The BRCT domain from the large subunit of human Replication Factor C

Kobayashi, Masakazu

Summary

Every time the cell divides, DNA, which carries the genetic information, must be duplicated. Accurate duplication of the genetic material is carried out by many proteins, each of which participates in a specific step of a complex process leading to successful DNA replication. Since the development of an *in vitro* replication system using the simian tumour virus SV40 in the late 1980s, it has been possible to study protein-protein interactions in eukaryotic DNA replication. Replication Factor C (RFC) is a five protein complex involved in initiating and regulating new DNA synthesis by loading another protein, called PCNA, onto DNA. RFC binds the replication site of DNA, and then using ATP-hydrolyses as energy sources RFC opens the ring shaped PCNA molecule and clamps it around the DNA strand. The loaded PCNA tethers other proteins to DNA, and as a result, PCNA forms an important platform for recruiting DNA modifying enzymes, such as DNA polymerases, to sites of replication or repair. Through stable anchoring to DNA *via* PCNA, DNA polymerases efficiently synthesize new DNA (**Chapter 1**).

Due to its obvious importance in DNA replication and repair, much of the studies on RFC have focused on its mechanism of PCNA loading. The work described in this thesis however, deals with DNA binding by the N-terminal portion of the large subunit of RFC, which has no apparent role in PCNA loading. This region of RFC (amino acids (aa) 403-480) shares sequence homology with the family of BRCT (*BR*CA1 *C-T*erminal) domains, which consist of roughly 90 aa's and are found throughout eukaryotic and prokaryotic DNA replication and repair proteins. Although the aa sequence of BRCT domains is poorly conserved, they share a highly conserved three dimensional protein fold. It is perhaps because of this sequence diversity that a variety of functions are now known to be associated with the members of this family. Such functions include acting as an adaptor module interacting with other proteins, peptide binding module and a DNA binding module (**Chapter 1**).

The first part of this thesis describes both the region of the large subunit of RFC and the structural determinants of the DNA that are required for productive interaction. In order to determine the optimal DNA binding region, several protein fragments encoding of the large subunit of RFC were generated and subsequently assessed for their double stranded DNA (dsDNA) binding. dsDNA binding is achieved only when the RFC constructs retained an intact BRCT domain and extra amino acids N-terminal to the

domain. This "BRCT region", encompassing residues 375-480 of the RFC large subunit (also called p140(375-480)), forms a tight 1:1 protein-DNA complex (Kd ~10nM) that binds specifically to the 5' phosphorylated end of dsDNA (**Chapter 2**). Only a single type of protein-DNA complex was observed. Surprisingly, however, when the protein fragment spanned an additional 60 amino acids at the C-terminus of the BRCT region, the DNA binding specificity was altered to bind indiscriminately to dsDNA, resulting in multiple types of protein-DNA complexes.

The replacement of several amino acids of the BRCT region using site-directed mutagenesis revealed a requirement for not only the few highly conserved aminoc acids, but also some non-conserved amino acids, for DNA binding, The non-conserved amino acids are found within the N-terminal sequence, while the conserved amino acids are exclusively found in the BRCT domain part of the BRCT region (**Chapter 2**). The conserved residues are also shared among BRCT domain from the bacterial DNA ligases, which have also been implicated in DNA binding. Furthermore, those conserved amino acids are positional equivalents of those used in binding to the phosphate-moiety by the tandem BRCT domains from BRCA1 when bound to a phospho-peptide ligand. Therefore it was speculated that the residues responsible for phosphate binding are conserved between the dsDNA and the phospho-peptide binding BRCTs.

To gain better insight as to how the protein-DNA complex is formed, it was essential to elucidate the three dimensional structure of the DNA-protein complex. For the structural elucidation, multidimensional NMR experiments were employed (**Chapter 4**). From the resulting NMR spectra, we obtained essentially complete ^{1}H, ^{15}N and ^{13}C resonance assignments of the protein moiety of the 19 kDa p140(375-480)-DNA complex. Prediction of the secondary structures based on the chemical shift indices and on the pattern of backbone NOEs from the NMR data indicated the presence of a consensus BRCT fold with an extra α helix within the aa's N-terminal to the BRCT domain.

Based on the chemical shift assignment of human RFC p140(375-480) bound to dsDNA (**Chapter 4**), with the aid of the computer program CANDID/CYANA, we calculated an ensemble of 20 structures of the protein (**Chapter 5**). The protein consisted of a well-defined core, corresponding to the consensus BRCT fold, and an N-terminal α helix, whose spatial orientation with respect to the core of the protein was less well defined.

The structure alignment between the RFC BRCT domain and the distantly related N-terminal BRCA1 BRCT domain (of the tandem domain structure) revealed a remarkable conservation of amino acids that specifically interact with the phosphate of the phosphopeptide bound to the BRCA1 BRCT domain. However, a lack of data from the DNA undermined the NMR-based structure determination of the complete protein-DNA complex. We therefore obtained a model of the protein-DNA complex using the empirical protocol, HADDOCK, which drives docking of molecules using ambiguous restraints derived from mutagenesis (**Chapter 3**), amino acid residue conservation, structural conservation and ambiguously assigned intermolecular NOEs (**Chapter 5**).

In our HADDOCK generated model of the protein-DNA complex, the BRCT region is bound to the DNA with the positively charged surface of the BRCT domain interacting with the 5' phosphate of the DNA while the N-terminal helix forms several interactions with the phosphate backbone, as well as the bases, of the DNA major groove. The model supports all of the biochemical and NMR spectroscopic data. DNA binding of mutants designed based on the model complex structure further supported the validity of this model (**Chapter 5**).

Although the approaches taken in this study did not successfully generate a three-dimensional structure of the protein-DNA complex based purely on NMR data, the evidence from other studies strongly supports the empirically determined model. Thus, if the model is indeed correct, then the remarkable conservation of the mechanism of recognition of the 5'-phospate of DNA and the phosphoserine moiety of the peptide noted in **Chapter 5** may help us to identify other BRCT bearing proteins that are phosphate-dependent DNA binders (**Chapter 6**).