The eubacterial DNA replicases have a highly conserved organisation with three main components – the replicative DNA polymerase, a sliding clamp (β2) and the clamp loader (γ complex), constituting the DNA polymerase III holoenzyme. The sliding clamp also interacts with a large number of other enzymes involved in DNA replication and repair and appears to be a key node in the protein interaction network associated with these functions (Fig. 1). Given the remarkable functional and structural conservation of the sliding clamps in both prokaryote and eukaryotic organisms, it is not surprising to observe good primary sequence conservation of the β2 subunit in eubacterial species. Consistent with this, all β2 subunits isolated or expressed thus far have been shown to be stable homodimers (1-3; Kongsuwan and Wijffels, unpublished data).

Clamp loader complexes
The sliding clamp, β2, is loaded by the clamp loader, which in E. coli consists of a minimal assembly of a number of subunits: (γ/τ)1-3δδ'. Orthologues of the δ' and τ/γ subunits have been easily identified in all eubacteria, but it is only very recently that orthologues of the δ subunit have been recognised in eubacteria outside of the gamma and beta subdivisions of the proteobacteria (1, 4). Indeed, although the δ subunit is required for loading β2 onto DNA, it is the least conserved of the essential components of the Pol III holoenzyme.

The assembly of the components into clamp loaders has also been demonstrated for Streptococcus pyogenes – a Gram positive organism (1), Helicobacter pylori (Kongsuwan, unpublished data) and two thermophiles, Thermus thermophilus (3) and Aquifex...
Table 1. \( \beta_2 \)-binding proteins.
The proteins are listed by family, function, gene location and copy number, the most common \( \beta_2 \)-binding motif in their family, the similarity of the motif to the consensus pentameric motif, QL(S/D)LF, and the location of the \( \beta_2 \)-binding motif in the protein. Abbreviations: na, not applicable; nd, not determined.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Function</th>
<th>Gene location</th>
<th>( \beta_2 )-binding motif</th>
<th>Motif ranking</th>
<th>Motif location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>Clamp loading</td>
<td>Single copy</td>
<td>SLF</td>
<td>na</td>
<td>Internal</td>
</tr>
<tr>
<td>DnaE1 – most species</td>
<td>Major replicative DNA polymerase</td>
<td>Single copy</td>
<td>QxDLF</td>
<td>Medium</td>
<td>Internal</td>
</tr>
<tr>
<td>DnaE1 – PolC containing species</td>
<td>Lagging strand replicative DNA polymerase and repair</td>
<td>Single copy</td>
<td>xxSLF</td>
<td>Low</td>
<td>Internal</td>
</tr>
<tr>
<td>DnaE2</td>
<td>Repair DNA polymerase</td>
<td>May be multiple copies, some episomal</td>
<td>QLPLF</td>
<td>Low</td>
<td>Internal</td>
</tr>
<tr>
<td>DnaE3</td>
<td>?</td>
<td>? Some episomal</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PolC</td>
<td>Major replicative DNA polymerase - leading strand</td>
<td></td>
<td>QLSLF</td>
<td>High</td>
<td>C-terminus</td>
</tr>
<tr>
<td>PolB</td>
<td>Repair DNA polymerase</td>
<td>Some episomal</td>
<td>QLGLF</td>
<td>High</td>
<td>C-terminus</td>
</tr>
<tr>
<td>UmuC</td>
<td>Repair DNA polymerase</td>
<td>Some episomal</td>
<td>QLNLF</td>
<td>High</td>
<td>Internal</td>
</tr>
<tr>
<td>DinB1</td>
<td>Repair DNA polymerase</td>
<td>Some episomal</td>
<td>QLPLF</td>
<td>Low</td>
<td>Internal, or C-terminus</td>
</tr>
<tr>
<td>DinB2</td>
<td>Repair DNA polymerase</td>
<td>Some episomal</td>
<td>QLSLF</td>
<td>High</td>
<td>Internal</td>
</tr>
<tr>
<td>DinB3</td>
<td>Repair DNA polymerase</td>
<td>Some episomal</td>
<td>QLSLF</td>
<td>Low</td>
<td>Internal</td>
</tr>
<tr>
<td>MutS</td>
<td>Mismatch repair</td>
<td></td>
<td>QLGLF</td>
<td>High</td>
<td>C-terminus</td>
</tr>
<tr>
<td>Hda</td>
<td>Initiation of replication</td>
<td></td>
<td>QLPLPL</td>
<td>Medium</td>
<td>N-terminus</td>
</tr>
<tr>
<td>Duf72</td>
<td>Unknown</td>
<td></td>
<td>QLPLF</td>
<td>High</td>
<td>C-terminus</td>
</tr>
<tr>
<td>RepA families</td>
<td>Initiation of replication</td>
<td>Episomal</td>
<td>Various</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>UmuD</td>
<td>Repair DNA polymerase subunit</td>
<td>Episomal</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DnaA</td>
<td>Initiation of replication</td>
<td>Episomal</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligase</td>
<td>Ligase</td>
<td>Episomal</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PolA</td>
<td>Repair DNA polymerase</td>
<td>Episomal</td>
<td>nd</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \alpha \)aeolicus (2). The distribution of the \( \gamma \) subunit (a truncated version of \( \tau \)) is variable. \( S. pyogenes, A. aeolicus \) and \( B. subtilis \) express only the complete DnaX product, the \( \tau \) subunit (1, 2, 5). On the other hand, the \( \gamma \) subunit is seen in \( T. thermophilus \), a number of Enterobacteriaceae, while \( A. aeromonas \) and \( V. vibrio \) may express an intermediate form (3, 6). The use of combinations of \( \tau \) and \( \gamma \) subunits in the clamp loaders may impact on its interactions with the replicative polymerases within the replisome complex.

Orthologues of the two smaller subunits of \( E. coli \) clamp loader, \( \chi \) and \( \psi \), or proteins taking a similar role, have only been observed in eubacterial species very closely related to \( E. coli \).

The DNA polymerase subunit

The \( \alpha \) subunit (expressed from the \( dnaE \) gene) is the main replicative DNA polymerase in \( E. coli \). It forms a core enzyme with the \( \theta \) and \( \epsilon \) (proofreading exonuclease) subunits. In \( T. thermophilus \), the \( \alpha \) subunit is devoid of both \( \epsilon \) and \( \theta \) subunits, whereas in \( A. aeolicus \), it has only a weak interaction with \( \epsilon \) (2, 3). The \( \theta \) subunit is not widespread in bacteria, but the \( \epsilon \) subunit (or similar proteins) is common. What does the lack of interaction with the exonuclease subunit suggest for the fidelity of replication in these species?

A second major replicative DNA polymerase

A related DNA polymerase, PolC, appears to be the major replicative polymerase in many Gram positive species. This enzyme possesses its own proofreading activity, is highly processive, and is essential in organisms containing the gene. The interaction of PolC with \( \tau \) would appear to be weaker than the binding between the \( \alpha \) and \( \tau \) subunits (1), suggesting that \( \tau \) does not secure this polymerase as tightly to the replisome. In low GC-content Gram positive bacterial species, both polymerases are essential. PolC is required for leading strand synthesis, while the DnaE polymerase, which is closely related to the \( E. coli \) \( \alpha \) subunit, is proposed to be responsible for lagging strand synthesis (7). Furthermore, the DnaE enzyme has been observed to be far less processive and more error prone in these species; a very different profile to that of the \( \alpha \) subunit of non-PolC containing bacteria, suggesting an additional role in DNA repair (8, 9).
The $\beta_2$ sliding clamp is used by many DNA polymerases and DNA repair proteins

It has only been recently recognised that throughout the life of a cell, the $\beta_2$ sliding clamp is harnessed by many DNA polymerases and other proteins involved in the synthesis and integrity of double stranded DNA. In fact, $\beta_2$, with its ability to slide along duplex DNA and to stall at sites of DNA damage, may act as a physical screening device that stops to recruit the appropriate repair complex where DNA integrity is compromised. The $\beta_2$ dimer interacts with several enzymes active in excision repair, mismatch repair and translesion DNA synthesis. These protein families include Pol II (or PolB), Pol IV (or DinB) and Pol V (UmuD and UmuC subunits), as well as MutS, a mismatch recognition protein (reviewed in ref. 10) (Fig. 1 and Table 1). Many of these proteins are induced during the SOS (DNA damage) response.

$\beta_2$-binding sites

By combining the experimental identification of $\beta_2$ binding proteins with the extensive protein sequence information provided by the eubacterial emerging genome sequences, and by analogy with known clamp binding motifs in the eukaryotes and Archea, we predicted that many $\beta_2$-binding proteins attach to $\beta_2$ via a short pentameric peptide sequence with the consensus sequence of QL[S/D]LF (11, 12). Small regions of the $\alpha$ subunit and a number of other proteins containing their putative $\beta_2$-binding sequences were shown to bind $\beta_2$ using a yeast two-hybrid system, and synthetic peptides containing the $\beta_2$-binding sequences inhibited $\alpha$-$\beta_2$ and $\delta$-$\beta_2$ interactions in vitro. Based on the identification of this sequence, a conserved SLF motif in the $\delta$ subunit was proposed to be involved in binding of $\delta$ to $\beta_2$ during loading (11). The subsequent crystallographic structure of the $\beta$-$\delta$ complex confirmed that L73F74 of the SLF motif in the E. coli $\delta$ subunit penetrate a hydrophobic pocket on the $\beta_2$ surface (13). Experimental evidence suggests that the $\alpha$ subunit binds at the same site in a similar manner (14). Binding of the two proteins to overlapping sites underpins the switch in replication whereby $\beta_2$ is passed in the clamp loader to $\alpha$.

Having identified and experimentally validated the pentapeptide motif we looked in families of proteins related to known $\beta_2$-binding proteins for conserved related motifs and for new families with such motifs located at the amino and carboxyl termini. Using this search strategy, we identified several new families of putative $\beta_2$-binding proteins, including new families of probable DNA repair polymerases DnaE2 and DinB3 and a family of unknown function Duf72. The binding of members of the DnaE2, DinB3, and Duf72 families to $\beta_2$ has not been demonstrated experimentally. Many species, such as Streptomyces species, Pseudomonas aeruginosa and Agrobacterium tumefaciens C58, contain one or more copies of members of DnaE2. This family is implicated in DNA repair and mutagenesis, and in Mycobacterium tuberculosis its expression is linked with the development of drug resistance (15). Mycobacteria species lack many of the inducible DNA repair enzymes.

During the course of the bioinformatics analysis a second possible $\beta_2$-binding motif emerged. Around this time a new family of putative $\beta_2$-binding proteins, the Hda family, was described. Most Hda family members have a hexapeptide motif close to the amino-terminus. The role of this hexapeptide motif has been confirmed experimentally, from Hda (16) and DinB (17). Mutagenesis of the proposed $\beta_2$-binding sites of the repair polymerases Pol II, Pol IV and Pol V ablated translesion synthesis in vivo (18). A recent structure of a fragment of a repair polymerase, Pol IV, containing a hexameric $\beta_2$-binding motif revealed that two leucines are involved in similar interactions with the $\beta$-binding surface as the L73F74 of the $\delta$ subunit (19). Thus, all three types of $\beta_2$-binding motif, $\delta$, pentapeptide and hexapeptide have a conserved core of interaction at the same site on $\beta_2$.

Are all $\beta_2$-binding peptides equal?

The conservation of the $\beta_2$-binding peptides and the binding site on $\beta_2$ allows the proteins from different species to interact. S. aureus, S. pyogenes and B. subtilis PolC subunits can all use E. coli $\beta_2$ as their sliding clamp (1, 20, 21). However, the interspecies exchange does not always apply. The E. coli polymerase (at) cannot use $\beta_2$ from the other species (21). The H. pylori $\delta$ subunit does not interact with E. coli $\beta_2$. The E. coli clamp loader cannot load S. aureus $\beta_2$ (21) and the S. pyogenes clamp loader cannot load E. coli $\beta_2$ (1). These results indicate that underlying mechanisms associated with some structural features of the E. coli $\beta_2$ clamp may not apply to bacteria outside the gamma proteobacteria.

Experimental data using peptides with native sequences derived from a number of $\beta_2$-binding proteins and modified peptides have shown that the peptides with sequences closest to the consensus sequence have the strongest $\beta_2$-binding ability (22). We have attempted to correlate the ranking of the $\beta_2$-binding sequences in the different families of $\beta_2$-binding proteins with their potential function and competition for the $\beta_2$-binding site. The distribution of conservation to the $\beta_2$-binding sequences compared to the consensus sequence was analysed on a protein family by family basis. Some protein families, such as PolC, PolB and DinB2, had very high ranking conserved motifs, whilst other families, such as MutS and DnaE2, had generally poorly conserved motifs (Table 1).

Although the $\alpha$ subunit (DnaE) is the major replicative DNA polymerase in most species, its $\beta_2$-binding sequence has an intermediate ranking of conservation of the consensus $\beta_2$-binding sequences relative to PolC proteins (Table 1). This was rather surprising as the DnaE enzyme probably occupies $\beta_2$ more than any other enzyme in the cell. In contrast to all of the other enzymes the $\beta_2$-binding peptide motif in the DnaE proteins is not located close to, or at, the amino or carboxyl terminus. In fact, the $\beta_2$-binding motif is flanked by two DNA-binding domains, suggesting that the polymerase subunit could contact
DNA on both sides of $\beta_2$ providing additional stability to the complex. Binding at this site may be supported by a second proposed $\beta_2$-binding site located at the carboxyl-terminus of the $\alpha$ subunit (23). However, this site is not conserved across a wide range of species. In contrast, the high ranking $\beta_2$ binding peptide in the PolC enzyme is located right at the carboxyl-terminus of the protein. The DnaE enzymes in PolC containing species have an even weaker predicted $\beta_2$-binding motif than the other DnaE enzymes. It is interesting to speculate that the distinct motif in most DnaE proteins represents a compromise between the $\beta_2$-binding requirements for leading and lagging strand synthesis. In PolC-containing species, DnaE's weaker interaction with $\beta_2$ is allowed because of its lagging strand specific role.

Plasmid-borne $\beta_2$ binding proteins

A number of families of plasmid-borne replication (Rep) proteins with various levels of ranking of $\beta_2$-binding motifs have also been identified. The distribution of families of Rep proteins containing potential $\beta_2$-binding sites is sporadic, but the sequences are conserved across members of a family or subfamily. To the best of our knowledge, the binding of plasmid Rep proteins to $\beta_2$ has not been investigated, although many Rep proteins have been shown to bind to other components of the DNA replication apparatus. Interestingly, no episome bearing both a Rep protein with a putative $\beta_2$-binding site and a repair DNA polymerase (of either the DinB/UmuC/Rad30/Rev1 or the DnaE family) has been described. It is not clear what benefit accrues to plasmids carrying repair DNA polymerases. Current theory is that the cells acquire increased capacity to repair damaged DNA, and that the enzyme is not involved in the replication of the plasmid DNA. Interestingly, the plasmid-borne enzymes appear to have $\beta_2$-binding motifs that more closely match the consensus sequence than do the chromosomal copies.

Summary

There is clear conservation of the general architecture of the clamp loader and the sliding clamp across all eubacterial species. The two apparently interchangeable peptide $\beta_2$-binding motifs and their common binding site on $\beta_2$, appear to represent a universal system across all eubacteria. The many interacting partners provide a strong selective pressure for conservation. The multiplicity of $\beta_2$-binding proteins in one cell, all predicted to bind at the same site on $\beta_2$, raise questions concerning the nature of the regulation of these interactions. However, the extent of $\beta_2$ sliding clamp use by polymerases involved in replication, repair or recombination of duplex DNA is quite different between different species.

Factors influencing competition for the site on $\beta_2$ may be: first, the site on the duplex DNA at which $\beta_2$ is located – a lesion, a mismatch site, or a replicative primer site; second, the dynamics of other proteins in proximity to the site – as part of transient active complexes or stalled, disassembling complexes; and third, the intrinsic ‘competitiveness’ of the $\beta_2$-binding motif of the protein. Peptides with sequences closest to the pentameric consensus motif, QL[D/S]FL, or the hexameric consensus motif, QLXLXL, appear to be bind $\beta_2$ more effectively (22).

These and other factors affecting $\beta_2$’s role in DNA metabolism may be studied in the various repertoires of $\beta_2$-binding proteins in different bacterial species. For example, H. pylori has only two identifiable $\beta_2$-binding proteins: the $\alpha$ and $\delta$ subunits of the replicase, which are the bare minimum required for DNA replication. M. tuberculosis lacks inducible DNA repair enzymes. Its DnaE2 protein may have evolved to assume part of this role. Both these bacteria are fastidious, slow growing organisms.

In contrast, E. coli, A. tumefaciens and B. subtilis have large cohorts of repair enzymes, some carried on plasmids. These species are well adapted to diverse environments. For DNA replication in the Gram positive organisms, PolC has acquired the role of a highly processive, accurate, replicative polymerase and has generally highly-ranked $\beta_2$-binding motifs. The DnaE enzyme in these organisms appears to have assumed the role of an error prone, less processive polymerase dedicated to the lagging strand. Are Gram positive organisms, with their ‘leakier’ cell membranes, more likely to suffer DNA damage? The Gram negative organisms have retained a more efficient DnaE enzyme as the major replicative enzyme.

Thus far, DNA replication, repair and recombination of few organisms have been studied in any depth. The environmental pressures on the evolution, selection and usage of the $\beta_2$-binding proteins are not understood. Uncovering the roles of these factors will become essential in understanding and predicting the evolution of new phenotypes in bacteria and their plasmids in medical and agricultural environments.

References


References continued on page 21