catalyzes translocation with concomitant phosphorylation of sugars and hexitols and it also regulates the metabolism in response to the availability of carbohydrates [7,8]. PTSs consist of two cytoplasmic proteins, enzyme I (EI) and HPr, and a variable number of sugar-specific transport complexes (Enzymes II^{sugar}) (Fig. 1B). EI transfers phosphoryl groups from PEP to the phosphoryl carrier protein HPr. HPr then transfers the phosphoryl groups to the different transport complexes. PTSs are ubiquitous in eubacteria but do not occur in archaebacteria and eukaryotes. The amino acid sequences of EI and HPr are strongly conserved in all bacteria. The number and structure of PTS trans-

*Corresponding author. Fax: (41)-31-631 4887.
E-mail address: bernhard.erni@ibc.unibe.ch (B. Erni).

0014-5793/01/$20.00 © 2001 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.
PII: S0014-5793(01)02705-3

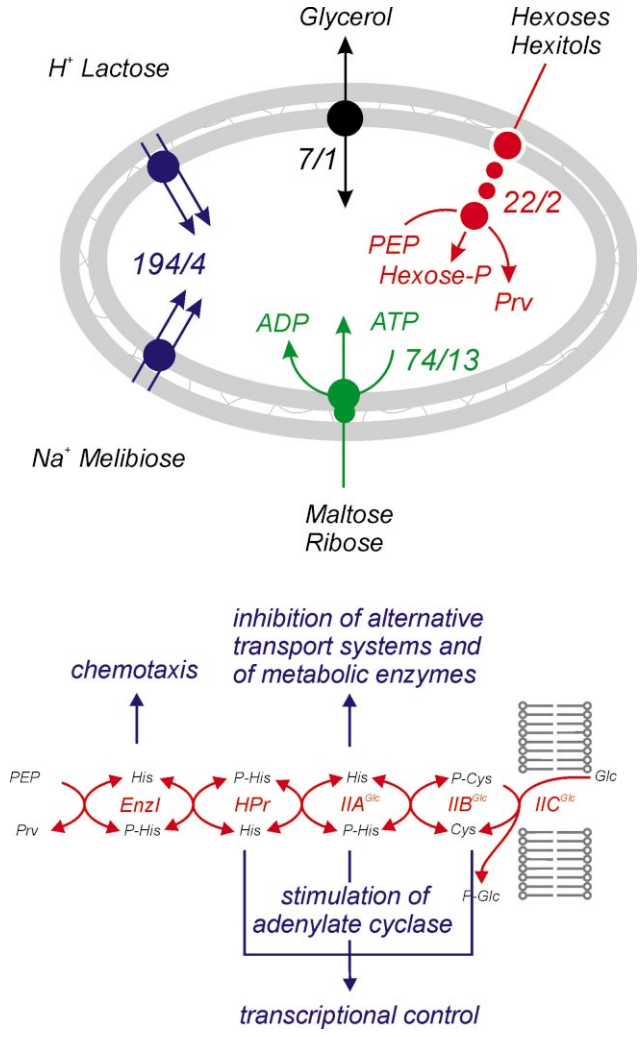

Fig. 1. A: Common transport mechanisms for small solutes in the inner membrane of bacteria. Facilitation (black), secondary ion-symport (blue), primary ATP-dependent transport (green), group translocation (red). The numbers refer to the number of paralogous proteins in *E. coli* (left) and *M. genitalium* (right). The genomes of *E. coli* and *M. genitalium* contain 4300 and 470 genes of which 25% to 40%, and 10% to 30% respectively, encode putative membrane proteins. B: Phosphotransfer chain of the glucose PTS. Phosphoryl groups are sequentially transferred from PEP to EI (63 kDa), HPr (9 kDa) and hence to the carbohydrate transporter IIA^{Glc} (18 kDa) and $IICB^{Glc}$ (51 kDa). Proteins are phosphorylated at His and Cys. The red arrows emphasize the reversibility of the phosphotransfer reaction. The blue arrows indicate regulatory interactions but do not imply direct protein–protein contacts between the PTS components and the regulated targets (see Fig. 4).

porters vary between species. They can be grouped by sequence comparison into four structurally very different families [9]. The *E. coli* genome encodes 38 different PTS proteins, 33 of which are subunits belonging to 22 different transporters. The *Mycoplasma genitalium* genome contains one gene for EI and HPr each and only two genes for PTS transporters (Fig. 1A). The genomes of *Treponema pallidum*, *Chlamydia trachomatis* and *Xylella fastidiosa* (a plant pathogen) contain EI- and HPr-like proteins but no PTS transporters. *Mycobacterium tuberculosis* has no PTS.

PTS transporters consist of three functional units, IIA, IIB and IIC, which occur either as protein subunits in a complex (e.g. IIA^{Glc}·$IICB^{Glc}$) or as domains of a single polypeptide chain (e.g. $IICBA^{GlcNAc}$) (Fig. 2). IIA and IIB sequentially transfer phosphoryl groups from HPr to the transported sugars. IIC contains the sugar binding site. EI, HPr und IIA are phosphorylated at His, IIB domains are phosphorylated at either Cys or His depending on the particular transporter.

The IIA und IIB domains of the different families of PTS transporters have completely different folds (Fig. 3, reviewed in [10]). IIA^{Glc} of *E. coli* is a sandwich of two β-sheets each consisting of six antiparallel β-strands and two very short α-helices. IIA^{Mtl} is a five-stranded mixed β-sheet with α-helices on both faces. IIA^{Man} of *E. coli* is a homodimer of open-twisted α/β subunits. The dimer is stabilized by a swap of β-strands between the subunits. IIA^{Lac} of *Lactococcus lactis* is a homotrimer of three-helix-bundle subunits. IIB^{Glc} consists

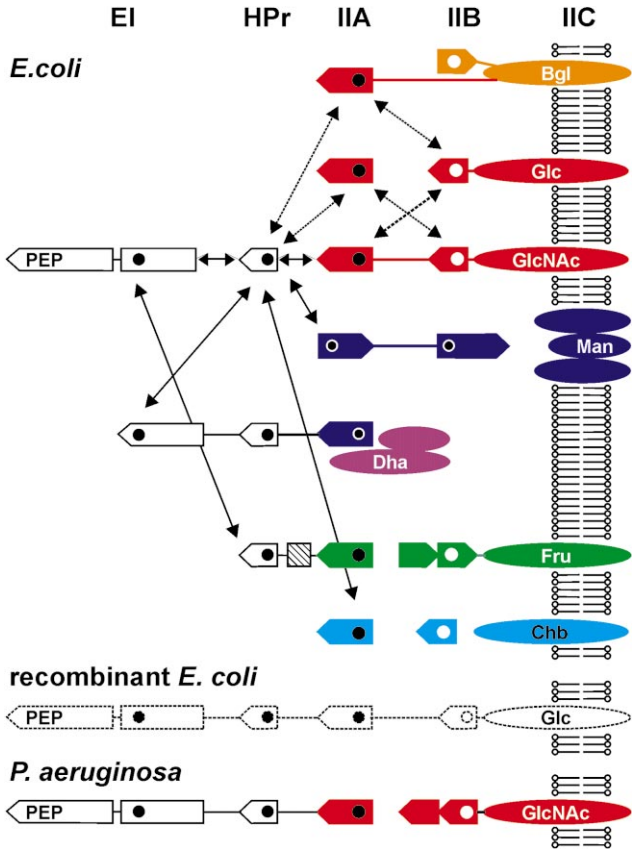

Fig. 2. Modular design of the PTS. Shown are representative examples of the glucose/β-glucoside (red, orange), mannose (blue), fructose/mannitol (green) and lactose/chitobiose (cyan) families from *E. coli* and *P. aeruginosa*. Also shown is the heterotrimeric dihydroxyacetone kinase (purple/blue). The functional units EI, HPr, IIA, IIB and IIC are vertically aligned. The FruA subunit of *E. coli* and $IICB^{GlcNAc}$ of *P. aeruginosa* have duplicated B domains. Solid and open circles indicate phosphorylated His and Cys, respectively. Arrows indicate phosphotransfer between HPr and the sugar-specific IIA domains, and between homologous IIA and IIB domains of different transporters. The recombinant multidomain protein was obtained by connecting the EI, HPr, IIA^{Glc} and $IICB^{Glc}$ units with the Ala–Pro rich linkers taken from the $IIAB^{Man}$ subunit of the mannose transporter [56]. The protein symbols are drawn to scale with the pointed end marking the C-terminus. Swiss-Prot accession codes are: *E. coli* $IIBCA^{Bgl}$, P08722; IIA^{Glc}/$IICB^{Glc}$, P08837/P05053; $IICBA^{GlcNAc}$, P09323; $IIAB^{Man}$/IIC^{Man}/IID^{Man}, P08186/P08187/P08188; Dha Kinase, P37349/P76015/P76015; FruB/FruA, P24217/P20966; IIA^{Cel}/IIC^{Cel}/IIC^{Cel}, P17335/P17409/P17334; *P. aeruginosa* MTP^{GlcNAc}/$IICB^{GlcNAc}$ Q9HXN5/Q9HXN4.

of a four-stranded antiparallel β-sheet covered on one face by three helices. IIB^{Cel} and IIB^{Man} have open-twisted α/β structures with four- and seven-stranded parallel β-sheets, respectively.

Some PTS proteins play a role in intracellular signal transduction in addition to their transport function. These subunits regulate their targets either allosterically or by phosphorylation. Their regulatory activity varies with the degree of phosphorylation (ratio of non-phosphorylated to phosphorylated form) which in turn varies with the ratio of sugar-dependent dephosphorylation and PEP-dependent rephosphorylation [11]. Examples of regulatory functions exerted by the glucose-specific PTS of *E. coli* are summarized in Fig. 4. Dephosphorylated EI inhibits the sensor kinase CheA of the bacterial chemotaxis machinery [12]. The dephosphorylated IIA^{Glc} subunit inhibits glycerolkinase (GK) [13], phosphorylated IIA^{Glc} activates adenylate cyclase (Cya) [14]. Dephosphorylated $IICB^{Glc}$ binds and thereby sequesters the transcriptional repressor protein Mlc. Free Mlc blocks the expression of genes involved in glucose transport and metabolism, and these genes become derepressed whenever glucose is transported and $IICB^{Glc}$ thereby is dephosphorylated [15–18,66,67]. Another regulatory mechanism (not shown) is the (in)activation of transcription factors (repressors and anti-terminators) by PTS protein-mediated phosphorylation (for a review see [19]).

The discovery of new PTS-related sequences in the bacterial genomes continues, and suggests, that PTSs might have additional and as yet unknown functions. For instance, the dihydroxyacetone kinase of *E. coli* utilizes a PTS protein instead of ATP as phosphoryl donor (Fig. 2, [20,21]). EI^{Ntr}, an EI-like protein consisting of a N-terminal NifA and a C-terminal EI-like domain [22] is a virulence factor in the *Pseudomonas aeruginosa* PA14- *Caenorhabditis elegans* slow killing model [23]. EI^{Ntr} mutants are less pathogenic not only in *C. elegans* but also in the plant *Arabidopsis thaliana* and in mice. In addition to EI^{Ntr} there are other PTS proteins which more tentatively have been associated with bacterial virulence [24–27].

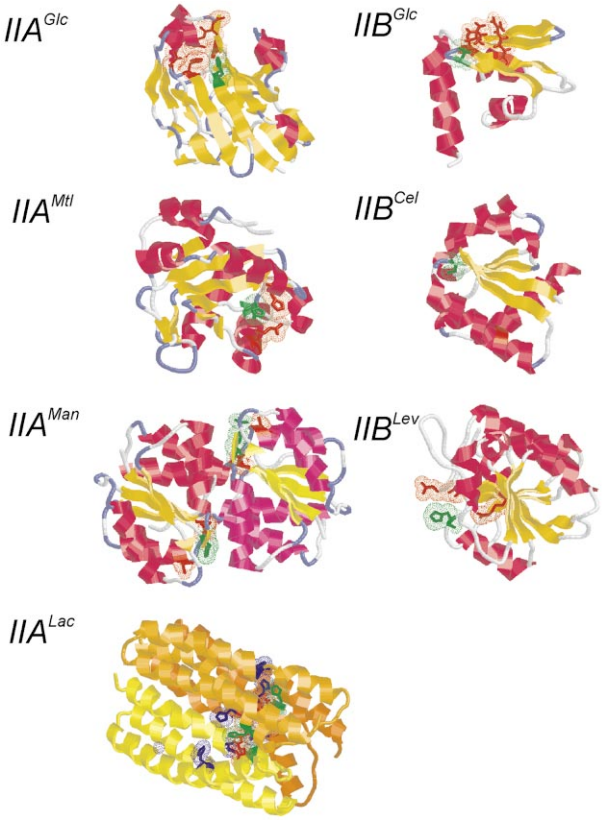

Fig. 3. Diversity of folds. Cartoons of IIA and IIB domains/subunits of different PTS transporters. Phosphorylatable residues are represented as green sticks, other active site residues in red and blue. IIA^{Lac} is a trimer, IIA^{Man} a dimer, the other domains are monomeric. The PDB access codes and references: IIA^{Glc} (1F3G, [57]); IIA^{Mtl} (1A3A, [58]); IIA^{Man} (1PDO, [59]); IIA^{Lac} (1E2A, [60]); IIB^{Glc} (1IBA, [61]); IIB^{Cel} (1IIB, [62]) IIB^{Lev} (1BLE, [63]).

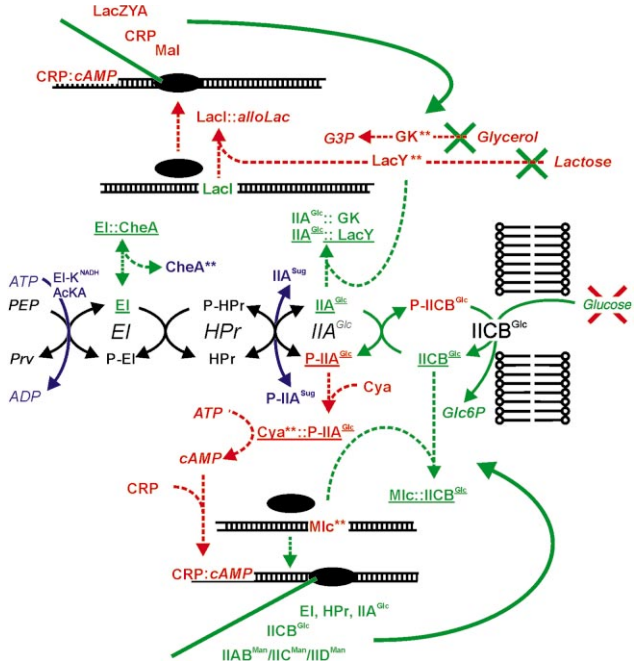

Fig. 4. Energy- and signal transduction by the glucose PTS. Regulatory events occurring in the presence of Glc are emphasized in green, those occurring in its absence in red. In the presence of Glc, a large fraction of IIA^{Glc} and $IICB^{Glc}$ is dephosphorylated in the steady state. $IICB^{Glc}$ binds the Mlc repressor, and genes encoding PTS proteins (general and Glc-specific) can be transcribed by RNA polymerase (black ovals). IIA^{Glc} binds and inhibits the lactose transporter (LacY) and prevents uptake of lactose (inducer exclusion). IIA^{Glc} also binds and inhibits other targets, e.g. GK. The (more) active forms of these targets are labelled with two stars. In the absence of Glc, IIA^{Glc} and $IICB^{Glc}$ mostly are in the phosphorylated form. Mlc is not sequestered by P-$IICB^{Glc}$. Mlc binds to operator DNA and represses the transcription of glucose-specific and other genes. P-IIA^{Glc} does not bind to GK and other targets of IIA^{Glc}, but instead activates Cya. cAMP binds to the catabolite repressor protein (CRP) which when complexed with cAMP activates the transcription of (catabolite repressed) genes. Activation occurs only under the condition that a specific inducer, e.g. lactose is present. CheA, the sensor kinase of the bacterial chemotaxis system is inhibited by dephosphorylated EI. Sources and sinks of phosphate are shown in blue. EI-kinase ($EI\text{-}K^{NADH}$) and acetate kinase (AckA) can feed phosphate from ATP into the phosphotransfer chain [64,65]. IIA^{Sug} are IIA subunits of other transporters (Fig. 2) which can compete with IIA^{Glc} for P-HPr. Work summarized in this figure is from the laboratories of S. Roseman, J. Plumbridge, H. Aiba, P. Postma, M. Saier and J. Lengeler (for references see text).

2. The $IICB^{Glc}$ subunit of the glucose transporter

2.1. Structure and function of $IICB^{Glc}$

The $IICB^{Glc}$ subunit (50.7 kDa) consists of two domains,

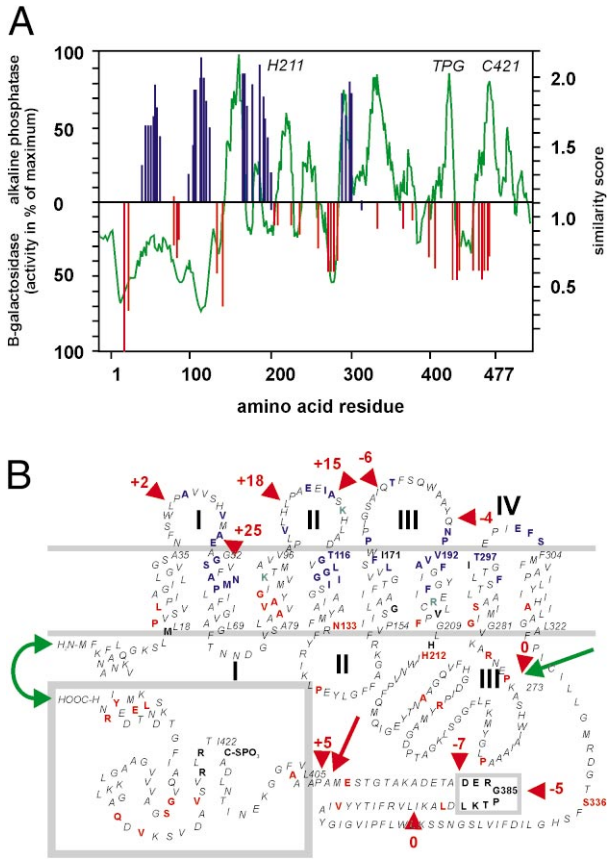

Fig. 5. Topology of the IICB^{Glc} subunit. A: Activity of fusion proteins between IICB^{Glc} and alkaline phosphatase (blue) or β-galactosidase (red). Activities are plotted against the residue at fusion point. Similarity plot of subunits from the glucose family of PTS transporters (green). Important residues at sites of maximum similarity are indicated. The average similarity of the entire sequence was 1.1. B: Transmembrane topology of IICB^{Glc}. Residues fused to alkaline phosphatase and β-galactosidase are blue and red. Red triangles and numbers indicate positions tolerant to insertion/deletions of the indicated length. The green arrow indicates the position at which IICB^{Glc} was split and circularly permuted. The red arrow indicates the chymotryptic cleavage site. The IIB domain, and the invariant motif of the linker region are framed.

the membrane bound C (41.1 kDa) and the cytoplasmic B domain (9.6 kDa). There exist over 30 sequences of orthologous and paralogous proteins with affinity for Glc or GlcNAc. Overall sequence similarities are stronger in the second half of the sequence from cytoplasmic loop 2 to the end (Fig. 5A).

The C domain contains the sugar binding site, as concluded from the specificity of a IICB^{Glc}/IICBA^{GlcNAc} chimera [28], and point mutations conferring relaxed substrate specificity [16,18,29–32]. It comprises eight putative membrane spanning segments (Fig. 5), a topology which was first derived from the activity of fusion proteins between IICB^{Glc} and β-galactosidase or alkaline phosphatase [33] and further confirmed by random linker insertion mutagenesis [34]. Mutants which retained between 50 and 100% activity carried inserts and deletions in the putatively periplasmic loops I, II and III, at the end of the cytoplasmic domain (loop III) and in the linker between the IIC and IIB domain. A projection map of 2D crystals [35] indicates a dimeric structure of IIC^{Glc}, in agreement with results of inter-allelic complementation (Fig. 6B), crosslinking and copurification of heterodimers [36–39].

The IIB domain consists of a four-stranded antiparallel β-sheet and three helices packed against one side of the sheet (Fig. 3). The phosphorylatable Cys421 is highly exposed in the loop between strands β1 and β2 which leads to artifactual disulfide crosslinking of subunits in the purified dimer [37]. Next to Cys421 on the surface of the β-sheet and also invariant in the transporters of the glucose family are Arg424 and Arg426. The two arginines are essential for Glc phosphorylation and transport, but not for the phosphorylation of IICB^{Glc} by IIA^{Glc} [39]. Only the C421S substitution also abolishes phosphorylation of IICB^{Glc} by IIA^{Glc} (Fig. 6A,B, [40]).

IIC and IIB are flexibly linked and complementation between domains of different inactive subunits is possible (Fig. 6B, [39]). At present we do not know, whether complementation occurs between colliding homodimers, or in heterodimers formed by subunit exchange. The interdomain region is susceptible to proteolysis around the position indicated with the red arrow in Fig. 5. Phosphorylation of IICB^{Glc} partially protects against chymotryptic cleavage indicating that the interdomain region undergoes a conformational change upon phosphorylation (Fig. 6C). This region comprises a short motif (LKTPGRED, Fig. 5) which is strongly conserved in all transporters of the glucose family. Surprisingly, most modifi-

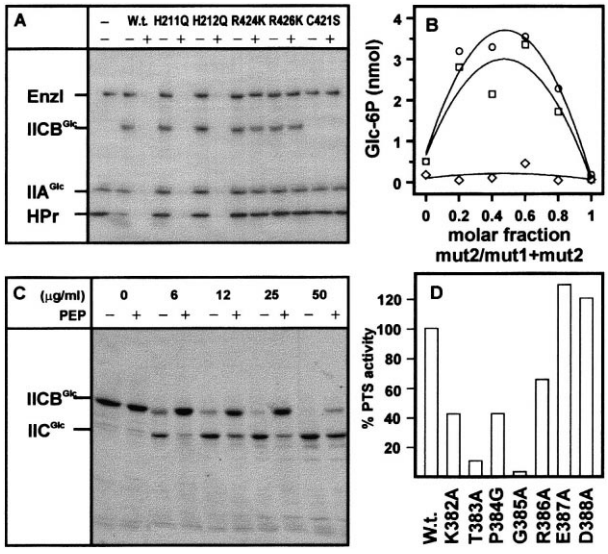

Fig. 6. Functional properties of IICB^{Glc} mutants. (A) Purified IICB^{Glc} mutants indicated at the top were incubated with [^{32}P]PEP and the purified PTS proteins indicated on the left in the absence (−) and presence (+) of Glc in excess over PEP. Mutants which are phosphorylated by IIA^{Glc}, but cannot transfer the phosphoryl group to Glc are labelled. Wild-type and mutants which are active become completely dephosphorylated in the presence of Glc. Shown is the autoradiogram of an polyacrylamide gel. B: Inter-allelic complementation of Glc phosphotransferase activity between purified H212Q/R424K (◇), H212Q/R426K (□) and R424K/R426K (7). The activity maximum corresponds to 10% of wild-type-specific activity. It is not clear whether complementation occurs between subunits in heterodimers or between homodimers of IICB^{Glc}. C: P-IICB^{Glc} is stabilized against chymotryptic cleavage of the interdomain linker region. Purified IICB^{Glc} was incubated for 10 min with (+) and without (−) PEP in the presence of catalytic concentrations of EI, HPr and IIA^{Glc}. 0.5 mg/ml IICB^{Glc} were digested with the indicated concentrations of chymotrypsin for 30 min. Shown is a Coomassie blue-stained gel of the proteolytic products. D: Alanine-scanning mutagenesis of the invariant motif in the interdomain linker region. Specific Glc phosphotransferase activity of the purified proteins in percent of wild-type.

cations at this motif do not compromise transport and phosphorylation activity. The IICB^{Glc} subunit can be split within this motif, and the combined purified domains have in vitro glucose phosphotransferase activity [41]. The motif also is the only crossover point for functional chimeric proteins between IICB^{Glc} and IICBA^{GlcNAc}, whereas the chimeras with fusion joints at Ser45, Ala68, Pro154 have spurious activity only ([28] and unpublished observations). Alanine mutations in the motif either have no effect or reduce PTS activity to between 10% and 60% (Fig. 6D, [39]). Only the G385A substitution and a deletion of the motif reduce phosphorylation and uptake activity to below 1% of wild-type. What function then does this motif have? One hypothesis suggests, that it is (part of) the binding site for the Mlc repressor (Fig. 3) in response to the phosphorylation state of IICB^{Glc} [18,42].

2.2. Mutants of IICB^{Glc}

Substrate specificity mutants catalyze phosphorylation and uptake of additional sugars (e.g. GlcNAc and Man) [16,18,29] or permit facilitated diffusion of substrates (ribose, fructose, mannitol) without concomitant phosphorylation [30–32]. The substrate specificities are only relaxed and uncoupling between translocation and phosphorylation is conditional. This means that the mutants retain nearly normal phosphorylation and transport activity towards Glc. K_M for the non-cognate substrates is high and the V_{max} low. Another group of mutants has reduced transport activity (less than 5%) for Glc but almost normal phosphorylation activity (20–70%). Uptake of α-methylglucoside is 4–10× more strongly inhibited than uptake of Glc, most likely because these mutants were selected against uptake of the toxic analog α-methylglucoside and for phosphorylation of intracellular glucose [43]. The H112Q mutation blocks transport but not phosphorylation. In H111Q both activities are moderately reduced. The mutants described above are distributed all over the IIC domain, and an obvious correlation between structure and function has yet to emerge.

3. Split, circularly permuted and cyclic forms of IICB^{Glc}

With the objective to stabilize the interaction between the IIC and IIB domains, circularly permuted [44] and cyclized variants of IICB^{Glc} (C. Siebold, unpublished results) were constructed. In an attempt to break the membrane spanning IIC domain into smaller parts, variants split in the IIC domain were also characterized [34].

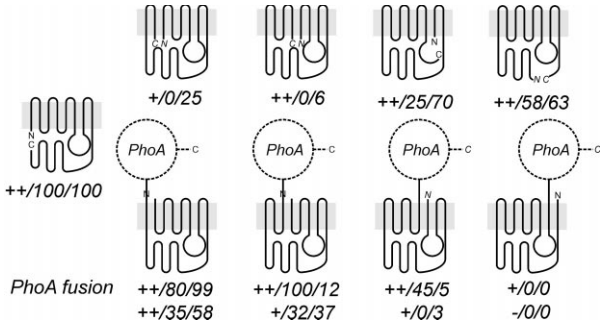

Fig. 7. Topology and activity of the circularly permuted variants of IICB^{Glc}. The + symbol indicates the amount of protein detected with an anti-IICB^{Glc} antibody on Western blots. The numbers indicate in percent of wild-type: transport activity of intact cells expressing the permuted IICB^{Glc}/specific Glc phosphotransferase activity of membrane fractions. The values are relative to total protein and are not corrected for the different expression level of IICB^{Glc}.

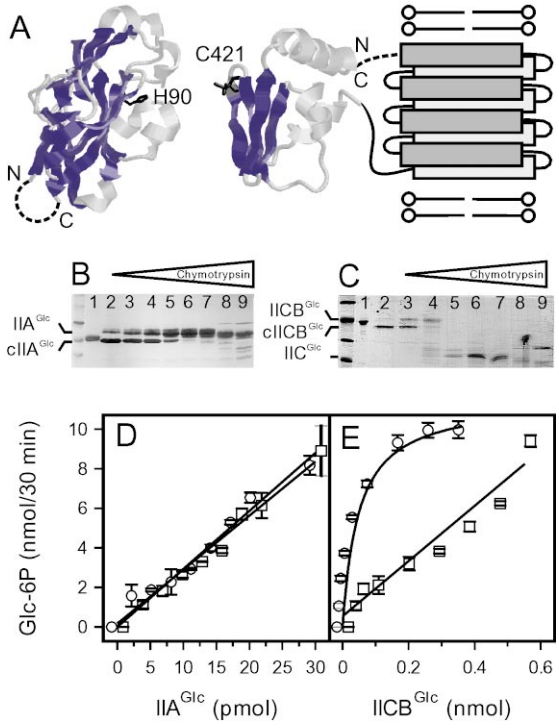

Fig. 8. Cyclized proteins. A: Model of the IIA^{Glc} and IICB^{Glc} subunits of the glucose transporter. The active site Cys421 of IICB^{Glc} (1IBA) and the active site His90 of IIA^{Glc} (PDB code 2F3G) are shown in stick representation. The sites of cyclization are shown as broken lines between the native N- and C-termini. (B, C) Linearization of cyclized IIA^{Glc} and IICB^{Glc} with chymotrypsin. Note the decrease of electrophoretic mobility upon linearization of the cyclic forms. Lanes 1: purified wild-type IIA^{Glc} and IICB^{Glc} (references). Lanes 2: No chymotrypsin. Lanes 3–8: with increasing concentrations of chymotrypsin (2 to 500 μg/ml in B; 2 ng/ml to 167 μg/ml in C. D, E: Glucose phosphotransferase activity purified cyclized IIA^{Glc} (D) and IICB^{Glc} (E). Cyclized forms (squares), wild-type forms (circles). Shown are means ± S.D. of two independent assays.

3.1. Split variants

The IIC domain was split at the sites which had previously been shown to tolerate amino acid insertions and deletions (Fig. 5B, Section 2.1). A discontinuity of the polypeptide at these sites reduced the phosphotransferase activity to between 7% and 32%. The variant split in the C-terminal end of the cytoplasmic domain (loop III) had the highest activity and was purified. The two polypeptides copurified and could not be separated without denaturation of the complex. The split variant is 30× more active when the proteins are encoded by a bicistronic operon than by genes on two compatible plasmids [34]. This indicates that spatial proximity during protein synthesis might favor assembly and membrane insertion of the two fragments. It is not known whether the two halves independently insert into the membrane and then assemble, or whether they associate first and then insert as a complex.

3.2. Circularly permuted variants

Native N- and C-termini of IICB^{Glc} were joined by an Ala–Pro rich linker and new N- and C-termini were introduced in all seven periplasmic and cytoplasmic loops (Fig. 7). Variants

with the termini in the periplasmic loops I and II had between 30% and 50% activity. The variants with termini in putative periplasmic loops III and IV were not stable. Of the variants with termini on the cytoplasmic side, only the one modified in loop III was active. Fusion of alkaline phosphatase to the C-termini in the periplasm stabilized membrane integration and increased activity, in particular of the variant with the N-terminus in periplasmic loop I (Fig. 7 bottom row). These results further confirm that internal signal anchor and stop transfer sequences can function as N-terminal signal sequences, and that the orientation of transmembrane segments is determined by the amino acid sequence and not by their sequential appearance during translation [45,46]. It appears that alkaline phosphatase helps to anchor the protein in the membrane possibly by preventing the last segment from slipping back. In the circularly permuted fusion protein the IIC domain is now sandwiched between two soluble domains, an 'immobilized' B-domain on the cytoplasmic and alkaline phosphatase on the periplasmic side. Whether they help protein crystallization as did the Fv fragments, remains to be demonstrated.

3.3. Cyclic $IICB^{Glc}$ and IIA^{Glc}

The N- and C-termini of proteins are often close together and have their surface exposed [47] and, therefore, can be joined together to afford a cyclized protein. Such proteins

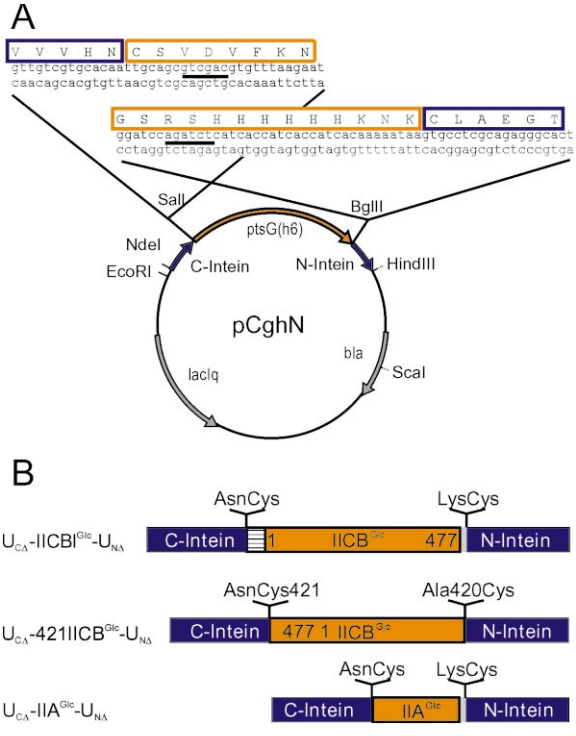

Fig. 9. Plasmid map and structure of the circularly permuted intein-$IICB^{Glc}$ (purple-orange) precursor. A: Plasmid map. The nucleotide and amino acid sequences of the splice sites are indicated. The two cysteine at the Asn/Cys and Lys/Cys splice junction catalyze the S–N acyl shift reactions. B: Structure of precursor proteins. The $IICB^{Glc}$ precursor contains an Ala–Pro linker (striped) and a His-tag (gray) at the splice junction. The $IICB^{Glc}$ variant with the active site Cys421 at the N-terminal splice junction did not afford a circularized protein. The IIA^{Glc} variant contained a His-tag only. Residue numbers refer to the $IICB^{Glc}$ wild-type sequence. (Swiss-Prot accession number P05053; IIA^{Glc} P08837).

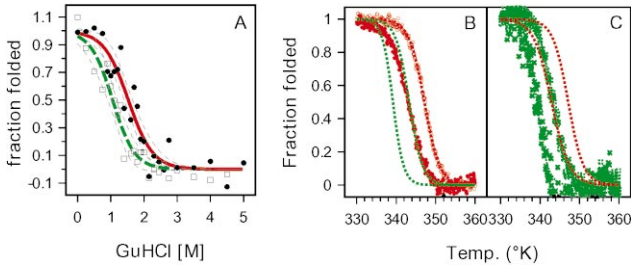

Fig. 10. GuHCl- and temperature-induced unfolding of cyclized and wild-type IIA^{Glc}. A: Cyclized IIA^{Glc} (red, closed circles); wild-type IIA^{Glc} (green, open squares). The smoothed curves show the mean and the 95% confidence band (broken lines). B: Cyclized IIA^{Glc}: Unfolding transition (red, open circles), refolding transition (red, closed circles). C: Wild-type IIA^{Glc}: unfolding transition (green, open squares), refolding transition (green, closed squares). To facilitate comparison, the smoothed curves drawn through the data points of C are also plotted in B and vice versa in C (broken lines).

are expected to be resistant against N- and C-peptidases and to be more stable because of the lower entropy of their unfolded states. Partial immobilization of unstructured ends by crosslinking could facilitate protein crystallization. Cyclized $IICB^{Glc}$ and IIA^{Glc} subunits were produced by in vivo splicing and characterized (Fig. 8A; C. Siebold, unpublished results).

Protein splicing is a posttranslational process, by which an internal sequence (intein) of a precursor protein is removed by cleavage of two peptide bonds and the two flanking sequences (N-extein and C-extein) are ligated through a native peptide bond. The splice reaction involves four nucleophilic substitution reactions (N–S, S–S, N–N and S–N acyl shifts) and is catalyzed by the intein (for reviews see [48,49]). The N- and C-exteins are the substrates of this reaction but do not have splicing activity of their own. Inteins can be used to prepare cyclized proteins [50–52] because the central portion of the intein is not necessary for function and the flanking regions do not need to be linked covalently for splicing activity. A circularly permuted linear precursor comprising in this order the C-terminal half of the *recA* intein of *M. tuberculosis*, $IICB^{Glc}$ or IIA^{Glc} to be cyclized and the N-terminal half of the *recA* intein were produced in vivo using a recombinant gene constructed according to previous methods [53–55]. A 20 residue long Ala–Pro rich linker peptide and a histidine tag were used to join the native N- and C-termini (Fig. 9).

The cyclic forms of IIA^{Glc} and $IICB^{Glc}$ complemented growth of glucose auxotrophic strains. Precursor and processed forms of $IICB^{Glc}$ were in the membrane fraction. The cyclized form could be solubilized completely in *n*-dodecyl-maltoside whereas the precursor was more or less insoluble in the non-ionic detergent. The linear IIA^{Glc} precursor was strongly expressed but most of it was in inclusion bodies. Cyclized IIA^{Glc}, however, was soluble. The cyclic proteins could be purified by metal chelate affinity chromatography with a yield of 10% compared to the wild-type proteins. Cyclic $IICB^{Glc}$ had 30%, cyclic IIA^{Glc} 100% of wild-type glucose phosphotransferase activity (Fig. 8D,E). Both proteins had an increased electrophoretic mobility which decreased upon linearization of the proteins with chymotrypsin (Fig. 8B,C). The stabilities of cyclized and linear IIA^{Glc} were compared by temperature and GuHCl-induced unfolding (Fig. 10). The transition midpoint for GuHCl-induced unfolding was shifted

from 1.1 M to 1.5 M for cyclized IIA^{Glc} (Fig. 6A). The transition temperature of unfolding was 75°C for cyclized IIA^{Glc} and 70°C for linear IIA^{Glc} (Fig. 10B,C). The midpoint temperatures of refolding were 71°C and 66°C respectively. Phosphotransferase activity of cyclized and linear proteins was recovered completely after cooling, indicating that both assumed the catalytically active fold.

In conclusion, IIA^{Glc} can be stabilized by cyclization. However the yield of cyclized protein is too low to start crystallization experiments with it. The objective of future experiments is to improve the efficiency of cyclization. Results reported for DnaE-mediated in vivo cyclization [51] suggest that this naturally split intein may be more suitable because its two halves have an intrinsically strong affinity. Alternatively, the association between the two RecA intein fragments might be strengthened by fusion of leucine zippers to the N- and C-termini of the precursor.

Acknowledgements: This research was supported by the Swiss National Science Foundation (Grant 31-45838.95), ARPIDA AG Münchenstein, and Ciba-Geigy Jubiläumsstiftung, Basel.

References

[1] De, G.J. and Luirink, J. (2001) Mol. Microbiol. 40, 314–322.
[2] Koch, H.G., Hengelage, T., Neumann-Haefelin, C., MacFarlane, J., Hoffschulte, H.K., Schimz, K.L., Mechler, B. and Muller, M. (1999) Mol. Biol. Cell 10, 2163–2173.
[3] Le, M.M., Champeil, P. and Moller, J.V. (2000) Biochim. Biophys. Acta 1508, 86–111.
[4] Ostermeier, C., Iwata, S., Ludwig, B. and Michel, H. (1995) Nat. Struct. Biol. 2, 842–846.
[5] Hunte, C., Koepke, J., Lange, C., Rossmanith, T. and Michel, H. (2000) Struct. Fold. Des. 8, 669–684.
[6] Kundig, W., Gosh, S. and Roseman, S. (1964) Proc. Natl. Acad. Sci. USA 52, 1067–1074.
[7] Postma, P.W., Lengeler, J.W. and Jacobson, G.R. (1993) Microbiol. Rev. 57, 543–594.
[8] Postma, P.W., Lengeler, J.W. and Jacobson, G.R. (1996) in: *Escherichia coli* and *Salmonella*: cellular and molecular biology (Neidhardt F.C., et al., Eds.), p. 1149, ASM, Washington, DC.
[9] Paulsen, I.T., Nguyen, L., Sliwinski, M.K., Rabus, R. and Saier, M.H. (2000) J. Mol. Biol. 301, 75–100.
[10] Robillard, G.T. and Broos, J. (1999) Biochim. Biophys. Acta 1422, 73–104.
[11] Saier, M.H., Chauvaux, S., Deutscher, J., Reizer, J. and Ye, J.-J. (1995) Trends Biochem. Sci. 20, 267–271.
[12] Lux, R., Jahreis, K., Bettenbrock, K., Parkinson, J.S. and Lengeler, J.W. (1995) Proc. Natl. Acad. Sci. USA 92, 11583–11587.
[13] Hurley, J.H., Faber, H.R., Worthylake, D., Meadow, N.D., Roseman, S., Pettigrew, D.W. and Remington, S.J. (1993) Science 259, 673–677.
[14] Peterkofsky, A., Seok, Y.-J., Amin, N., Thapar, R., Lee, S.Y., Klevit, R.E., Waygood, E.B., Anderson, J.W., Gruschus, J., Huq, H. and Gollop, N. (1995) Biochemistry 34, 8950–8959.
[15] Kim, S.Y., Nam, T.W., Shin, D., Koo, B.M., Seok, Y.J. and Ryu, S. (1999) J. Biol. Chem. 274, 25398–25402.
[16] Notley-McRobb, L. and Ferenci, T. (2000) J. Bacteriol. 182, 4437–4442.
[17] Plumbridge, J. (1999) Mol. Microbiol. 33, 260–273.
[18] Zeppenfeld, T., Larisch, C., Lengeler, J.W. and Jahreis, K. (2000) J. Bacteriol. 182, 4443–4452.
[19] Stülke, J., Arnaud, M., Rapoport, G. and Martin-Verstraete, I. (1998) Mol. Microbiol. 28, 865–874.
[20] Beutler, R., Kämpfer, U., Schaller, J. and Erni, B. (2001) Microbiology 147, 249–250.
[21] Gutknecht, R., Beutler, R., Garcia Alles, L.F., Baumann, U. and Erni, B. (2001) EMBO J. 20, 2481–2486.
[22] Powell, B.S., Court, D.L., Inada, T., Nakamura, Y., Michotey, V., Cui, X., Reizer, A., Saier, M.H. and Reizer, J. (1995) J. Biol. Chem. 270, 4822–4839.
[23] Tan, M.W., Rahme, L.G., Sternberg, J.A., Tompkins, R.G. and Ausubel, F.M. (1999) Proc. Natl. Acad. Sci. USA 96, 2408–2413.
[24] Khan, M.A. and Isaacson, R.E. (1998) J. Bacteriol. 180, 4746–4749.
[25] Lycheva, T.A., Bondarenko, V.M., Burd, G.I., Umiarov, A.M. and Bol'shakova, T.N. (1980) Zh. Mikrobiol. Epidemiol. Immunobiol. 58–63.
[26] Saier, M.H. and Chin, A.M. (1990) in: Bacterial energetics (Krulwich, T.A., Ed.), p. 273, Academic Press, San Diego, CA.
[27] Turner, A.K., Lovell, M.A., Hulme, S.D., Zhang-Barber, L. and Barrow, P.A. (1998) Infect. Immun. 66, 2099–2106.
[28] Hummel, U., Nuoffer, C., Zanolari, B. and Erni, B. (1992) Protein Sci. 1, 356–362.
[29] Begley, G.S., Warner, K.A., Arents, J.C., Postma, P.W. and Jacobson, G.R. (1996) J. Bacteriol. 178, 940–942.
[30] Ruijter, G.J.G., van Meurs, G., Verwey, M.A., Postma, P.W. and Van Dam, K. (1992) J. Bacteriol. 174, 2843–2850.
[31] Oh, H., Park, Y. and Park, C. (1999) J. Biol. Chem. 274, 14006–14011.
[32] Kornberg, H.L., Lambourne, L.T. and Sproul, A.A. (2000) Proc. Natl. Acad. Sci. USA 97, 1808–1812.
[33] Buhr, A. and Erni, B. (1993) J. Biol. Chem. 268, 11599–11603.
[34] Beutler, R., Kaufmann, M., Ruggiero, F. and Erni, B. (2000) Biochemistry 39, 3745–3750.
[35] Zhuang, J.P., Gutknecht, R., Flükiger, K., Hasler, L., Erni, B. and Engel, A. (1999) Arch. Biochem. Biophys. 372, 89–96.
[36] Erni, B. (1986) Biochemistry 25, 305–312.
[37] Meins, M., Zanolari, B., Rosenbusch, J.P. and Erni, B. (1988) J. Biol. Chem. 263, 12986–12993.
[38] Waeber, U., Buhr, A., Schunk, T. and Erni, B. (1993) FEBS Lett. 324, 109–112.
[39] Lanz, R. and Erni, B. (1998) J. Biol. Chem. 273, 12239–12243.
[40] Nuoffer, C., Zanolari, B. and Erni, B. (1988) J. Biol. Chem. 263, 6647–6655.
[41] Buhr, A., Flükiger, K. and Erni, B. (1994) J. Biol. Chem. 269, 23437–23443.
[42] Lee, S.J., Boos, W., Bouché, J.P. and Plumbridge, J. (2000) EMBO J. 19, 5353–5361.
[43] Buhr, A., Daniels, G.A. and Erni, B. (1992) J. Biol. Chem. 267, 3847–3851.
[44] Beutler, R., Ruggiero, F. and Erni, B. (2000) Proc. Natl. Acad. Sci. USA 97, 1477–1482.
[45] Gafvelin, G. and Von Heijne, G. (1994) Cell 77, 401–412.
[46] Sato, M., Hresko, R. and Mueckler, M. (1998) J. Biol. Chem. 273, 25203–25208.
[47] Thornton, J.M. and Sibanda, B.L. (1983) J. Mol. Biol. 167, 443–460.
[48] Liu, X.Q. (2000) Annu. Rev. Genet. 34, 61–76.
[49] Paulus, H. (2000) Annu. Rev. Biochem. 69, 447–496.
[50] Iwai, H. and Pluckthun, A. (1999) FEBS Lett. 459, 166–172.
[51] Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. and Benkovic, S.J. (1999) Proc. Natl. Acad. Sci. USA 96, 13638–13643.
[52] Evans, T.C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X.Q. and Xu, M.Q. (2000) J. Biol. Chem. 275, 9091–9094.
[53] Lew, B.M., Mills, K.V. and Paulus, H. (1999) Biopolymers 51, 355–362.
[54] Mills, K.V., Lew, B.M., Jiang, S.Q. and Paulus, H. (1998) Proc. Natl. Acad. Sci. USA 95, 3543–3548.
[55] Shingledecker, K., Jiang, S.Q. and Paulus, H. (1998) Gene 207, 187–195.
[56] Mao, Q., Schunk, T., Gerber, B. and Erni, B. (1995) J. Biol. Chem. 270, 18295–18300.
[57] Worthylake, D., Meadow, N.D., Roseman, S., Liao, D.I., Herzberg, O. and Remington, S.J. (1991) Proc. Natl. Acad. Sci. USA 88, 10382–10386.
[58] Van Montfort, R.L.M., Pijning, T., Kalk, K.H., Hangyi, I., Kouwijzer, M.L.C.E., Robillard, G.T. and Dijkstra, B.W. (1998) Structure 6, 377–388.
[59] Nunn, R.S., Markovic-Housley, Z., Génovésio-Taverne, J.C., Flükiger, K., Rizkallah, P.J., Jansonius, J.N., Schirmer, T. and Erni, B. (1996) J. Mol. Biol. 259, 502–511.

[60] Sliz, P., Engelmann, R., Hengstenberg, W. and Pai, E.F. (1997) Structure 5, 775–788.
[61] Eberstadt, M., Grdadolnik, S.G., Gemmecker, G., Kessler, H., Buhr, A. and Erni, B. (1996) Biochemistry 35, 11286–11292.
[62] Van Montfort, R.L.M., Pijning, T., Kalk, K.H., Reizer, J., Saier, M.H., Thunnissen, M.M.G.M., Robillard, G.T. and Dijkstra, B.W. (1997) Structure 5, 217–225.
[63] Schauder, S., Nunn, R.S., Lanz, R., Erni, B. and Schirmer, T. (1998) J. Mol. Biol. 276, 591–602.
[64] Dannelly, H.K. and Roseman, S. (1992) Proc. Natl. Acad. Sci. USA 89, 11274–11276.
[65] Fox, D.K., Meadow, N.D. and Roseman, S. (1986) J. Biol.Chem. 261, 13498–13503.
[66] Nam, T.W., Cho, S.H., Shin, D., Kim, J.H., Jeong, J.Y., Lee, J.H., Roe, J.H., Peterkofsky, A., Kang, S.O., Ryu, S. and Seok, Y.J. (2001) EMBO J. 20, 491–498.
[67] Tanaka, Y., Kimata, K. and Aiba, H. (2000) EMBO J. 19, 5344–5352.
[68] Hunte, C. (2001) FEBS Lett. 504, 126–132.